Dottorati di ricerca in Puglia XXXIII XXXIV XXXV ciclo è un intervento cofinanziato dall'Unione Europea a valere sul POR Puglia 2014-2020, Asse Prioritario OT X "*Investire nell'istruzione, nella formazione e nella formazione professionale per le competenze dell'apprendimento permanente*" - Azione 10.4 - DGR 2228/2019 – Avviso 1/FSE/2019 n. 20 del 24/02/2020 (BURP N. 25 DEL 27/02/2020)

UNIVERSITY OF FOGGIA

PhD course in
Translational Medicine and Management of Health Systems

XXXV Cycle

Coordinator: Prof. Teresa Antonia Santantonio

PhD Thesis

Decision Support System (DSS) for policy formulation in the Apulian regional health system

Tutor: Prof.ssa Caterina Tricase

Co-Tutor: Prof. Nicola Faccilongo

PhD candidate: Vito Santamato

Academic years: 2020/2023

To my father, my mother, and my sister,
pillars of my life and inexhaustible
sources of love and inspiration.

"Posso fare cose che tu non puoi,
tu puoi fare cose che io non posso;
insieme possiamo fare grandi cose".

Madre Teresa di Calcutta.

CONTENTS

SESSION III
ASSESSMENT OF PUBLIC HEALTH PERFORMANCE IN RELATION TO HOSPITAL ENERGY DEMAND, SOCIO-ECONOMIC EFFICIENCY AND QUALITY OF SERVICES 118

SESSION IV
ANALYSIS OF OPERATIONAL EFFICIENCY IN PUBLIC HOSPITALS: AN INNOVATIVE MACHINE LEARNING APPROACH 138

GLOSSARY OF ABBREVIATIONS

ABBREVIATION	MEANING
ANOVA	Analysis of Variance
ASL	Azienda Sanitaria Locale
AUC	Area Under the ROC Curve
AUSL	Azienda Unità Sanitaria Locale
CFA	Confermative Factor Analysis
CPD	Cluster - Principal Component - Data Envelopment
CPDA	Cluster - Principal Component - Data Envelopment - ANOVA
CRS	Constant Return of Scale
CSBO	Circulatory System Based Optimization
DEA	Data Envelopment Analysis
DMU	Decision Making Units
DRG	Diagnosis Related Groups
DSCF	Dwass-Steel-Critchlow-Flinger
DSS	Decision Support System
EFA	Exploratory Factor Analysis
IRCCS	Istituti Di Ricovero e Cura a Carattere Scientifico
IRKO	Improved Runge-Kutta Optimization
ISTAT	Istituto Nazionale di Statistica
KMO	Kaiser–Meyer–Olkin
LEA	Livelli Essenziali di Assistenza
MSA	Measure of Sampling Adequacy
MSE	Mean Squared Error
NHS	National Health Service
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
PTE	Pure Technical Efficiency
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
SE	Scale Efficiency
SSN	Servizio Sanitario Nazionale
TE	Technical Efficiency
TLDEA	Two Levels Data Envelopment Analysis
VRS	Variable Return of Scale

ABSTRACT

Introduction: Crossing the boundaries of innovation in healthcare, this research delves into the depths of hospital efficiency and health policies in the Apulia region of Italy. With an approach that skillfully intertwines advanced machine learning techniques and data analysis, the beating heart of this work is the adoption of the revolutionary Cluster Principal Data-Envelopment and ANOVA (CPDA) method. This methodology not only promises a holistic evaluation of hospital efficiency but also pays meticulous attention to the perceived quality of services and their resilience in the face of the population's evolving needs.

Materials and Methods: The journey begins with a thorough analysis of healthcare performance in Apulia, where CPDA becomes the tool to decipher efficiency and quality perception, unveiling the vital importance of service adaptability. This initial phase opens the door to a detailed comparison between the healthcare systems of Apulia and Emilia-Romagna, where efficiency and quality parameters intertwine to explore operational practices and resource management in different regional contexts. Moving forward, attention shifts to hospital energy efficiency and its socio-economic impact, bridging the gap between energy resource management, healthcare economics, and service quality. A further qualitative leap is achieved with the introduction of neural network models for an in-depth examination of operational efficiency in hospitals, considering variables such as energy costs, personnel costs, and the effectiveness of medical device utilization.

Results: Efficient structures emerge at various levels, while technical efficiency is decomposed into pure technical efficiency (PTE) and allocative efficiency (SE), painting a landscape of significant differences in efficiency across different hospital levels. The integration of the Particle Swarm Optimization (PSO) algorithm into the CPDA model elevates the model's discriminative capacity, refining the performance evaluation. Furthermore, a direct correlation between hospital efficiency and the perceived quality of healthcare is revealed, indicated by a negative linear relationship between scale efficiency and patients' propensity for hospitalization. The analysis then delves into the complex interaction between hospital organizational structures, patients' propensity for hospitalization, and the resulting energy costs. The increase in medical devices in public hospitals in Apulia is directly linked to rising energy costs, highlighting the importance of a balanced approach towards the adoption of new medical technologies.

Conclusions: The research proposes a decision support system for healthcare in Apulia, based on advanced analytical methodologies and data-driven decisions. This system aims to optimize the efficiency, effectiveness, and sustainability of healthcare services in the region, representing a significant contribution to the field of healthcare analysis. Demonstrating how the integration of advanced machine learning techniques can improve operational efficiency in hospitals and positively influence health policies, the work emphasizes the crucial importance of technological innovation for optimized resource management and evidence-based decision support.

Keywords:
Hospital Efficiency, Machine Learning, CPDA, Neural Networks, Optimization Algorithms, Resource Management.

RIASSUNTO IN ITALIANO

Introduzione: Attraversando i confini dell'innovazione in campo sanitario, la presente ricerca si immerge nelle profondità dell'efficienza ospedaliera e delle politiche sanitarie nella regione Puglia in Italia. Con un approccio che intreccia abilmente tecniche avanzate di machine learning e analisi dei dati, il cuore pulsante di questo lavoro è l'adozione del rivoluzionario metodo Cluster Principal Data-Envelopment e ANOVA (CPDA). Questa metodologia non solo promette una valutazione olistica dell'efficienza delle strutture ospedaliere ma pone anche un'attenzione scrupolosa sulla qualità percepita dei servizi e sulla loro resilienza di fronte alle esigenze in continua evoluzione della popolazione.

Materiali e metodi: Il viaggio inizia con un'analisi meticolosa della performance sanitaria in Puglia, dove il CPDA diventa lo strumento per decifrare l'efficienza e la percezione della qualità dell'assistenza, svelando l'importanza vitale dell'adattabilità dei servizi sanitari. Questa fase iniziale apre le porte a un confronto dettagliato tra i sistemi ospedalieri di Puglia ed Emilia-Romagna, dove parametri di efficienza e qualità si intrecciano per esplorare le pratiche operative e la gestione delle risorse in contesti regionali diversi.
Proseguendo, l'attenzione si sposta sull'efficienza energetica ospedaliera e il suo impatto socio-economico, tracciando un ponte tra la gestione delle risorse energetiche, l'economia sanitaria e la qualità del servizio. Un ulteriore salto qualitativo si realizza con l'introduzione di modelli di reti neurali per una disamina approfondita dell'efficienza operativa ospedaliera, prendendo in considerazione variabili quali i costi energetici, i costi del personale e l'efficacia nell'utilizzo dei dispositivi medici.

Risultati: Strutture efficienti si delineano a vari livelli, mentre l'efficienza tecnica si scompone in efficienza tecnica pura (PTE) ed efficienza allocativa (SE), disegnando un panorama di differenze significative nell'efficienza tra i vari livelli ospedalieri. L'integrazione dell'algoritmo di ottimizzazione PSO nel modello CPDA eleva la capacità discriminante del modello, affinando la valutazione delle prestazioni ospedaliere. Inoltre, emerge una correlazione diretta tra l'efficienza ospedaliera e la qualità percepita dell'assistenza sanitaria, rivelata da una relazione lineare negativa tra l'efficienza di scala e la propensione dei pazienti all'ospedalizzazione. L'analisi si addentra poi nella complessa interazione tra le strutture organizzative ospedaliere, la propensione dei pazienti all'ospedalizzazione e i costi energetici risultanti. L'incremento dei dispositivi medici negli ospedali pubblici pugliesi si lega direttamente all'aumento dei costi energetici, sottolineando l'importanza di un approccio bilanciato verso l'adozione di nuove tecnologie mediche.

Conclusioni: La ricerca propone un sistema di supporto decisionale per l'assistenza sanitaria in Puglia, basato su metodologie analitiche avanzate e decisioni guidate dai dati. Questo sistema mira a ottimizzare l'efficienza, l'efficacia e la sostenibilità dei servizi sanitari nella regione, rappresentando un contributo significativo al campo dell'analisi sanitaria. Dimostrando come l'integrazione di tecniche avanzate di machine learning possa migliorare l'efficienza operativa ospedaliera e influenzare positivamente le politiche sanitarie, il lavoro enfatizza l'importanza cruciale dell'innovazione tecnologica per una gestione ottimizzata delle risorse sanitarie e un supporto decisionale basato su prove.

Parole chiave
Efficienza ospedaliera, Machine learning, CPDA, Reti neurali, Algoritmi di Ottimizzazione, Gestione delle risorse.

SESSION I

HEALTHCARE PERFORMANCE ANALYTICS BASED ON THE NOVEL CPDA METHODOLOGY FOR ASSESSMENT OF EFFICIENCY AND PERCEIVED QUALITY OUTCOMES: A MACHINE LEARNING APPROACH

1. INTRODUCTION

Identifying the quality of healthcare is more complex compared to other services because the evaluation is based on the patients themselves and their quality of life (Eiriz & António Figueiredo, 2005). The patient's perceived quality is not a straightforward definition, but rather structured by multiple components that complete its explanatory model. A significant component of perceived quality is the sense of inclusion generated between patient and caregiver. The possible sense of solidarity produced strengthens the patient's well-being (Nygren Zotterman et al., 2016).

Certainly, the ability of the hospital to effectively treat complex and particular conditions, hence a high ratio between the hospital's specialization and the complexity of clinical cases treated, is a determining factor. Another determining factor is the reputation of the doctors who work there. These components can be emphasized by marketing policies pursued (Falavigna & Ippoliti, 2013).

Assessing the determinants of patient choice is challenging due to the multidimensional nature of quality and the limited observability of important attributes. The choice of hospital, influenced by changes in clinical quality, suggests that promoting an informed patient choice, such as disseminating information to the public about hospital quality, can produce beneficial effects even in highly regulated contexts. Patients' sensitivity to changes in quality makes hospitals with better health outcomes more attractive (Lippi Bruni et al., 2021).

The study of patient care facility choices is integrated into the identification of determinants of hospital mobility. Policy makers assume that patients "actively" choose the facility they go to, simultaneously seeking the best quality of care and minimizing the cost to obtain it. The determinants of international hospital mobility are quite heterogeneous: each health system, in a legal-formal sense, is not in a vacuum but rather reflects political choices, social objectives, and territorial peculiarities. Thus, when a health system fails to meet, quantitatively or qualitatively, all the needs expressed by the target population, it endogenously generates a "demand" for mobility, that is, a certain rate of escape (Evangelista, 2016).

Many countries have introduced competition in the hospital care market to improve quality and reduce costs (Oliver & Mossialos, 2005), but the effectiveness of this strategy is still a matter of debate (Berta et al., 2016). Perceived quality of care is an important factor in user choice and drives migration from the South to the North in Italy (Berta et al., 2021). The Covid-19 pandemic has further influenced the resources available for health systems, affecting their performance (Mirmozaffari et al., 2022). To ensure high-quality health services for customers and save costs, some institutions are implementing process optimization and automation strategies. In summary, providing high-quality health services requires continuous adaptation

to the changing needs and priorities of the population, as well as effective management of available resources.
The use of data mining algorithms in this context allows for the analysis of large amounts of data, identification of patterns and correlations, and provides recommendations for improving healthcare and healthcare system efficiency. Through data analysis, it is possible to improve the quality of care and healthcare system efficiency, taking into account patient satisfaction as an indicator of quality (Koh & Tan, 2005). Furthermore, the use of advanced computer tools enables the efficient and precise management and analysis of large amounts of data, discussing vital issues involving Data Mining as an important applied technique in solving healthcare problems (Ekwonwune et al., 2022).
Computer science and data mining can help identify the factors that influence patient choice regarding hospital mobility. Analysis of healthcare data can identify patient preferences and hospital performance in terms of efficiency, quality of care, and patient satisfaction. Moreover, collected data can be used to make informed decisions on hospital mobility and the distribution of healthcare resources in a fair and efficient manner. Information technology has enabled the analysis of patient mobility on a large spatial scale, with significant practical implications. Particularly in developing countries with limited healthcare resources, optimizing patient mobility is a crucial goal for policymakers (Ding et al., 2023).
The present paper is organized as follows. After this introduction, which describes the problem and the goals of the work, the next paragraph presents the methodological background in which the work is based, and a description of the context to which it has been applied. This is followed by a chapter with the details of the study conducted and the most interesting results. These are then critically discussed in the next chapter, along with an evaluation of the innovativeness, potential, and limitations of the proposed approach. The conclusions complete the work by evaluating the implications of the methodology for decision support at various levels and outlining possible future developments.

2. BACKGROUND

Within the healthcare sector, the ongoing quest to optimize operational efficiency and ensure high-quality patient care has become an increasingly pressing challenge. In this context, efforts to evaluate the efficiency of healthcare processes and the perceived quality of outcomes have taken on fundamental importance. However, traditional evaluation methodologies often encounter limitations in capturing the complex interplay between efficiency and quality.
To address this complexity, opportunities have emerged in recent years through the application of advanced data analysis techniques, machine learning, and innovative methodologies. This study focuses on introducing the "Cluster Principal Data-envelopment and Anova analysis" (CPDA) methodology, an innovative approach designed to thoroughly assess the efficiency of healthcare operations and the perceived quality of results by harnessing the potential of machine learning.
Efficiency in the healthcare sector translates to resource optimization, improved workflows, cost reduction, and waste minimization. Simultaneously, quality of care encompasses parameters such as patient satisfaction, adherence to clinical guidelines, and achieving positive outcomes. However, the interaction between efficiency and quality is complex and often involves balanced choices among different objectives.

Incorporating machine learning into the CPDA methodology offers the opportunity to analyze substantial and heterogeneous data, extract models and insights that might otherwise remain hidden. Machine learning algorithms excel in identifying non-linear relationships, predicting trends, and identifying factors influencing both efficiency and quality of outcomes. Through the implementation of machine learning, the CPDA methodology can provide a data-based foundation for informed decision-making. This supports healthcare administrators, policymakers, and professionals in seeking optimal decisions to optimize care processes and enhance patient experiences.
Furthermore, the CPDA methodology's emphasis on perceived outcome quality acknowledges the importance of patient-centered care. Patient experiences, their satisfaction, and achieved outcomes constitute valuable perspectives that complement traditional clinical measurements. The CPDA approach, which merges clinical data with patient perspectives, promises a comprehensive evaluation of healthcare performance.
This research study contributes to the field of healthcare analysis, presenting the CPDA methodology as an innovative and comprehensive approach to assessing healthcare performance. By combining the power of data analysis and machine learning, this methodology based on Cluster, Principal components, Data-envelopment and Anova analysis has the potential to reshape how efficiency and quality are evaluated within the healthcare context.
A key factor in our investigation was to conduct analyses within a machine learning environment for even more rigorous and accurate insights. This approach, involving the fusion of PCA-DEA and statistical algorithms with machine learning, has the potential to radically reshape how efficiency and quality are assessed in the hospital context.
Through the application of cluster analysis, factorial analysis, and reliability analysis, a comprehensive investigation was conducted on the structure and coherence of the data. Cluster analysis revealed clear structures and significant correlations among variables, while factorial analyses identified consistent latent factors supported by robust correlations within each cluster. Reliability analysis further reinforced these results, demonstrating high internal consistency of measures within both clusters. Building on these findings, subsequent PCA analysis was performed considering a single principal component for each cluster, providing a robust methodological foundation for data interpretation.
The subsequent sections of this study will delve into the components of the methodology, its application in real healthcare scenarios, and the potential impact it could have on the entire healthcare sector. Through this research, our aim is to lay the groundwork for a more efficient, patient-centered, and data-driven healthcare system.

2.1 RELATED WORKS

Previous studies in the field of healthcare efficiency measurement have predominantly relied on Data Envelopment Analysis (DEA) compared to deterministic and stochastic frontier analysis methods. This preference for DEA is attributed to its flexibility in specifying inputs and outputs and formulating production correspondences, which proves advantageous when data availability is limited. The application of DEA in healthcare efficiency measurement has been widely adopted in various studies. A significant step forward was taken in the study conducted by Hajiagha et al. in 2023, which examined the application of the three-

stage, Principal Components – Factor – Two Levels Data Envelopment Analysis (PCA-FA-TLDEA) methodology. This study catalyzed our approach, leading us to use the PCA-DEA combination in our research effort. This decision conferred greater discriminatory power within the evaluation model, further enriched by the incorporation of ANOVA statistical algorithms. These were leveraged to examine the efficiency outcomes attributed to hospitals based on their network membership and hospital type.
Data Envelopment Analysis (DEA) has been applied to evaluate the effectiveness of COVID-19 pandemic management strategies. The study conducted by Mohanta et al. in 2021 provides valuable insights into varying levels of efficiency in COVID-19 crisis management among different Indian states. By applying DEA, the study contributes data-driven understanding of the effectiveness of strategies employed during the pandemic, aiding policymakers in identifying successful approaches and optimizing resource allocation for a more efficient response to similar challenges in the future.
Several studies have extensively utilized DEA to evaluate the efficiency of public healthcare systems, while highlighting the importance of refining methodologies, considering various factors, and staying updated on evolving research trends to effectively measure and enhance healthcare service efficiency in the public sector (Jung et al., 2023).
The present study is following the same patterns as described by Hagjiagha et al. (2023) in their study on Iranian hospitals, demonstrates its suitability for accurately calculating the technical efficiency of hospitals in the Apulian region, and also applied to the New Zealand District Health Boards (Andrews, 2022). It includes statistical methodologies, including the combination of Principal Component Analysis (PCA) and Data Envelopment Analysis (DEA).
In the healthcare sector and other social fields, the application of principal component analysis is a widely used approach in the literature to calculate the efficiency of Decision-Making Units (DMU). This approach provides a comprehensive perspective to evaluate the performance and efficiency of healthcare services provided by hospitals.
The combination of PCA and DEA is particularly suitable for measuring the efficiency of complex systems, such as investment efficiency in hospital construction (Lan et al., 2021).
Combining machine learning techniques, such as clustering, with the traditional DEA approach is important for gaining a more nuanced view of learning process performance. This hybrid approach demonstrates how it can aid hospital administrators in recognizing best practices, efficiently allocating resources, and enhancing learning outcomes in teaching hospitals (Hasni et al., 2021). A recent study aimed to examine the efficiency of technological, healthcare, and consumer funds through a global DEA approach, providing key insights into how resources are utilized in these sectors (Proença et al., 2023).
A recent study conducted in Italy evaluates the performance of public healthcare services considering factors like hospital energy demand, socioeconomic efficiency, and service quality. The machine learning workflow composed of Principal Component Analysis, Linear Regression, and ANOVA Analysis algorithms, applied to key variables used in calculating hospital efficiency using the DEA method as identified in literature, offers a valuable perspective for administrators and healthcare policy makers, enabling informed decisions that promote efficiency, sustainability, and improved service quality in the public healthcare sector (Santamato et al., 2023).

Building upon this context, to further enrich our analysis, we have chosen to utilize the CPDA methodology within a machine learning framework to achieve a more thorough and accurate assessment of learning process performance. This hybrid approach promises to provide a strong methodological foundation for interpreting data and obtaining more reliable and relevant results, further contributing to enhancing healthcare service efficiency and quality within hospital settings.

2.2 CPDA IN HOSPITAL EFFICIENCY EVALUATION: ADVANTAGES, IMPLICATIONS, AND LIMITATIONS

The CPDA methodology, harnessing soft computing techniques like cluster analysis, Principal Component Analysis (PCA), Data Envelopment Analysis (DEA), and Analysis of Variance (ANOVA), emerges as a pioneering approach in hospital efficiency evaluation. This synergistic blend of techniques offers a plethora of advantages, setting the stage for profound insights and potential implications in the results.

Advantages of adopting the CPDA methodology:

- **Analytical Versatility:** The fusion of these techniques provides a multifaceted view, capturing both macro and micro aspects of hospital efficiency.
- **Advanced Segmentation:** Cluster analysis allows for segmentation of hospitals based on similar characteristics, facilitating homogeneous and relevant comparisons.
- **Complexity Reduction:** PCA condenses essential information by reducing data dimensionality, making the analysis more manageable and less prone to multicollinearity errors.
- **Relative Efficiency:** DEA evaluates the relative efficiency of decision-making units, providing a clear picture of how each unit performs relative to "best practices."
- **Identification of Significant Variables:** ANOVA allows determining which factors significantly impact efficiency, providing valuable insights into potential areas of intervention.

Implications for Results:

- **In-depth Analysis:** Due to its comprehensive nature, the CPDA methodology can uncover interactions and trends that might remain hidden with traditional methodologies.
- **Robust Results:** The integration of multiple techniques enhances result robustness, reducing the risk of drawing incorrect conclusions.
- **Interpretative Complexity:** While the CPDA methodology offers a rich overview, its complexity might make interpreting results challenging, demanding a sound understanding of each involved technique.

Additional Details:

Through the combined use of these techniques, CPDA provides a comprehensive view of hospital efficiency, not only evaluating efficiency itself but also underlying factors and interactions influencing it. For instance, the PCA and DEA fusion allows not only for efficiency evaluation but also for identifying the primary variance directions in the data contributing to such efficiency. Simultaneously, ANOVA's integration allows isolating and evaluating specific factors or variables' importance.

Despite its myriad advantages, the CPDA methodology also has inherent limitations:
Computational Complexity: The combined use of various techniques can escalate computational complexity, potentially making the analysis longer and more error-prone, especially with large datasets.
Interpretation: Combining multiple techniques might complicate result interpretation. A deep understanding of each technique is essential for accurate interpretation.
Outlier Sensitivity: Techniques like DEA are especially sensitive to outliers. A single anomalous data point can significantly skew results, leading to potentially misleading conclusions.
Efficiency Assumption: DEA, in particular, assumes decision-making units are efficient, which might not always be the case in reality. This can lead to some units' efficiency overestimation.
Segmentation Limitations: While cluster analysis provides segmentation based on similar characteristics, there's always a risk of suboptimal segmentation or misinterpreting the resulting clusters.
Data Dependency: Like all analysis techniques, CPDA heavily relies on input data quality. Inaccurate, incomplete, or misleading data can yield incorrect results.
Generalizability: Due to the analysis's specificity and depth, results obtained in one context or dataset might not easily generalize to other contexts or datasets.
Overfitting Risk: Using advanced techniques, especially when integrated into a machine-learning framework, can lead to overfitting, where the model too closely fits the training data and loses its ability to generalize over new data.

2.3 APPLICATION CONTEXT

The region of Apulia represents an extremely interesting context for studying hospital efficiency in terms of quality perceived by resident patients, for several reasons. Firstly, the region has undergone a significant healthcare system reform in recent years, resulting in a complete reorganization of the entire system with the aim of improving efficiency and quality of healthcare services. Additionally, Apulia has a diverse population, composed of a variety of heterogeneous groups, making it an ideal area for studying the relationship between quality of care and socioeconomic differences.
The Regional Healthcare Service of Apulia is composed of six Local Health Authorities (ASL) covering the entire region. Hospital facilities are divided between the regional public network and the accredited private network of the Local Health Authority, representing 52.5% and 47.5% of the total respectively. The regional public network includes 24 ASL Direct Hospital Facilities, an Integrated Hospital Company with the National Health System (NHS), an Integrated Hospital Company with the University, and a Public Institute for Research and Treatment. The accredited private network includes 26 accredited private clinics, two Private Scientific Institutes for Research, Treatment and Care, a Public Scientific Institute for Research, Treatment and Care, and a Classified or Assimilated Hospital. To classify the hospital facilities in Apulia, there are 5 second-level hospitals, 17 first-level hospitals, 4 Scientific Institutes for Research, Treatment and Care, 9 basic hospitals, and 24 private nursing homes.
With reference to the regional healthcare system, it refers to an organized set of structures, both public and private, accredited and present within a specific region.

This system is considered as a regional healthcare industry, as defined by Falavigna and Ippoliti in 2013.
A considerable and constant flow of patients (and money) moves from southern Italy (especially from Calabria, Campania, Apulia, and Sicily) to selected regions in central and northern Italy. At the micro level, patients migrate when the perceived quality of care in another region compensates for the costs of migration (Brenna & Spandonaro, 2015). The Apulia Region has also been selected as the sample region in southern Italy, the subject of our study, to study hospital efficiency in terms of perceived quality by patients residing within the regional borders.

2.4 ASSESSING THE EFFICIENCY AND PERCEIVED QUALITY OF HOSPITALS: A CLOSER LOOK AT APULIA

In the realm of academic research, choosing a specific region as a case study is pivotal to ensure relevance and applicability of findings. Apulia, with its unique blend of urban and rural settings, stands out as an ideal benchmark to scrutinize the Italian healthcare system. A significant variable in this backdrop is the quality of care as perceived by local residents. Their insights provide a lens to gauge the efficacy of healthcare facilities and adopted policies. Apulia's distinct geography, coupled with the availability of detailed data, facilitates an in-depth exploration of healthcare dynamics, even when incorporating this dimension. However, it's crucial to weigh the generalizability of outcomes. While conclusions drawn from Apulia might resonate in regions with similar dynamics, they might not be entirely transferable to different settings. Alternatives like Lombardy, Lazio, or regions outside Italy would present varied challenges and opportunities, swayed by factors such as culture and socioeconomic context. Assessing perceived quality in these alternate realities might add nuanced layers to the discourse. Despite potential alternatives, Apulia offers a valuable framework, but caution is essential when extrapolating findings to broader scopes, considering variables that could influence outcomes. The quality of care as perceived by local residents significantly influences the outcomes of any study that assesses hospital efficiency. Here's how it might affect results:
Objective vs. Subjective Measure: While hospital efficiency can be gauged through objective metrics, like length of stay or cost per patient, perceived quality introduces a subjective dimension grounded in patients' firsthand experiences. This difference in perspectives might lead to misaligned results.
Depth of Analysis: Incorporating perceived quality, the study can offer a more holistic view of hospital efficiency. A hospital might seem efficient in objective metrics, but if patients don't perceive high-quality care, it might point to underlying issues or areas for improvement.
Generalizability: Quality perception might vary significantly across different regions or countries due to cultural variances, expectations, and prior experiences. Thus, while results related to hospital efficiency might be generalizable to regions with similar healthcare setups, quality perceptions might not.
Policy Implications: If research indicates a disparity between hospital efficiency and perceived quality, it might suggest the need for policy interventions to enhance quality perception without compromising efficiency.
Practical Relevance: Perceived quality directly impacts patients' trust in the healthcare system. Therefore, even if a hospital is efficient, a low-quality perception might lead to diminished trust and, consequently, reduced adherence to proposed treatments.

Incorporating the perceived quality of care by local residents into a study on hospital efficiency can enrich the findings, providing a more comprehensive view of hospital performance. However, it also introduces additional complexities that must be carefully considered in the evaluation and interpretation of results.

2.5 DATA ANALYSIS TECHNIQUES USED IN THE PROPOSED METHOD

The data manipulation algorithm used to create the dataset is crucial for data preprocessing. There are numerous techniques for data manipulation, such as variable selection, missing value replacement, data transformation, and adding new variables (Kumar et al., 2022) .The main objective is to create a clean and consistent dataset with relevant variables and all necessary information for data analysis. This phase is important to ensure data confidence and the accuracy and significance of subsequent analysis.

The cluster analysis will be carried out on a set of 17 healthcare variables previously identified in the literature, in order to filter them consistently and partition them into clusters that confirm their respective membership in the input and output of the subsequent DEA analysis for evaluating hospital efficiency based on patient-perceived quality. Data clustering algorithms, serve as a valuable tool for grouping data homogeneously, thereby enhancing the overall accuracy of the models employed in data analysis (Dansana et al., 2023).

The standardization algorithm for numerical health variables is a process that transforms numerical variables so that they have a mean of 0 and standard deviation of 1. This is important because when analyzing data from different sources, the units of measurement may be different, and therefore variables may have different scales. Standardization allows all variables to be put on the same scale, so that they can be compared fairly and accurately. Furthermore, standardization is often used as a first step before applying multivariate analysis techniques such as PCA, in order to have a common starting point for all numerical variables. Standardizing data is an important step in data analysis, including the use of the data mining algorithm for PCA. Standardization is necessary to ensure that different variables within the dataset have the same scale, to avoid variables with higher values dominating those with lower values. This can affect PCA and produce inconsistent or misleading results. By standardizing the dataset, a more accurate analysis and better understanding of the data can be obtained (Mohammed et al., 2023).

To ensure the robustness of the model and guide the selection of principal components, we employed two statistical techniques: exploratory factor analysis (EFA) and reliability analysis. EFA allowed us to identify latent factors and assess the internal consistency of measurements within each cluster. Reliability analysis evaluated the internal consistency of the selected variables. These analyses provided a solid foundation for the principal component analysis.

The PCA (Principal Component Analysis) algorithm is a multivariate analysis technique that can be used to reduce the dimensionality of a dataset by identifying linear combinations of input variables that capture most of the variance in the data. This allows for simplification of data understanding and improvement in visualizing relationships between variables.

In the present study, the PCA algorithm was applied to input variables to identify the key factors that influence hospital performance. Additionally, the PCA algorithm was applied to output variables to identify the key factors that influence the quality of hospital care. In particular, using PCA to reduce the number of

original variables into two principal components simplifies the understanding of health data and provides useful information on the effectiveness of hospital organization and the quality of healthcare provided by the hospital.
Applying the DEA output-oriented algorithm to evaluate hospital efficiency (Ferreira et al., 2023) in producing Propensity for hospitalization, considering hospital organization as input and Propensity for hospitalization as output. Patient mobility can be used as an indicator of the quality of service provided (Falavigna and Ippoliti, 2013).
The Analysis of Variance (ANOVA) applied in hospital settings (Noudeh et al., 2022), allows for the evaluation of whether there are significant differences in efficiency values assigned to hospital facilities through DEA analysis, based on the hospital level.
Finally, the use of the CPDA methodology, which combines various data mining algorithms such as CLUSTER, PCA, DEA, and ANOVA, allows for a more complete evaluation of hospital efficiency, integrating the analysis of various aspects such as hospital organization, propensity for hospitalization, hospital size, and type of management (public or private).
The inclusion of neural networks and the Particle Swarm Optimization (PSO) algorithm represents a significant advancement in the evaluation of hospital efficiency. The use of neural networks enhances the model's discriminative power, allowing for a more accurate classification of hospitals in terms of efficiency. At the same time, PSO optimizes the model's parameters, strengthening the robustness and reliability of the efficiency scores. The combined use of the two algorithms serves to optimize the model (Ma et al., 2023).
This complex approach not only provides a more comprehensive evaluation but also serves as a cross-validation tool, making it highly useful for resource management in the healthcare sector.
However, the effectiveness of the CPDA methodology depends on the quality and completeness of the data used in the analysis and the correct application of data mining algorithms. Therefore, it is important to ensure that the data used is reliable and that the algorithms are applied correctly to obtain meaningful and useful information for improving hospital performance.

2.6 DEFINING AND ANALYZING RESEARCH QUESTIONS FOR HOSPITAL EFFICIENCY EVALUATION

In this context, the study aims to present an innovative computer methodology "CPDA" (Cluster-PCA-DEA-ANOVA Analysis), in a data mining environment that, through a single workflow resulting from the combination of statistical-economic algorithms, can answer the following research questions:
Q_1: How does the efficiency of hospitals in the Apulia region, calculated in terms of efficiency scores based on hospital organization, influence the perceived quality of healthcare by resident patients?

Q_2: Can the differences in efficiency and inefficiency among different levels of hospitals in Apulia, based on calculated hospital efficiency scores, provide support for managerial decisions with policy implications?

Q_3: How can the application of the CPDA method, based on the combination of analysis (Cluster-Pca-Dea-Anova) in a data mining environment, enhance the discriminant capability compared to traditional DEA analysis models for hospital efficiency evaluation?

Advantages of the adopted research questions:

- **Specificity:** The proposed questions are closely aligned with the CPDA methodology and the data mining environment, ensuring that the results obtained are directly relevant and applicable.
- **Scope of Analysis:** These questions cover a broad spectrum, from the perception of care quality to hospital efficiency, providing a holistic view of hospital dynamics.
- **Relevance to Decision-Makers:** The questions are crafted to yield results that can have direct policy and managerial implications, offering valuable tools for decision-makers.

Impact on Results:

The adoption of these specific research questions will influence the nature and scope of the results. As the questions are closely tied to the CPDA methodology and the data mining environment, the results will be particularly suited to pinpointing specific patterns and trends in hospital efficiency and the perception of care quality. However, this might also limit the generalizability of the findings to other contexts or methodologies.

In conclusion, while the chosen research questions offer a range of distinct advantages and are closely tied to the study's methodology, it's essential to consider potential limitations and implications for the interpretation and application of the results.

Alternatives to the proposed research questions:

1. How do regional health policies influence the perception of quality and hospital efficiency in the Apulia region?
2. To what extent do the infrastructure and resources available in different hospitals in Apulia impact their operational efficiency?
3. What are the main factors contributing to variability in the perception of care quality among resident patients from different areas of Apulia?

While our primary research questions drive the essence of our study and outline our central focus, we acknowledge the importance of exploring and reflecting upon alternative angles. The complexity and multifaceted nature of factors influencing hospital efficiency and the perception of care quality demand holistic and multi-dimensional consideration. Presenting alternative research questions not only showcases our depth of thought but also offers insights for future investigations in the field. Although these alternative questions are not the main crux of our study, they provide additional perspectives that could further enrich discussions in the realm of hospital efficiency.

The core research questions provide a comprehensive exploration of hospital efficiency and perceived quality of care in Apulia, delving into how efficiency influences perception and the implications for managerial and policy decisions. Highlighting the application of the innovative CPDA methodology in a data mining context underscores the study's cutting-edge relevance. The inclusion of alternative questions showcases a rigorous critical assessment, offering diverse perspectives on regional health policies and hospital infrastructure. These alternatives not only enhance the paper's transparency and depth but also suggest potential avenues for future research, emphasizing the study's thoroughness and the broader applicability of its findings.

3. METHODOLOGY

3.1. VARIABLE SELECTION

The data were collected from various sources including the National Health Service Database (NHS Database) of the Ministry of Health, the National Outcomes Program of the National Agency for Regional Health Services, and the new National Statistical Institute Database (ISTAT Database), for the year 2020. These data contain information on various inputs and outputs considered in various studies on hospital efficiency.

We initially divided all the variables under study into two groups: the group of input variables and the group of output variables. This division was based on a review of the scientific literature. The choice of input and output variables generally depends on the research goals and data availability. In our case, we selected the variables that appeared most relevant for evaluating hospital efficiency in Apulia in 2020, based on previous studies and the scientific literature (Table 1).

Table 1 Hospital Efficiency Evaluation Measures.

Variables	Definition	References	Data sources
INP1_BP	Number of beds provided *	(Santamato et al., 2023; Bağci et al., 2022; Briestensky et al., 2021; Cordero et al., 2023; Falavigna & Ippoliti, 2013; Gutierrez-Romero et al., 2021; Hajiagha et al., 2023; Henriques & Gouveia, 2022)	NHS Database
INP2_BU	Number of beds used *		
INP3_DP	Number of departments planned	(Santamato et al.; 2023; Colombi et al., 2017)	NHS Database
INP4_DU	Number of departments used		
INP5_MN	Number of male nurses	(Santamato et al., 2023; Bağci et al., 2022; Briestensky et al., 2021; Cordero et al., 2023; Falavigna & Ippoliti, 2013; Gutierrez-Romero et al., 2021; Hajiagha et al., 2023; Henriques & Gouveia, 2022)	NHS Database
INP6_FN	Number of female nurses		
INP7_MP	Number of male physicians		
INP8_FP	Number of female physicians		
INP9_HS	Total number of hospital staff		
INP10_RESP	Apulian resident population distributed by hospital physicians	(Santamato et al., 2023; Gutierrez-Romero et al., 2021)	ISTAT/ NHS Database
OUT1_HOS	Number of hospitalizations reported by the Ministry of Health	(Santamato et al., 2023; Bağci et al., 2022; Hajiagha et al., 2023; Kucsma & Varga, 2021)	NHS Database
OUT2_MOB	Intra-regional mobility active by territorial scope	(Santamato et al., 2023; Falavigna & Ippoliti, 2013)	National Outcomes Plan
OUT3_DEA	Number of deaths at 30 days after hospitalization according to the NOP	(Santamato et al., 2023; Briestensky et al., 2021)	

OUT4_INT	Number of interventions according to the NOP	(Santamato et al., 2023; Bağci et al., 2022; Hajiagha et al., 2023; (Kucsma & Varga, 2021)	
OUT5_REA	Number of hospital readmissions at 30 days after hospital discharge	(Santamato et al., 2023; Hajiagha et al., 2023)	
OUT6_INP	Inpatient days	(Fazria & Dhamayanti, 2021; Alharbi et al., 2023)	NHS Database
OUT7_AVA	Available days		

* Day Hospital beds are not included in the analysis. Data collected in February 2022.

We also included the number of physicians, nurses, and total staff as explanatory variables to assess the organization and human resources of the hospital. Additionally, we considered the number of departments and the actual and expected bed utilization as additional input variables to evaluate hospital efficiency. These input variables are crucial in understanding the working environment and available material resources.

On the other hand, we included several output variables that reflect the outcomes of hospital services and have an impact on patient choice. These output variables include 30-day mortality rates after discharge, the number of readmissions within 30 days, the total number of hospitalizations, the number of surgical procedures and the number of impatient/available days. These measures were selected as they represent important indicators for evaluating the quality of care provided by the hospital and contribute to patients' decisions in choosing healthcare facilities.

The selection of both input and output variables was based on previous studies and scientific literature, aiming to provide a comprehensive assessment of hospital efficiency and the factors influencing patient choice.

We also created two derived variables: one for the inputs, representing the catchment area of the hospital, and one for the outputs, expressing the intra-regional active mobility of patients residing in Apulia (par.3.2.).

To confirm the correct assignment of variables to the input and output clusters, we conducted a comparison among various clustering algorithms. The aim was to obtain a data clustering that confirmed their respective belonging to the input and output clusters for the subsequent DEA analysis. The cluster analysis involved a set of 17 healthcare variables selected based on their relevance for the evaluation of hospital efficiency. The results of the analysis confirmed the division of data into two distinct clusters, one containing the input variables and the other containing the output variables, filtering out the variables that conform to the selection from the literature (par.3.5.).

3.2 DERIVEDVARIABLES

Apulian resident population distributed by hospital physicians:

Regarding the input variables, we used the resident population of Apulia as of December 31, 2020, distributed based on the number of hospital physicians, to estimate the size of the facility in terms of user and service basin.

To calculate the Apulian resident population distributed by hospital physicians, we first identified the number of residents in each municipality in Apulia for the year 2020. Next, we identified the municipalities that make up each ASL (Local Health Authority) and summed up the number of residents to obtain the total population for each ASL ($Tot_{res_{Asl}}$). Then, we determined the total number of physicians for each ASL by summing up the physicians working in hospitals within each ASL ($Tot\ Physicians_{ASL}$). Finally, we calculated the Apulian resident population distributed based on the number of hospital physicians for 2020 using the following formula:

$$INP10_{RESP} = Population_{res} = \frac{Tot_{res_{Asl}}}{Tot\ Physicians_{ASL}} x\ Physicians_{Hospitals} \quad (1)$$

Intra-regional active mobility of patients residing in Apulia:
In the present study, a patient's decision to seek treatment where they perceive better quality, subject to their economic availability and the medical offer proposed, is considered. 'Positive mobility' is defined as the flow of 'immigrants,' residents in the Apulia Region in 2020, who reach a hospital located in a different ASL from the one where the patient is a resident. Only intra-regional movements, i.e., within the region, made by patients resident in the region, have been evaluated. Therefore, admissions of non-resident patients are not considered.
To calculate intra-regional active mobility in kilometers, we first calculated the interpolated distance between the patient's ASL of residence and the city where the hospital providing the service is located ($Dist\_km_{Hospital}$). Next, we summed the total number of active hospitalizations for each ASL (($Hospi_{ASL}$) and for each territorial area ($Hospi_{Area}$) within the region. Finally, we calculated intra-regional active mobility in kilometers using the following formula:

$$OUT2\_MOB = Active\ mobility_{infra-regional} = (Hospi_{ASL} + Hospi_{Area})$$
$$x\ \ Dist\_km_{Hospital} \quad (2)$$

This variable of intra-regional active mobility in kilometers represents the distance traveled by patients within the same region to access hospital services provided by different ASLs. It can be used to assess patient preference in choosing a hospital and may be correlated with the perceived quality of hospital services.

3.3 DATA SOURCES AND VARIABLE SELECTION IN HOSPITAL EFFICIENCY ASSESSMENT

Alternative Data Sources:
While experiments drew from recognized databases such as the National Health Service, the National Institute of Statistics, and the National Outcomes Program, there are alternative data sources that might be explored:

- **Surveys and Interviews:** Direct data collection from hospital staff, patients, or administrators could provide detailed, specific insights.
- **Hospital Registries:** These might contain granular data on patient care, resources, and outcomes.
- **Governmental Health Reports:** Often offering a comprehensive overview of regional health metrics and benchmarks.

Advantages of Chosen Data Sources:
The chosen datasets offer several benefits:

- **Reliability and Credibility:** Being official databases, they assure data accuracy.
- **Completeness:** These databases encompass a broad range of variables essential for evaluating hospital efficiency.

- **Standardization:** Data from these sources often follow standard metrics, facilitating comparison and analysis.

Implications on Results:
Adopting data from these established sources ensures robustness in the outcomes. However, results might lean towards macro insights, possibly overlooking finer details that alternative sources like direct surveys might capture.
Alternative Variables:
Beyond the chosen input and output variables, potential alternatives include:
- **Input Variables:** Number of specialized equipment, patient-to-nurse ratio, or details on hospital funding and budget.
- **Output Variables:** Patient satisfaction scores, post-treatment recovery rates, or the frequency of medical errors.

Advantages of Selected Variables:
The variables picked for the study offer:
- **Relevance:** They are directly related to the primary goal of measuring hospital efficiency in Apulia.
- **Literature Backing:** The selection is based on extensive research and prior studies, assuring their relevance.
- **Holistic View:** Collectively, these variables provide a comprehensive view of both the operational aspects (input) and the outcomes (output) of hospitals.

Implications on Results:
The choice of these specific variables ensures that the findings are tailored to the study's objectives. While conclusions will provide a detailed understanding of hospital efficiency as defined by these variables, there might be other facets of efficiency or patient care that the study might not delve into due to variable selection.
In conclusion, while the chosen data sources and variables are suited to the study's objectives, it's crucial to acknowledge the potential nuances and insights that other sources or variables might offer.

3.4 THE METHODOLOGICAL WORKFLOW

Robust data mining tools were incorporated into the methodological flow of the present study (Mirmozaffari et al., 2022). Knime, an open-source data mining tool with graphical capabilities that is widely used by various organizations, was employed during the data processing and transformation phase to generate the dataset with original input and output variables (Fig. 1). Orange, a Python-based software with a graphical front-end design for analyzing experimental data, was used for implementing the CPDA methodology.
The data was extracted and organized into tables using spreadsheets, and then analyzed using KNIME, an open-source platform for data modeling and analysis. The workflow consisted of three sections representing different data sources: the National Outcomes Plan for output data, the National Health System for input data, and the derived indicator for the distribution of the resident population in Apulia from ISTAT source. Various tools and KNIME nodes were used to clean, merge, and filter the data to obtain the final dataset for analysis.

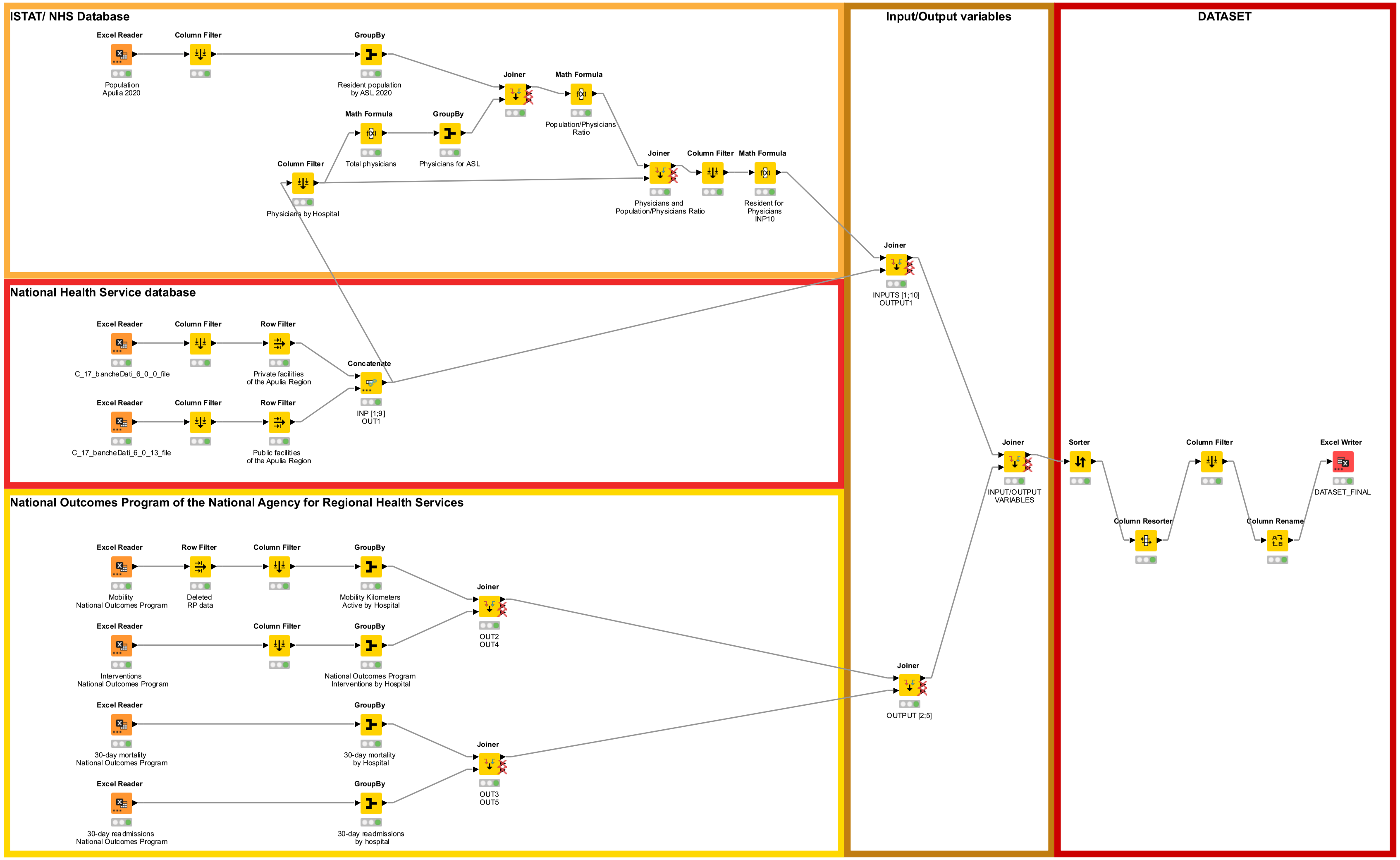


Figure 1 Knime workflow.

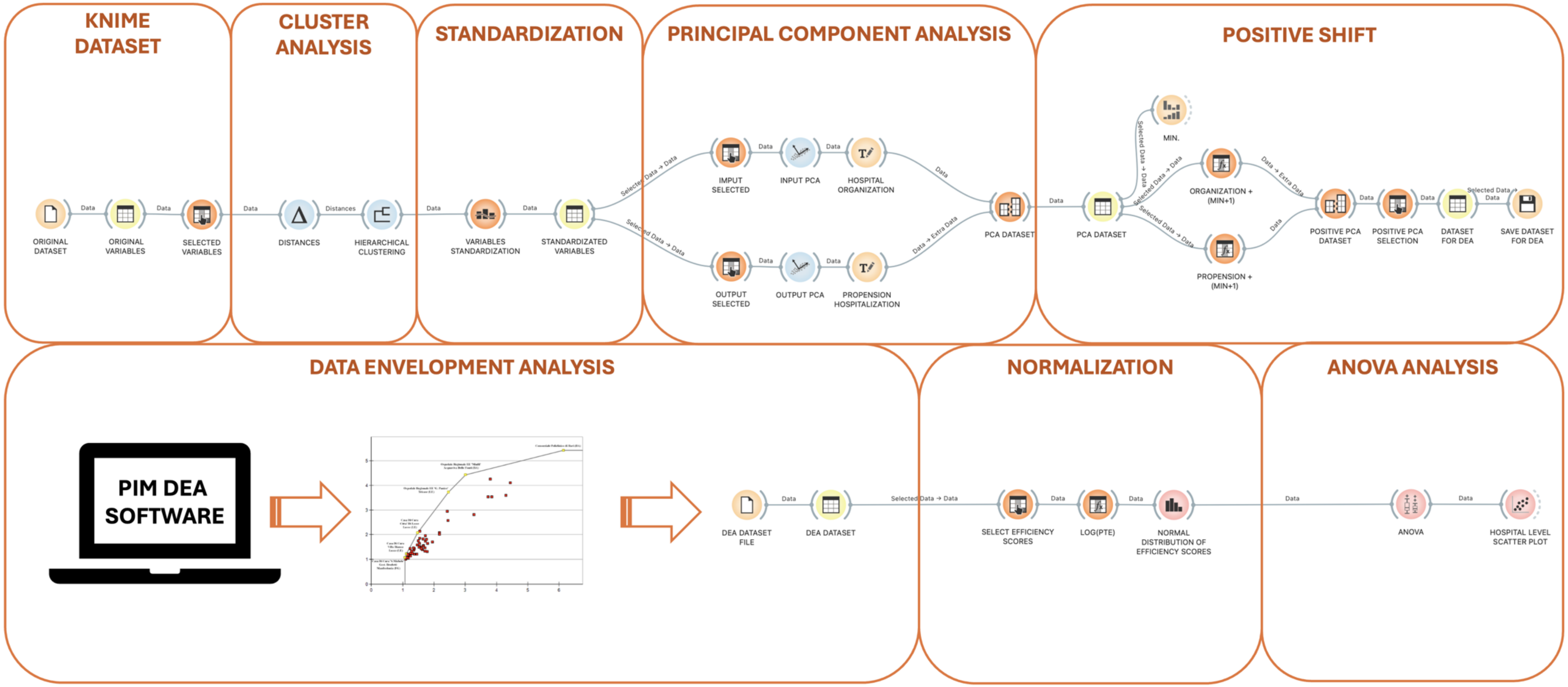


Figure 2 Orange Workflow.

Using Orange, the original variables identified by Knime were first analyzed using cluster analysis to identify input and output groups. Then, the variables were standardized to optimize two separate Principal Component Analyses on the identified clusters. The resulting two components were adjusted to ensure positivity for the subsequent DEA analysis. Finally, ANOVA analyses was carried out using Orange for different hospital networks and levels, respectively (Fig. 2). The PIM-DEA software was utilized for calculating hospital efficiency, utilizing the most recent theoretical developments in Data Envelopment Analysis (DEA) for optimal data analysis (Emrouznejad & Thanassoulis, 2013).
We used the statistical software Jamovi to perform the exploratory factor analysis and reliability analysis applied to the clusters identified by the cluster analysis.
The methodological workflow of the analyses used in this study is depicted in Figure 3.

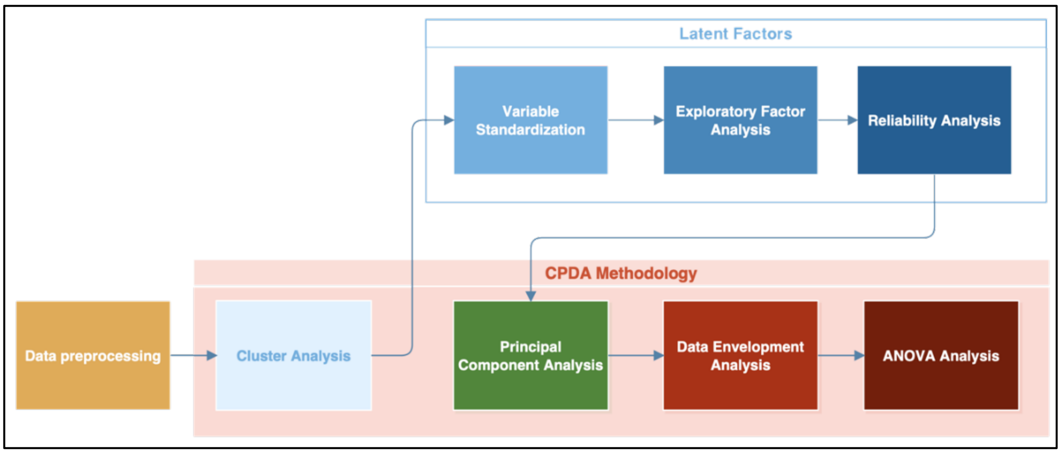


Figure 3 The workflow of the CPDA Methodology.

The "CPDA" analysis (is the acronym for Cluster, Principal Component, Data Envelopment, Anova Analysis) aims to overcome data instability caused by information variation (Xu et al., 2015), by providing an accurate selection of inputs and outputs through the application of cluster analysis to the variables identified in the literature. Subsequently, the application of exploratory factor analysis and reliability analysis to the two identified clusters has allowed for the identification of latent factors that best represent the variables and confirm their coherence within the identified clusters. Furthermore, in the present study, data accuracy has been improved by integrating variables on patients' perceived quality, expressed as propensity for hospitalization and choice of healthcare facility.

3.5 DATA PREPROCESSING PHASE WITH KNIME SOFTWARE

Knime is an open-source data analysis platform. Within it, operations are represented by "nodes". Each node performs a specific function, such as reading data or filtering results. By connecting nodes, users construct visual, modular workflows to efficiently process data. The workflow in Knime (Figure 1) is divided into color-coded subsections for easier comprehension. The first three subsections are categorized based on the data source.

3.5.1. YELLOW BOX: NATIONAL OUTCOME PLAN DATA

Input: 4 Excel tables with data on mortality 30 days post-hospitalization, readmissions 30 days post-discharge, number of interventions, and regional mobility.

Operations Executed:

- Column filtering with Column Filter node.
- Data aggregation with GroupBy node.
- Tables union with Joiner node.
- Ratio computation between aggregated data with Math Formula node.

Output: A table [59x6]. Rows: Name of 59 Apulian Health Facilities. Columns: Facility code, Facility name, Total km covered, Total interventions, Total mortality, Total readmissions.

Table 2 uses a color-coding system ranging from red to green to represent the values in each column. In this coding, red indicates low values while green represents high values. Each variable in the column has a specific meaning: the "Facility Code" is a unique identifier for each healthcare facility, while the "Facility Name" indicates the actual name of the facility. The "Total km covered" represents the total distance traveled for healthcare interventions, highlighting regional mobility. The "Total Mortality" reports the total number of deaths occurring within 30 days of hospitalization, the "Total Interventions" shows the total number of healthcare interventions performed, and the "Total Readmissions" indicates the total number of patients readmitted within 30 days of discharge.

The use of the color-coded matrix in Table 2 offers several advantages. It facilitates an intuitive visualization, allowing for the quick identification of extreme values in various categories and enabling the immediate recognition of facilities with the best or worst performance. Additionally, it allows for a direct visual comparison between different healthcare facilities for each variable, helping to highlight trends and anomalies. This visual support is particularly useful for decision-makers, who can use this information to direct interventions and resources where necessary. Finally, the color-coded matrix condenses a large amount of data into an easily interpretable format, enhancing the overall understanding of the dataset without the need to analyze each numerical value individually.

Table 2. Color-Coded Matrix of the Yellow Box Output Dataset.

Facility Code	Facility Name	Total km covered	Total mortality	Total interventions	Total readmissions
160078	Ospedale Regionale EE 'Miulli' Acquaviva Delle Fonti (BA)	346856,74	2704	2914	1543
160157	Ospedale Della Murgia - Perinei Altamura (BA)	64934,02	988	199	341
160140	Casa Di Cura Anthea Bari (BA)	72932,11	1333	810	309
160147	Casa Di Cura C.B.H. Mater Dei Hospital Bari (BA)	248510,43	2341	1620	457
160087	Casa Di Cura Santa Maria Bari (BA)	130806,77	1472	1331	433
160907	Consorziale Policlinico Bari Bari (BA)	314578,53	4826	2928	1358
160906	Ics Maugeri SPA Societa' Benefit Bari (BA)	35292,39	378	0	258
160901	Istituto Tumori Giovanni Paolo II Bari (BA)	31141,29	764	954	49
160169	Ospedale Di Venere Bari (BA)	181312,93	2036	860	545
160158	Ospedale San Paolo Bari (BA)	196367,61	2283	1019	762
160902	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	56820,26	506	395	9
160098	Casa Di Cura - Villa Lucia Hospital Conversano (BA)	54380,73	73	890	447

160159	Ospedale Monopoli Monopoli (BA)	60873,78	709	100	241
160100	Casa Di Cura ' Monte Imperatore' - Noci (BA)	2673,45	46	0	34
160160	Ospedale Putignano Putignano (BA)	25905,12	726	34	250
160101	Casa Di Cura 'Salus' Brindisi (BR)	20698,19	123	242	204
160151	IRCCS 'E.Medea' - Brindisi (BR)	0,00	0	0	0
160170	Ospedale Perrino Brindisi (BR)	208664,34	2567	1314	774
160162	Ospedale Francavilla Fontana Francavilla Fontana (BR)	87457,47	466	161	170
160161	Ospedale Ostuni Ostuni (BR)	36934,17	527	88	439
160174	Ospedale Andria Andria (BT)	64846,20	1749	925	397
160177	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	49933,13	1270	387	464
160178	Ospedale Bisceglie Bisceglie (BT)	16644,86	458	77	170
160180	Ospedale Opera Don Uva Bisceglie (BT)	3089,58	194	0	113
160105	Casa Di Cura Leonardo De Luca Castelnuovo Della Daunia (FG)	2439,68	38	0	36
160047	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	66195,18	336	90	186
160102	Casa Di Cura Prof. Brodetti Foggia (FG)	14439,10	68	67	36
160125	Casa Di Cura Universo Salute - Don Uva Foggia (FG)	7047,20	27	0	124
160181	Case Cura Riunite Villa Serena-S. Francesco Foggia (FG)	52795,15	2	246	410
160910	Ospedali Riuniti Di Foggia Foggia (FG)	423560,38	1438	1687	998
160102	Casa Di Cura 'S.Michele' Gest. Brodetti Manfredonia (FG)	3128,35	79	3	110
160164	Ospedale Manfredonia Manfredonia (FG)	15887,80	90	27	103
160905	Ospedale Casa Sollievo Della Sofferenza San Giovanni Rotondo (FG)	94278,58	1577	2033	724
160163	Ospedale San Severo - Teresa Masselli San Severo (FG)	96295,50	401	265	292
160152	Casa Di Cura Riabilitativa Euroitalia - Casarano (LE)	180,18	1	0	1
160167	Ospedale Casarano Casarano (LE)	66074,08	743	171	453
160165	Ospedale Copertino Copertino (LE)	24130,71	671	180	355
160110	Casa Di Cura San Francesco Galatina (LE)	11501,62	84	83	73
160063	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	119352,60	985	237	351
160107	Casa Di Cura 'Prof. Petrucciani' SRL Lecce (LE)	37515,01	90	332	54
160150	Casa Di Cura Citta' Di Lecce Lecce (LE)	119174,63	1645	1500	426
160108	Casa Di Cura Villa Bianca Lecce (LE)	20237,78	13	340	271
160109	Casa Di Cura Villa Verde - Lecce (LE)	0,00	0	0	0
160171	Ospedale Lecce 'V. Fazzi' Lecce (LE)	300858,92	4336	1627	1100
160166	Ospedale Scorrano Scorrano (LE)	134705,94	1191	234	528
160080	Ospedale Regionale EE 'G. Panico' Tricase (LE)	474965,79	2581	2083	931
160062	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	24816,02	247	12	73
160168	Ospedale Castellaneta Castellaneta (TA)	73903,08	537	268	215
160146	Centro Medico Riabilitazione Ics Maugeri Ginosa (TA)	12664,25	99	0	94
160074	Ospedale Manduria 'Giannuzzi' Manduria (TA)	52257,42	676	164	248
160141	Casa Di Cura Villa Bianca SRL - Martina Franca (TA)	65,58	0	0	0
160075	Ospedale Civile Martina Franca (TA)	115830,21	971	384	400
160111	Casa Di Cura Bernardini Taranto (TA)	29805,28	154	460	412
160112	Casa Di Cura D'Amore SRL Taranto (TA)	17041,74	8	438	157
160114	Casa Di Cura San Camillo Taranto (TA)	13939,63	227	214	325
160115	Casa Di Cura Santa Rita SRL Taranto (TA)	23,74	169	0	39
160116	Casa Di Cura Villa Verde SRL Taranto (TA)	70016,76	1712	591	289
160149	Fondazione Cittadella Della Carita` Taranto (TA)	4890,44	2	0	0

160172	Presidio Ospedaliero centrale Taranto (TA)	276536,88	1783	1431	1053

3.5.2. RED BOX: NATIONAL HEALTH SERVICE DATA

Input: Two tables on private and public health facilities.
Operations Executed:

- Filtering rows and columns for the Apulia region with Column Filter node.
- Concatenation of the filtered tables with Concatenate node.

Output: A table [59x16]. Rows: Name of 59 Apulian Health Facilities. Columns: Facility code, Facility name, Predicted bed capacity, Predicted departments, Used bed capacity, Used departments, Total staff, Male doctors, Female doctors, Male nurses, Female nurses, Hospitalizations, ASL name, Hospital level, Inpatient and Available days.

Table 3 referred to as the employs a color-coding system from red to green to represent the values in each column. In this system, red indicates low values while green represents high values. Each column variable has a specific meaning: the "Facility Code" is a unique identifier for each healthcare facility, and the "Facility Name" indicates the actual name of the facility. The "Predicted Bed Capacity" and "Predicted Departments" represent the forecasted number of beds and departments, respectively, while "Used Bed Capacity" and "Used Departments" show the actual usage. "Total Staff" indicates the total number of staff members, with further breakdowns into "Male Doctors," "Female Doctors," "Male Nurses," and "Female Nurses." "Hospitalizations" denotes the total number of hospital admissions, "ASL Name" identifies the associated health service area, "Hospital Level" categorizes the level of care provided, and "Inpatient Days" and "Available Days" capture the total inpatient days and the availability of the facility.

The color-coded matrix in Table 3 is particularly useful because it immediately highlights the differences in data among various healthcare facilities. This visual approach allows for the quick identification of anomalies and trends that might not be readily apparent from numerical data alone. Decision makers can leverage this representation to pinpoint areas needing improvement or additional resources. Moreover, the color visualization aids in understanding the relative performance of facilities, making it easier to compare and analyze the overall data. Ultimately, the matrix transforms a vast set of complex data into a user-friendly analytical tool, enhancing the ability to make informed decisions.

Table 3. Color-Coded Matrix of the Red Box Output Dataset.

ASL	Hospital Level	Facility Code	Facility Name	Predicted bed capacity	Used bed capacity	Predicted departments	Used departments	Male nurses	Female nurses	Male doctors	Female doctors	Total staff	Hospitalizations	Inpatient days	Available days
ASL BA	FIRST LEVEL	160078	Ospedale Regionale EE 'Miulli' Acquaviva Delle Fonti (BA)	659	544	31	31	212	356	152	98	1091	20075	140287	198887
ASL BA	FIRST LEVEL	160157	Ospedale Della Murgia - Perinei Altamura (BA)	238	154	20	19	68	209	79	71	652	4586	32889	55812
ASL BA	PRIVATE NURSING HOMES	160140	Casa Di Cura Anthea Bari (BA)	107	107	8	8	55	115	68	22	357	3098	19264	34009
ASL BA	FIRST LEVEL	160147	Casa Di Cura C.B.H. Mater Dei Hospital Bari (BA)	479	461	25	23	85	195	150	45	770	10114	71635	154026
ASL BA	PRIVATE NURSING HOMES	160087	Casa Di Cura Santa Maria Bari (BA)	175	175	13	13	40	95	18	10	257	5289	25117	59568
ASL BA	SECOND LEVEL	160907	Consorziale Policlinico Bari Bari (BA)	1355	884	79	77	480	1158	477	395	4050	29673	245097	321831
ASL BA	IRCCS	160906	Ics Maugeri SPA Societa' Benefit Bari (BA)	226	230	9	7	28	71	23	16	261	2681	75130	84180
ASL BA	IRCCS	160901	Istituto Tumori Giovanni Paolo II Bari (BA)	118	95	11	11	60	163	74	66	569	3344	24614	34587
ASL BA	FIRST LEVEL	160169	Ospedale Di Venere Bari (BA)	292	241	20	20	168	349	137	143	1223	10223	76955	87809
ASL BA	FIRST LEVEL	160158	Ospedale San Paolo Bari (BA)	610	291	39	35	225	563	178	196	1783	11372	72528	106412
ASL BA	IRCCS	160902	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	99	81	5	5	32	91	41	15	319	2527	21487	29463
ASL BA	PRIVATE NURSING HOMES	160098	Casa Di Cura - Villa Lucia Hospital Conversano (BA)	188	180	9	8	20	40	37	17	145	2712	13067	29251
ASL BA	BASE LEVEL	160159	Ospedale Monopoli Monopoli (BA)	152	120	10	10	50	177	59	67	452	4464	23183	43583
ASL BA	PRIVATE NURSING HOMES	160100	Casa Di Cura ' Monte Imperatore' - Noci (BA)	123	95	3	2	9	19	11	6	103	731	21238	34770
ASL BA	BASE LEVEL	160160	Ospedale Putignano Putignano (BA)	233	90	13	12	54	115	47	32	336	2142	21157	32785
ASL BR	PRIVATE NURSING HOMES	160101	Casa Di Cura 'Salus' Brindisi (BR)	60	60	5	5	10	28	45	7	137	1426	7029	21960
ASL BR	IRCCS	160151	IRCCS 'E.Medea' - Brindisi (BR)	30	28	1	1	2	12	8	14	74	295	5681	10156
ASL BR	SECOND LEVEL	160170	Ospedale Perrino Brindisi (BR)	775	609	42	42	226	624	193	173	1619	15238	154479	222312
ASL BR	FIRST LEVEL	160162	Ospedale Francavilla Fontana Francavilla Fontana (BR)	122	105	7	7	55	163	55	39	451	4794	26326	38277
ASL BR	BASE LEVEL	160161	Ospedale Ostuni Ostuni (BR)	160	60	8	8	0	0	30	19	272	2280	14390	21776
ASL BT	FIRST LEVEL	160174	Ospedale Andria Andria (BT)	187	152	17	17	166	279	122	82	859	7001	44784	54775
ASL BT	FIRST LEVEL	160177	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	221	191	16	16	146	211	110	99	744	8264	62070	69659
ASL BT	BASE LEVEL	160178	Ospedale Bisceglie Bisceglie (BT)	191	71	15	12	85	152	61	46	443	2596	21224	25710
ASL BT	PRIVATE NURSING HOMES	160180	Ospedale Opera Don Uva Bisceglie (BT)	100	100	4	4	29	76	15	4	520	1141	26853	35075
ASL FG	PRIVATE NURSING HOMES	160105	Casa Di Cura Leonardo De Luca Castelnuovo Della Daunia (FG)	51	51	3	3	1	4	13	4	68	417	6729	18666
ASL FG	FIRST LEVEL	160047	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	155	139	15	14	95	137	62	39	454	4224	22252	50691
ASL FG	PRIVATE NURSING HOMES	160102	Casa Di Cura Prof. Brodetti Foggia (FG)	53	53	4	4	4	28	30	5	89	2389	11120	14152
ASL FG	PRIVATE NURSING HOMES	160125	Casa Di Cura Universo Salute - Don Uva Foggia (FG)	80	80	4	4	19	39	5	10	360	879	14308	28670
ASL FG	PRIVATE NURSING HOMES	160181	Case Cura Riunite Villa Serena-S. Francesco Foggia (FG)	78	78	5	5	10	13	30	6	121	3673	12074	27389
ASL FG	SECOND LEVEL	160910	Ospedali Riuniti Di Foggia Foggia (FG)	856	738	56	52	322	790	323	226	2676	20628	167966	270167
ASL FG	PRIVATE NURSING HOMES	160102	Casa Di Cura 'S.Michele' Gest. Brodetti Manfredonia (FG)	31	31	1	1	5	6	8	4	37	761	5088	11346
ASL FG	BASE LEVEL	160164	Ospedale Manfredonia Manfredonia (FG)	123	104	9	8	52	87	32	18	319	1817	14953	37942
ASL FG	FIRST LEVEL	160905	Ospedale Casa Sollievo Della Sofferenza San Giovanni Rotondo (FG)	990	628	41	40	399	837	330	154	2738	21993	172164	228253
ASL FG	FIRST LEVEL	160163	Ospedale San Severo - Teresa Masselli San Severo (FG)	210	155	18	14	85	180	61	33	550	4528	28931	56394
ASL LE	PRIVATE NURSING HOMES	160152	Casa Di Cura Riabilitativa Euroitalia - Casarano (LE)	66	66	1	1	12	10	11	6	96	569	18046	24156
ASL LE	BASE LEVEL	160167	Ospedale Casarano Casarano (LE)	162	168	14	12	78	153	63	34	490	4672	35040	61488
ASL LE	BASE LEVEL	160165	Ospedale Copertino Copertino (LE)	138	125	7	7	56	131	45	25	392	2715	21330	45628
ASL LE	PRIVATE NURSING HOMES	160110	Casa Di Cura San Francesco Galatina (LE)	60	60	6	6	3	15	50	5	111	1585	7448	21960
ASL LE	FIRST LEVEL	160063	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	182	182	12	12	62	196	77	31	552	4440	32691	66612

ASL LE	PRIVATE NURSING HOMES	160107	Casa Di Cura 'Prof. Petrucciani' SRL Lecce (LE)	73	71	7	7	10	21	49	14	150	1695	10248	25986
ASL LE	PRIVATE NURSING HOMES	160150	Casa Di Cura Citta' Di Lecce Lecce (LE)	118	118	10	10	38	74	53	12	262	3592	19447	43188
ASL LE	PRIVATE NURSING HOMES	160108	Casa Di Cura Villa Bianca Lecce (LE)	28	28	1	1	9	12	34	4	87	825	2429	10248
ASL LE	PRIVATE NURSING HOMES	160109	Casa Di Cura Villa Verde - Lecce (LE)	46	46	3	3	15	16	20	3	103	849	16235	16836
ASL LE	SECOND LEVEL	160171	Ospedale Lecce 'V. Fazzi' Lecce (LE)	747	647	39	35	222	609	261	220	1834	19092	162279	237046
ASL LE	FIRST LEVEL	160166	Ospedale Scorrano Scorrano (LE)	171	149	11	11	67	198	61	47	547	5404	41436	54320
ASL LE	FIRST LEVEL	160080	Ospedale Regionale EE 'G. Panico' Tricase (LE)	379	393	26	26	149	237	112	58	875	14581	86570	143777
ASL LE	BASE LEVEL	160062	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	201	147	14	12	54	187	42	31	475	3161	26725	53679
ASL TA	FIRST LEVEL	160168	Ospedale Castellaneta Castellaneta (TA)	120	95	7	7	55	263	45	49	611	2959	18732	34739
ASL TA	PRIVATE NURSING HOMES	160146	Centro Medico Riabilitazione Ics Maugeri Ginosa (TA)	67	67	2	2	10	23	10	6	91	906	21375	24522
ASL TA	BASE LEVEL	160074	Ospedale Manduria 'Giannuzzi' Manduria (TA)	146	80	10	9	78	218	35	22	476	2263	17780	29158
ASL TA	FIRST LEVEL	160141	Casa Di Cura Villa Bianca SRL - Martina Franca (TA)	64	64	1	1	5	11	15	4	103	575	17697	23424
ASL TA	FIRST LEVEL	160075	Ospedale Civile Martina Franca (TA)	129	119	10	9	56	274	52	39	571	5249	32516	43584
ASL TA	PRIVATE NURSING HOMES	160111	Casa Di Cura Bernardini Taranto (TA)	96	94	7	6	13	46	54	11	187	2702	13779	34404
ASL TA	PRIVATE NURSING HOMES	160112	Casa Di Cura D'Amore SRL Taranto (TA)	40	40	2	2	8	14	43	8	95	1316	3730	14122
ASL TA	PRIVATE NURSING HOMES	160114	Casa Di Cura San Camillo Taranto (TA)	93	93	6	6	8	29	30	3	146	1597	8240	33245
ASL TA	PRIVATE NURSING HOMES	160115	Casa Di Cura Santa Rita SRL Taranto (TA)	80	80	3	3	7	14	9	2	77	642	5047	13115
ASL TA	PRIVATE NURSING HOMES	160116	Casa Di Cura Villa Verde SRL Taranto (TA)	167	164	9	8	46	88	56	15	318	3536	35503	59231
ASL TA	PRIVATE NURSING HOMES	160149	Fondazione Cittadella Della Carita` Taranto (TA)	54	54	2	2	6	18	12	3	82	749	13165	19764
ASL TA	SECOND LEVEL	160172	Presidio Ospedaliero centrale Taranto (TA)	711	489	42	28	270	916	240	212	2217	16451	134991	178911

3.5.3. ORANGE BOX: DEMOGRAPHIC AND HEALTH DATA

Input: A table with demographic data of Apulia residents in 2020.

Operations Executed:

- Column filtering and aggregation by ASL, using Column filter and Joiner nodes.
- Computation of the ratio of population/doctors and residents distributed per doctor with Math Formula node.

Output: A table [59x5]. Rows: Name of 59 Apulian Health Facilities. Columns: Facility code, Facility name, Residents per doctor, ASL, Network.

Table 4 uses a color-coding system ranging from red to green to represent the values in each column. In this system, red indicates low values while green represents high values. Each column variable has a specific meaning: the "Facility Code" is a unique identifier for each healthcare facility, and the "Facility Name" indicates the actual name of the facility. "Residents per Doctor" represents the number of residents assigned to each doctor, "ASL" identifies the associated health service area, and "Network" indicates the healthcare network to which the facility belongs. The color-coded matrix in Table 4 offers numerous advantages, allowing immediate visualization of differences in data between healthcare facilities. This visual approach facilitates the rapid identification of patterns and outliers that might not be easily detectable through simple numerical data analysis. It is an essential tool for decision-makers, as it clearly highlights areas that require attention or additional resources. Furthermore, the color representation enhances the understanding of the relative performance of various facilities, simplifying the process of comparing and evaluating overall data. Essentially, the matrix transforms a complex set of information into an intuitively visual format, improving analysis capabilities and the speed of data-driven decision-making.

Table 4. Color-Coded Matrix of the Orange Box Output Dataset.

ASL	Facility Code	Network	Facility Name	Residents per Doctor
ASL BA	160078	PRIVATE	Ospedale Regionale EE 'Miulli' Acquaviva Delle Fonti (BA)	111833
ASL BA	160157	PUBLIC	Ospedale Della Murgia - Perinei Altamura (BA)	67100
ASL BA	160140	PRIVATE	Casa Di Cura Anthea Bari (BA)	40260
ASL BA	160147	PRIVATE	Casa Di Cura C.B.H. Mater Dei Hospital Bari (BA)	87229
ASL BA	160087	PRIVATE	Casa Di Cura Santa Maria Bari (BA)	12525
ASL BA	160907	PUBLIC	Consorziale Policlinico Bari Bari (BA)	390072
ASL BA	160906	PRIVATE	Ics Maugeri SPA Societa' Benefit Bari (BA)	17446
ASL BA	160901	PUBLIC	Istituto Tumori Giovanni Paolo II Bari (BA)	62626
ASL BA	160169	PUBLIC	Ospedale Di Venere Bari (BA)	125252
ASL BA	160158	PUBLIC	Ospedale San Paolo Bari (BA)	167301
ASL BA	160902	PUBLIC	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	25050
ASL BA	160098	PRIVATE	Casa Di Cura - Villa Lucia Hospital Conversano (BA)	24156
ASL BA	160159	PUBLIC	Ospedale Monopoli Monopoli (BA)	56364
ASL BA	160100	PRIVATE	Casa Di Cura ' Monte Imperatore' - Noci (BA)	7605
ASL BA	160160	PUBLIC	Ospedale Putignano Putignano (BA)	35339
ASL BR	160101	PRIVATE	Casa Di Cura 'Salus' Brindisi (BR)	34067
ASL BR	160151	PRIVATE	IRCCS 'E.Medea' - Brindisi (BR)	14413
ASL BR	160170	PUBLIC	Ospedale Perrino Brindisi (BR)	239781
ASL BR	160162	PUBLIC	Ospedale Francavilla Fontana Francavilla Fontana (BR)	61583
ASL BR	160161	PUBLIC	Ospedale Ostuni Ostuni (BR)	32102
ASL BT	160174	PUBLIC	Ospedale Andria Andria (BT)	144235
ASL BT	160177	PUBLIC	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	147770
ASL BT	160178	PUBLIC	Ospedale Bisceglie Bisceglie (BT)	75653
ASL BT	160180	PRIVATE	Ospedale Opera Don Uva Bisceglie (BT)	13434
ASL FG	160105	PRIVATE	Casa Di Cura Leonardo De Luca Castelnuovo Della Daunia (FG)	7352
ASL FG	160047	PUBLIC	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	43677

ASL FG	160102	PRIVATE	Casa Di Cura Prof. Brodetti Foggia (FG)	15136
ASL FG	160125	PRIVATE	Casa Di Cura Universo Salute - Don Uva Foggia (FG)	6487
ASL FG	160181	PRIVATE	Case Cura Riunite Villa Serena-S. Francesco Foggia (FG)	15568
ASL FG	160910	PUBLIC	Ospedali Riuniti Di Foggia Foggia (FG)	237412
ASL FG	160102	PRIVATE	Casa Di Cura 'S.Michele' Gest. Brodetti Manfredonia (FG)	5189
ASL FG	160164	PUBLIC	Ospedale Manfredonia Manfredonia (FG)	21622
ASL FG	160905	PRIVATE	Ospedale Casa Sollievo Della Sofferenza San Giovanni Rotondo (FG)	209303
ASL FG	160163	PUBLIC	Ospedale San Severo - Teresa Masselli San Severo (FG)	40650
ASL LE	160152	PRIVATE	Casa Di Cura Riabilitativa Euroitalia - Casarano (LE)	9646
ASL LE	160167	PUBLIC	Ospedale Casarano Casarano (LE)	55040
ASL LE	160165	PUBLIC	Ospedale Copertino Copertino (LE)	39719
ASL LE	160110	PRIVATE	Casa Di Cura San Francesco Galatina (LE)	31208
ASL LE	160063	PUBLIC	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	61281
ASL LE	160107	PRIVATE	Casa Di Cura 'Prof. Petrucciani' SRL Lecce (LE)	35747
ASL LE	160150	PRIVATE	Casa Di Cura Citta' Di Lecce Lecce (LE)	36882
ASL LE	160108	PRIVATE	Casa Di Cura Villa Bianca Lecce (LE)	21562
ASL LE	160109	PRIVATE	Casa Di Cura Villa Verde - Lecce (LE)	13051
ASL LE	160171	PUBLIC	Ospedale Lecce 'V. Fazzi' Lecce (LE)	272929
ASL LE	160166	PUBLIC	Ospedale Scorrano Scorrano (LE)	61281
ASL LE	160080	PRIVATE	Ospedale Regionale EE 'G. Panico' Tricase (LE)	96461
ASL LE	160062	PUBLIC	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	41422
ASL TA	160168	PUBLIC	Ospedale Castellaneta Castellaneta (TA)	54179
ASL TA	160146	PRIVATE	Centro Medico Riabilitazione Ics Maugeri Ginosa (TA)	9222
ASL TA	160074	PUBLIC	Ospedale Manduria 'Giannuzzi' Manduria (TA)	32853
ASL TA	160141	PRIVATE	Casa Di Cura Villa Bianca SRL - Martina Franca (TA)	10951
ASL TA	160075	PUBLIC	Ospedale Civile Martina Franca (TA)	52449
ASL TA	160111	PRIVATE	Casa Di Cura Bernardini Taranto (TA)	37464
ASL TA	160112	PRIVATE	Casa Di Cura D'Amore SRL Taranto (TA)	29395
ASL TA	160114	PRIVATE	Casa Di Cura San Camillo Taranto (TA)	19020
ASL TA	160115	PRIVATE	Casa Di Cura Santa Rita SRL Taranto (TA)	6340
ASL TA	160116	PRIVATE	Casa Di Cura Villa Verde SRL Taranto (TA)	40922
ASL TA	160149	PRIVATE	Fondazione Cittadella Della Carita` Taranto (TA)	8646
ASL TA	160172	PUBLIC	Presidio Ospedaliero centrale Taranto (TA)	260518

3.5.4. BROWN BOX: INPUT/OUTPUT VARIABLES

Input: Two output tables from the yellow, red, and orange boxes.
Operations Executed:

- Tables union with Joiner node.

Output: A table [59 x18]. Rows: Name of 59 Apulian Health Facilities. Columns: The columns of the output tables from the yellow, red, and orange boxes.

3.5.5. DARK RED BOX: DATASET

Operations Executed:

- Rows and columns sorting with Sorter and Column Resorter nodes.
- Column filtering and renaming with column Filter and column Rename nodes.
- Dataset saving in Excel format with Excel Writer node.

Output: Dataset in Excel format with Rows: Name of 59 Apulian Health Facilities. Columns: ASL, Network, Hospital level, Facility Name, 17 input/output variables (renamed as described in Table 1). The final dataset, resulting from the preprocessing operations in the Brown Box and Dark Red Box, is presented in Table 5. This table uses a color-coding system where red indicates low values and green represents high values. Each column variable contributes to the understanding of the performance of Apulian health facilities. The color-coded matrix allows for quick identification of trends and anomalies, aiding decision-makers in effectively allocating resources and interventions. This visual approach enhances data comprehension and facilitates easier comparison between different facilities.

Table 5. Color-Coded Matrix of the Final Dataset

ASL	NETWORK	HOSPITAL LEVEL	FACILITY CODE	FACILITY NAME	INP1_BP	INP2_BU	INP3_DP	INP4_DU	INP5_MN	INP6_FN	INP7_MP	INP8_FP	INP9_HS	INP10_RESP	OUT1_HOS	OUT2_MOB	OUT3_DEA	OUT4_INT	OUT5_REA	OUT6_INP	OUT7_AVA
ASL BA	PRIVATE	FIRST LEVEL	160078	Ospedale Regionale EE 'Miulli' Acquaviva Delle Fonti (BA)	659	544	31	31	212	356	152	98	1091	111833	20075	346856,74	2704	2914	1543	140287	198887
ASL BA	PUBLIC	FIRST LEVEL	160157	Ospedale Della Murgia - Perinei Altamura (BA)	238	154	20	19	68	209	79	71	652	67100	4586	64934,02	988	199	341	32889	55812
ASL BA	PRIVATE	PRIVATE NURSING HOMES	160140	Casa Di Cura Anthea Bari (BA)	107	107	8	8	55	115	68	22	357	40260	3098	72932,11	1333	810	309	19264	34009
ASL BA	PRIVATE	FIRST LEVEL	160147	Casa Di Cura C.B.H. Mater Dei Hospital Bari (BA)	479	461	25	23	85	195	150	45	770	87229	10114	248510,43	2341	1620	457	71635	154026
ASL BA	PRIVATE	PRIVATE NURSING HOMES	160087	Casa Di Cura Santa Maria Bari (BA)	175	175	13	13	40	95	18	10	257	12525	5289	130806,77	1472	1331	433	25117	59568
ASL BA	PUBLIC	SECOND LEVEL	160907	Consorziale Policlinico Bari Bari (BA)	1355	884	79	77	480	1158	477	395	4050	390072	29673	314578,53	4826	2928	1358	245097	321831
ASL BA	PRIVATE	IRCCS	160906	Ics Maugeri SPA Societa' Benefit Bari (BA)	226	230	9	7	28	71	23	16	261	17446	2681	35292,39	378	0	258	75130	84180
ASL BA	PUBLIC	IRCCS	160901	Istituto Tumori Giovanni Paolo II Bari (BA)	118	95	11	11	60	163	74	66	569	62626	3344	31141,29	764	954	49	24614	34587
ASL BA	PUBLIC	FIRST LEVEL	160169	Ospedale Di Venere Bari (BA)	292	241	20	20	168	349	137	143	1223	125252	10223	181312,93	2036	860	545	76955	87809
ASL BA	PUBLIC	FIRST LEVEL	160158	Ospedale San Paolo Bari (BA)	610	291	39	35	225	563	178	196	1783	167301	11372	196367,61	2283	1019	762	72528	106412
ASL BA	PUBLIC	IRCCS	160902	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	99	81	5	5	32	91	41	15	319	25050	2527	56820,26	506	395	9	21487	29463
ASL BA	PRIVATE	PRIVATE NURSING HOMES	160098	Casa Di Cura - Villa Lucia Hospital Conversano (BA)	188	180	9	8	20	40	37	17	145	24156	2712	54380,73	73	890	447	13067	29251
ASL BA	PUBLIC	BASE LEVEL	160159	Ospedale Monopoli Monopoli (BA)	152	120	10	10	50	177	59	67	452	56364	4464	60873,78	709	100	241	23183	43583
ASL BA	PRIVATE	PRIVATE NURSING HOMES	160100	Casa Di Cura ' Monte Imperatore' - Noci (BA)	123	95	3	2	9	19	11	6	103	7605	731	2673,45	46	0	34	21238	34770
ASL BA	PUBLIC	BASE LEVEL	160160	Ospedale Putignano Putignano (BA)	233	90	13	12	54	115	47	32	336	35339	2142	25905,12	726	34	250	21157	32785
ASL BR	PRIVATE	PRIVATE NURSING HOMES	160101	Casa Di Cura 'Salus' Brindisi (BR)	60	60	5	5	10	28	45	7	137	34067	1426	20698,19	123	242	204	7029	21960
ASL BR	PRIVATE	IRCCS	160151	IRCCS 'E.Medea' - Brindisi (BR)	30	28	1	1	2	12	8	14	74	14413	295	0,00	0	0	0	5681	10156
ASL BR	PUBLIC	SECOND LEVEL	160170	Ospedale Perrino Brindisi (BR)	775	609	42	42	226	624	193	173	1619	239781	15238	208664,34	2567	1314	774	154479	222312
ASL BR	PUBLIC	FIRST LEVEL	160162	Ospedale Francavilla Fontana Francavilla Fontana (BR)	122	105	7	7	55	163	55	39	451	61583	4794	87457,47	466	161	170	26326	38277
ASL BR	PUBLIC	BASE LEVEL	160161	Ospedale Ostuni Ostuni (BR)	160	60	8	8	0	0	30	19	272	32102	2280	36934,17	527	88	439	14390	21776
ASL BT	PUBLIC	FIRST LEVEL	160174	Ospedale Andria Andria (BT)	187	152	17	17	166	279	122	82	859	144235	7001	64846,20	1749	925	397	44784	54775
ASL BT	PUBLIC	FIRST LEVEL	160177	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	221	191	16	16	146	211	110	99	744	147770	8264	49933,13	1270	387	464	62070	69659
ASL BT	PUBLIC	BASE LEVEL	160178	Ospedale Bisceglie Bisceglie (BT)	191	71	15	12	85	152	61	46	443	75653	2596	16644,86	458	77	170	21224	25710
ASL BT	PRIVATE	PRIVATE NURSING HOMES	160180	Ospedale Opera Don Uva Bisceglie (BT)	100	100	4	4	29	76	15	4	520	13434	1141	3089,58	194	0	113	26853	35075
ASL FG	PRIVATE	PRIVATE NURSING HOMES	160105	Casa Di Cura Leonardo De Luca Castelnuovo Della Daunia (FG)	51	51	3	3	1	4	13	4	68	7352	417	2439,68	38	0	36	6729	18666
ASL FG	PUBLIC	FIRST LEVEL	160047	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	155	139	15	14	95	137	62	39	454	43677	4224	66195,18	336	90	186	22252	50691
ASL FG	PRIVATE	PRIVATE NURSING HOMES	160102	Casa Di Cura Prof. Brodetti Foggia (FG)	53	53	4	4	4	28	30	5	89	15136	2389	14439,10	68	67	36	11120	14152
ASL FG	PRIVATE	PRIVATE NURSING HOMES	160125	Casa Di Cura Universo Salute - Don Uva Foggia (FG)	80	80	4	4	19	39	5	10	360	6487	879	7047,20	27	0	124	14308	28670
ASL FG	PRIVATE	PRIVATE NURSING HOMES	160181	Case Cura Riunite Villa Serena-S. Francesco Foggia (FG)	78	78	5	5	10	13	30	6	121	15568	3673	52795,15	2	246	410	12074	27389
ASL FG	PUBLIC	SECOND LEVEL	160910	Ospedali Riuniti Di Foggia Foggia (FG)	856	738	56	52	322	790	323	226	2676	237412	20628	423560,38	1438	1687	998	167966	270167
ASL FG	PRIVATE	PRIVATE NURSING HOMES	160102	Casa Di Cura 'S.Michele' Gest. Brodetti Manfredonia (FG)	31	31	1	1	5	6	8	4	37	5189	761	3128,35	79	3	110	5088	11346
ASL FG	PUBLIC	BASE LEVEL	160164	Ospedale Manfredonia Manfredonia (FG)	123	104	9	8	52	87	32	18	319	21622	1817	15887,80	90	27	103	14953	37942
ASL FG	PRIVATE	FIRST LEVEL	160905	Ospedale Casa Sollievo Della Sofferenza San Giovanni Rotondo (FG)	990	628	41	40	399	837	330	154	2738	209303	21993	94278,58	1577	2033	724	172164	228253
ASL FG	PUBLIC	FIRST LEVEL	160163	Ospedale San Severo - Teresa Masselli San Severo (FG)	210	155	18	14	85	180	61	33	550	40650	4528	96295,50	401	265	292	28931	56394
ASL LE	PRIVATE	PRIVATE NURSING HOMES	160152	Casa Di Cura Riabilitativa Euroitalia - Casarano (LE)	66	66	1	1	12	10	11	6	96	9646	569	180,18	1	0	1	18046	24156
ASL LE	PUBLIC	BASE LEVEL	160167	Ospedale Casarano Casarano (LE)	162	168	14	12	78	153	63	34	490	55040	4672	66074,08	743	171	453	35040	61488
ASL LE	PUBLIC	BASE LEVEL	160165	Ospedale Copertino Copertino (LE)	138	125	7	7	56	131	45	25	392	39719	2715	24130,71	671	180	355	21330	45628
ASL LE	PRIVATE	PRIVATE NURSING HOMES	160110	Casa Di Cura San Francesco Galatina (LE)	60	60	6	6	3	15	50	5	111	31208	1585	11501,62	84	83	73	7448	21960

ASL LE	PUBLIC	FIRST LEVEL	160063	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	182	182	12	12	62	196	77	31	552	61281	4440	119352,60	985	237	351	32691	66612
ASL LE	PRIVATE	PRIVATE NURSING HOMES	160107	Casa Di Cura 'Prof. Petrucciani' SRL Lecce (LE)	73	71	7	7	10	21	49	14	150	35747	1695	37515,01	90	332	54	10248	25986
ASL LE	PRIVATE	PRIVATE NURSING HOMES	160150	Casa Di Cura Citta' Di Lecce Lecce (LE)	118	118	10	10	38	74	53	12	262	36882	3592	119174,63	1645	1500	426	19447	43188
ASL LE	PRIVATE	PRIVATE NURSING HOMES	160108	Casa Di Cura Villa Bianca Lecce (LE)	28	28	1	1	9	12	34	4	87	21562	825	20237,78	13	340	271	2429	10248
ASL LE	PRIVATE	PRIVATE NURSING HOMES	160109	Casa Di Cura Villa Verde - Lecce (LE)	46	46	3	3	15	16	20	3	103	13051	849	0,00	0	0	0	16235	16836
ASL LE	PUBLIC	SECOND LEVEL	160171	Ospedale Lecce 'V. Fazzi' Lecce (LE)	747	647	39	35	222	609	261	220	1834	272929	19092	300858,92	4336	1627	1100	162279	237046
ASL LE	PUBLIC	FIRST LEVEL	160166	Ospedale Scorrano Scorrano (LE)	171	149	11	11	67	198	61	47	547	61281	5404	134705,94	1191	234	528	41436	54320
ASL LE	PRIVATE	FIRST LEVEL	160080	Ospedale Regionale EE 'G. Panico' Tricase (LE)	379	393	26	26	149	237	112	58	875	96461	14581	474965,79	2581	2083	931	86570	143777
ASL LE	PUBLIC	BASE LEVEL	160062	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	201	147	14	12	54	187	42	31	475	41422	3161	24816,02	247	12	73	26725	53679
ASL TA	PUBLIC	FIRST LEVEL	160168	Ospedale Castellaneta Castellaneta (TA)	120	95	7	7	55	263	45	49	611	54179	2959	73903,08	537	268	215	18732	34739
ASL TA	PRIVATE	PRIVATE NURSING HOMES	160146	Centro Medico Riabilitazione Ics Maugeri Ginosa (TA)	67	67	2	2	10	23	10	6	91	9222	906	12664,25	99	0	94	21375	24522
ASL TA	PUBLIC	BASE LEVEL	160074	Ospedale Manduria 'Giannuzzi' Manduria (TA)	146	80	10	9	78	218	35	22	476	32853	2263	52257,42	676	164	248	17780	29158
ASL TA	PRIVATE	FIRST LEVEL	160141	Casa Di Cura Villa Bianca SRL - Martina Franca (TA)	64	64	1	1	5	11	15	4	103	10951	575	65,58	0	0	0	17697	23424
ASL TA	PUBLIC	FIRST LEVEL	160075	Ospedale Civile Martina Franca (TA)	129	119	10	9	56	274	52	39	571	52449	5249	115830,21	971	384	400	32516	43584
ASL TA	PRIVATE	PRIVATE NURSING HOMES	160111	Casa Di Cura Bernardini Taranto (TA)	96	94	7	6	13	46	54	11	187	37464	2702	29805,28	154	460	412	13779	34404
ASL TA	PRIVATE	PRIVATE NURSING HOMES	160112	Casa Di Cura D'Amore SRL Taranto (TA)	40	40	2	2	8	14	43	8	95	29395	1316	17041,74	8	438	157	3730	14122
ASL TA	PRIVATE	PRIVATE NURSING HOMES	160114	Casa Di Cura San Camillo Taranto (TA)	93	93	6	6	8	29	30	3	146	19020	1597	13939,63	227	214	325	8240	33245
ASL TA	PRIVATE	PRIVATE NURSING HOMES	160115	Casa Di Cura Santa Rita SRL Taranto (TA)	80	80	3	3	7	14	9	2	77	6340	642	23,74	169	0	39	5047	13115
ASL TA	PRIVATE	PRIVATE NURSING HOMES	160116	Casa Di Cura Villa Verde SRL Taranto (TA)	167	164	9	8	46	88	56	15	318	40922	3536	70016,76	1712	591	289	35503	59231
ASL TA	PRIVATE	PRIVATE NURSING HOMES	160149	Fondazione Cittadella Della Carita` Taranto (TA)	54	54	2	2	6	18	12	3	82	8646	749	4890,44	2	0	0	13165	19764
ASL TA	PUBLIC	SECOND LEVEL	160172	Presidio Ospedaliero centrale Taranto (TA)	711	489	42	28	270	916	240	212	2217	260518	16451	276536,88	1783	1431	1053	134991	178911

3.6 WIDGETS AND CPDA METHODOLOGY: THE WORKFLOW APPROACH IN ORANGE SOFTWARE

Within the workflow of the Orange software, the use of various widgets is crucial, especially in implementing the CPDA methodology and in machine learning techniques. Widgets, seen as modular and interactive components, make data analysis dynamic and responsive, facilitating the adoption of machine learning algorithms. Users can visually construct and customize the flow, ensuring that both the requirements of the CPDA methodology and machine learning specifics are optimally met. Each widget, from data input to visualization, plays a pivotal role in integrating and executing the CPDA steps, ensuring a smooth and consistent analysis within the Orange environment.

The workflow in Orange software, as depicted in Figure 2, is divided into sections for easy comprehension:

1. Knime Dataset: Using the File widget, we loaded the dataset output from Knime. This was then linked to the DataTable widget for tabular visualization and connected to the Select Column widget to identify the 15 input/output variables as features to analyze.
2. Cluster Analysis: The dataset was linked to the Distances widget to specify the metric and method for the dataset columns. This was then connected to the Hierarchical Clustering widget to perform hierarchical cluster analysis. The dendrogram was visualized, cut at 80%, and the Ward method was indicated.
3. Standardization: The dataset was linked to the Continuize widget for variable standardization.
4. Principal Component Analysis: Two separate Column Filter widgets were used to divide the input variables belonging to the first cluster and the output variables belonging to the second cluster. Both were connected to two separate PCA widgets to conduct two distinct principal component analyses. These, in turn, were linked to Edit Domain widgets to rename the identified components and subsequently merged using the MergeData widget to form a single table.
5. Positive Shift: Initially, the Features Statistic widget was used to identify the minimum value between the two components. Subsequently, the Feature Constructor widget was employed to add this value to the components, making them positive. The dataset with the positive components was then saved using the SaveData widget.
6. Data Envelopment Analysis (DEA): The previously saved dataset was loaded into the PIM DEA software to compute the DEA Pure Technical Efficiency (PTE) scores with variable return to scale (VRS) output oriented. Once these scores were obtained, they were re-imported into Orange using the File widget.
7. Normalization: Using the Feature Constructor widget, the logarithm of the efficiency scores was calculated to achieve a normalized distribution. This distribution was then graphically displayed using the Distribution widget.
8. ANOVA Analysis: The efficiency scores were linked to the ANOVA widget for a variance analysis, considering hospital levels as groups. The results were graphically displayed using the Scatter Plot widget.

3.7 CPDA WORKFLOW IN HOSPITAL EFFICIENCY EVALUATION: ADVANTAGES, ALTERNATIVES, AND IMPACTS

Alternatives to the CPDA Workflow

Several alternatives exist to the CPDA workflow. Common methodologies include the use of techniques such as multivariate regression, discriminant analysis, neural networks, support vector machines, and time series analysis. These techniques can be used individually or in combination, depending on the nature of the data and research objectives.

Advantages of the CPDA Workflow

The CPDA framework offers numerous advantages in the context of this study:

- **Comprehensiveness:** The combination of clustering techniques, principal component analysis, DEA, and ANOVA provides a holistic analysis, allowing exploration of both relationships among variables and relative efficiencies among decision-making units (like hospitals).
- **Health Sector Specificity:** The workflow has been designed considering the specific challenges and characteristics of the healthcare sector, making it particularly suitable for analyzing hospital efficiency.
- **Flexibility:** The CPDA workflow can be easily adapted to include or exclude variables, offering some flexibility in designing the analysis.
- **Integration of Diverse Data Sources:** The CPDA methodology is designed to integrate data from various sources, ensuring a comprehensive view of hospital efficiency.

Impact on Results

The adoption of the CPDA workflow will influence the results in various ways:

- **Relevance:** The results will be closely tied to the selected variables and analysis techniques, making them particularly relevant for hospital efficiency evaluation.
- **Depth:** The combination of various techniques will provide a deeper insight into the relationships among variables and relative efficiencies.
- **Validity:** The use of established techniques like DEA and ANOVA ensures the validity of the results.

3.8 ANALYSIS OF HOSPITAL NETWORKS AND LEVELS IN APULIA

In the analysis of hospital structures in the Apulia region, a contingency table will be used to examine the relationship between two key variables: the nature of the hospital network (private or public) and the hospital level (secondary, primary, base, IRCCS, private nursing home) (table 6). The row variable will represent the nature of the hospital network, while the column variable will be dedicated to the hospital level. This table will provide us with an overall and detailed view of the composition of the Apulia hospital network in terms of private or public affiliation and hospital level. Through the analysis of data obtained from the contingency table, we will gain significant insights into the regional hospital structure and identify any disparities or areas for improvement. This decision support will be invaluable in optimizing resource allocation and promoting a more equitable and sustainable distribution of hospital care in the Apulia region.

In accordance with the Ministry of Health:

Hospitals of First Level: They provide basic services such as emergency care, diagnostics, regular hospitalization, and outpatient services. They are usually present in various regions of Italy and provide primary level care to the local community.

Hospitals of Second Level: They are more specialized than first-level hospitals. They offer more complex services such as specialized surgery, intensive care, and hemodynamics services. They are present in numerous regions of Italy and serve as reference points for the provision of advanced care.

Basic Hospitals: They primarily perform primary care functions. They provide basic care, outpatient services, and primary level diagnostics. They are present in various regions of Italy and serve as a link between primary care and more specialized hospital facilities.
Institutes of Scientific Research and Care (IRCCS): They are specialized hospital facilities dedicated to scientific research and highly specialized healthcare. They are present in various regions of Italy and offer highly specialized care, playing an important role in medical research and the development of new therapies.
Accredited Private Healthcare Facilities: These are private healthcare facilities that have obtained accreditation from the Italian National Health Service (Servizio Sanitario Nazionale or SSN). They offer care and rehabilitation services to patients in need of medical assistance. They collaborate with the public healthcare system and operate in compliance with the quality and safety requirements established by the SSN.

Table 6. Contingency Table: Relationship between Hospital Network Nature and Hospital Level in the Apulia Region.

		HOSPITAL LEVEL					
NETWORK		BASE LEVEL	FIRST LEVEL	IRCCS	PRIVATE NURSING HOMES	SECOND LEVEL	Total
PRIVATE	Observed	0	5	2	24	0	31
	% of total	0.0 %	8.5 %	3.4 %	40.7 %	0.0 %	52.5 %
PUBLIC	Observed	9	12	2	0	5	28
	% of total	15.3 %	20.3 %	3.4 %	0.0 %	8.5 %	47.5 %
Total	Observed	9	17	4	24	5	59
	% of total	15.3 %	28.8 %	6.8 %	40.7 %	8.5 %	100.0 %

The p-value < 0.001 indicates that the association between the two variables is highly significant. In other words, it is extremely unlikely that the observed association between hospital level and hospital network is due to chance (Table 7).

Table 7. χ^2 Tests for Hospital Distribution.

	Value	df	p
χ^2	40.8	4	< .001
N	59		

Interpreting the results of the contingency table, we can observe the following:
"Hospital levels" are associated with "hospital networks". For example, all 9 base-level hospitals and 12 first-level hospitals belong to the public network, while all 5 first-level hospitals belong to the private network. This association between hospital level and hospital network is highlighted by the significant chi-square test value.
IRCCS (Scientific Research Institutes) are present in both the public and private networks, with 2 hospitals in each network.
Accredited private nursing homes are exclusively present in the private network, with a total of 24 hospitals.
Five second-level hospitals are present in the public network.
In summary, the results indicate significant differences in the distribution of hospitals across different levels depending on the hospital network (public or private). These differences may be influenced by factors such as management, access to resources, and the type of services offered.

The private network represents 52.5% of the analyzed hospitals, while the public network represents 47.5%. It is important to balance the role of the public and private sectors to ensure comprehensive healthcare coverage. Policies that promote collaboration between the two networks can contribute to a more efficient and sustainable healthcare system.

3.9 EXPERIMENTAL ASSESSMENT OF CLUSTERING ALGORITHMS

In the realm of hospital efficiency evaluation, the complexity and heterogeneity of data present a significant challenge. The variables involved can range from financial measures to service quality indicators and may even include demographic or geographic variables. This is why we have selected three distinct clustering algorithms for our CPDA model (Cluster - PCA - DEA - ANOVA), each with its strengths in specific application areas:
DBSCAN: This algorithm is particularly well-suited for identifying clusters of arbitrary shape in data, a feature that can be invaluable when dealing with hospital data that doesn't always follow symmetric distributions or predictable clustered forms. Its ability to handle "noise" and outlier points is also advantageous when dealing with hospital data that may include errors or outliers.
Louvain Clustering: The strength of this algorithm lies in its ability to detect community structures in large networks. In the hospital setting, where complex relationships exist between various departments, services, and units, the ability to identify such communities can be extremely useful for understanding how resources are efficiently allocated and utilized.
Hierarchical Clustering: This algorithm excels at exploring structural relationships within data. In the hospital context, it could reveal hidden hierarchies or relationships between various sectors or units, like the relationships between different types of healthcare services or performance indicators. Its ability to provide a hierarchical output makes it easier for administrators and policymakers to interpret the results.
The selection of these algorithms was not only aimed at establishing a robust methodological framework to tackle the complexity and diversity of hospital data, but also at leveraging the clustering method as an effective filtering tool. Starting from a comprehensive set of available variables, it then focused on identifying those that did not significantly align with relevant patterns or with selections found in existing literature.

3.9.1. EXPERIMENTATION GOAL AND CONSTRAINTS

The primary goal of our experimentation was to identify the most suitable clustering algorithm for the CPDA model. This was done while considering two specific constraints: the formation of at least two clusters and the correct assignment of input and output variables into distinct clusters, in accordance with existing literature in the field of hospital efficiency evaluation.
The rationale behind the need for a minimum of two clusters stems from the methodological requirements of the DEA model, which is an integral component of our CPDA approach. Specifically, having at least one cluster for input variables and another for output variables is crucial for efficiency calculations. This ensures that the model has the necessary information to perform a meaningful and robust efficiency analysis.

Moreover, we chose to experiment with multiple clustering algorithms to explore potential methodological alternatives. Each algorithm has its own unique strengths and limitations, and the use of multiple algorithms allowed us to evaluate how different clustering combinations might interact with the remaining components of the CPDA workflow. This also enabled us to ascertain whether the incorporation of various clustering techniques could lead to refinements or variations in hospital efficiency evaluation.
To this end, we implemented a machine learning-based workflow, applying three distinct clustering algorithms to our preprocessed dataset (as detailed in Section 3.4.1). The first step of our selection process focused on variable filtering. We utilized the results of the clustering to identify and remove variables that did not conform to our predetermined criteria, thereby simplifying the dataset and retaining only the most relevant variables for subsequent analysis.
The second step involved the representation of the identified clusters, to confirm the expectations based on literature. In this phase, we evaluated how effectively the algorithms formed two distinct clusters, ensuring precise assignment of input and output variables. The algorithm selected was the one that best demonstrated this capability, affirming the results from the literature and integrating harmoniously with the CPDA workflow.
The parameters of the algorithm were optimised by determining at least 2 clusters for the different metrics as follows (Table 8):

Table 8. Parameter Optimization for Clustering Algorithms in Two-Step Process.

I STEP	II STEP
DBSCAN	
Distance metric: Cosine Core point neighbors: 4 Neighborhood distance: 1,90 Normalize features **Distance metric: Euclidean** Core point neighbors: 4 Neighborhood distance: 19,35 Normalize features **Distance metric: Manhattan** Core point neighbors: 4 Neighborhood distance: 134,77 Normalize features	**Distance metric: Cosine** Core point neighbors: 4 Neighborhood distance: 0,11 Normalize features **Distance metric: Euclidean** Core point neighbors: 4 Neighborhood distance: 1,74 Normalize features **Distance metric: Manhattan** Core point neighbors: 4 Neighborhood distance: 12,38 Normalize features
LOUVAIN CLUSTERING	
Distance metric: Cosine Apply PCA preprocessing Normalize data PCA Components: 17 K neighbors: 13 Resolution: 0.9	**Distance metric: Cosine** Apply PCA preprocessing Normalize data PCA Components: 15 K neighbors: 13 Resolution: 0.9
Distance metric: Euclidean Apply PCA preprocessing Normalize data PCA Components: 17 K neighbors: 13 Resolution: 0.9	**Distance metric: Euclidean** Apply PCA preprocessing Normalize data PCA Components: 15 K neighbors: 13 Resolution: 0.9
Distance metric: Manhattan Apply PCA preprocessing Normalize data PCA Components: 17 K neighbors: 13 Resolution: 0.9	**Distance metric: Manhattan** Apply PCA preprocessing Normalize data PCA Components: 15 K neighbors: 13 Resolution: 0.9
HIERARCHICAL CLUSTER	
Distance metric: Cosine	**Distance metric: Cosine**

Linkage: Ward Height ratio: 80% Distance between rows	Linkage: Ward Height ratio: 80% Distance between rows
Distance metric: Euclidean Linkage: Ward Height ratio: 80% Distance between rows Normalize features	**Distance metric: Euclidean** Linkage: Ward Height ratio: 80% Distance between rows Normalize features
Distance metric: Manhattan Linkage: Ward Height ratio: 80% Distance between rows Normalize features	**Distance metric: Manhattan** Linkage: Ward Height ratio: 80% Distance between rows Normalize features
Distance metric: Spearman Linkage: Ward Height ratio: 80% Distance between rows	**Distance metric: Spearman** Linkage: Ward Height ratio: 80% Distance between rows

3.9.2. RESULTS EVALUATION

We conducted a series of experiments utilizing various clustering algorithms and metrics with the aim of identifying the most effective combination for our CPDA model. Existing literature on hospital efficiency evaluation recommends the use of 17 variables, comprising 10 input and 7 output variables (Table 1).
Figure 4 illustrates the clustering results for the three algorithms in the first step of our analysis using various metrics. The affiliation matrix in Figure 4 displays the clustering assignments of the 17 variables. Input variables (INP) are highlighted in beige, while output variables (OUT) are in blue. Each column represents the outcomes of a different algorithm and/or distance metric setting, with "C1" or "C2" indicating the assignment of each variable to the respective cluster.

	DBSCAN			LOUVAIN			HIERARCHICAL			
VARIABLES	EUCLIDEAN	COSINE	MANHATTAN	EUCLIDEAN	COSINE	MANHATTAN	EUCLIDEAN	COSINE	MANHATTAN	SPEARMAN
INP1_PLP	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP2_PLU	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP3_RP	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP4_RU	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP5_INFU	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP6_INFD	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP7_MEDU	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP8_MEDD	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP9_PO	C1	C1	C1	C1	C1	C1	C1	C1	C1	C2
INP10_RESMED	C1	C1	C1	C2	C2	C2	C2	C1	C2	C2
OUT1_RIC	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1
OUT2_MOB	C1	C1	C1	C2	C2	C2	C2	C2	C2	C1
OUT3_MORT	C1	C1	C1	C1	C1	C1	C1	C2	C1	C1
OUT4_INT	C1	C1	C1	C1	C1	C1	C1	C2	C1	C1
OUT5_RIA	C1	C1	C1	C1	C1	C1	C1	C2	C1	C1
OUT6_DEG	C1	C1	C1	C2	C2	C2	C2	C1	C2	C2
OUT7_DISP	C1	C1	C1	C2	C2	C2	C2	C1	C2	C2
Percentage of Assignment Error	Unclassifiable			29,41	29,41	29,41	29,41	17,65	29,41	11,76

Figure 4. First Step Cluster Assignment Results.

For DBSCAN with the Euclidean, Manhattan, and Cosine metrics, it was observed that all 17 variables were assigned to a single cluster, not meeting the requirements of the CPDA model, which necessitates a clear separation of input and output variables into distinct clusters. Louvain Clustering and Hierarchical Clustering with Euclidean and Manhattan metrics showed an error rate of 29.41%, while Hierarchical Clustering with the Cosine metric had an error rate of 17.65%. Remarkably, Hierarchical Clustering with the Spearman metric exhibited the lowest error rate at 11.76%, indicating a better alignment with the CPDA model's criteria. To refine our model and enhance its accuracy, we decided to eliminate two variables, OUT6 and OUT7, which had been incorrectly assigned by the selected algorithm. This pruning will allow us to focus on the most significant variables and

ensure that the CPDA model closely captures the actual dynamics of hospital efficiency, as suggested by the reference literature.
Following the conclusive results of the first step of analysis, we proceeded with the second step, focusing on the refined dataset consisting of 15 variables: 10 inputs and 5 outputs. The scatterplots representing the clusters for the three algorithms in the second step of analysis across various metrics are depicted in Figure 5.
For the algorithms with Euclidean and Manhattan metrics: We found that one cluster was comprised of a single input variable and a single output variable. The second cluster contained all other variables. This distribution does not align with the constraints set forth by our CPDA model, which requires a separation of input and output variables into distinct clusters.
DBSCAN with Cosine Metric: Similarly, one cluster was composed of a single input variable and a single output variable, while the second cluster included all remaining variables. This configuration does not meet the requirements of our model.
Louvain Clustering with Cosine Metric: This algorithm produced a cluster made up of 3 output variables (OUT2, OUT3, OUT4) and a second cluster containing all remaining variables. Again, this clustering was not in line with the model's constraints as it mixed input and output variables.
Hierarchical Clustering with Cosine Metric: The algorithm produced a cluster containing 4 output variables (OUT2, OUT3, OUT4, OUT5), and a second cluster containing the remaining variables. Although closer to our constraints, this configuration still mixed input and output variables.
Hierarchical Clustering with Spearman Metric: Significantly, this algorithm was the only one to align with the constraints of our model and existing literature. It generated a cluster containing all 10 input variables and a second cluster containing all 5 output variables. This result suggests that the Spearman metric is particularly effective for our CPDA model.

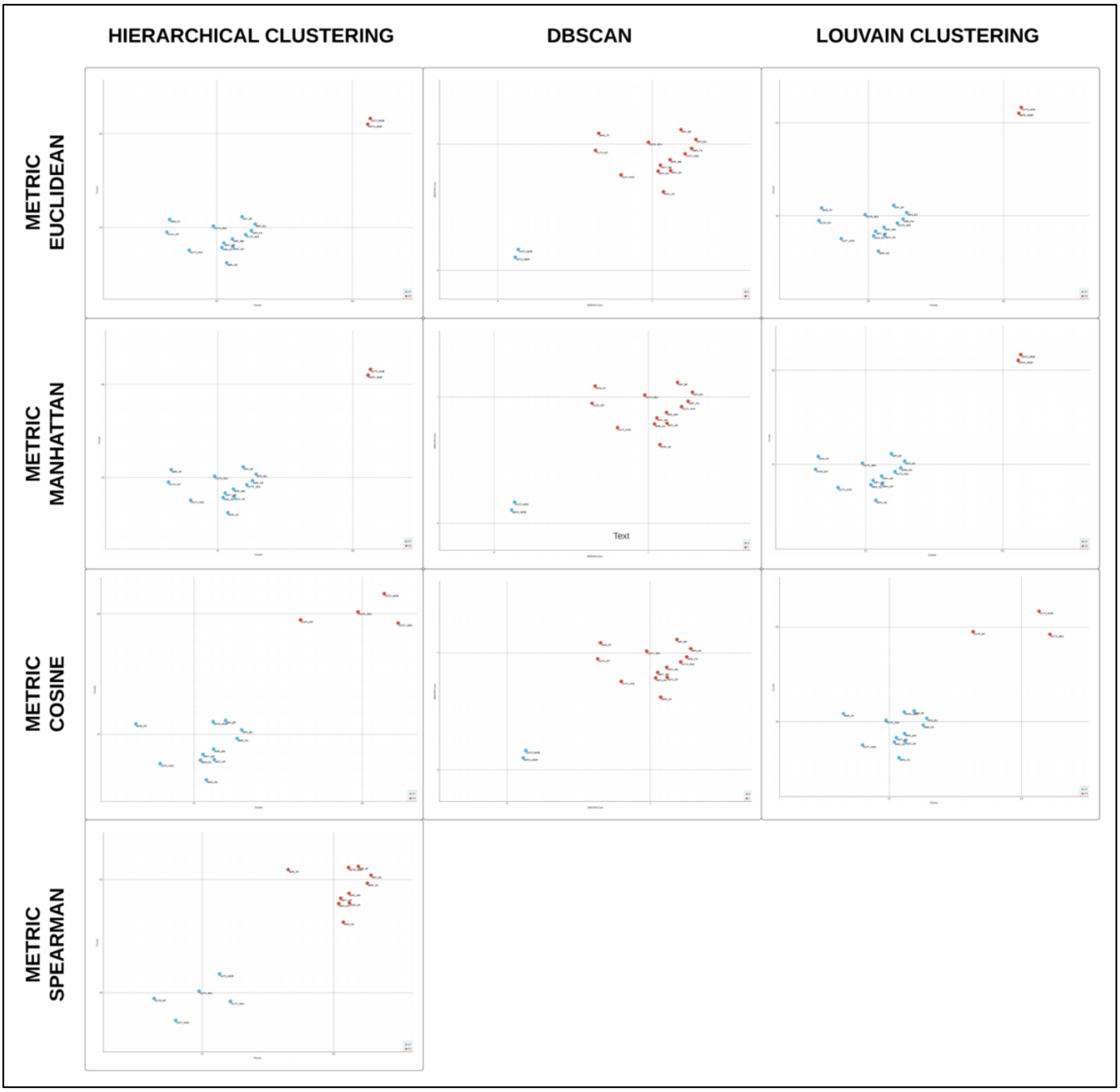


Figure 5 Scatterplots for the Three Clustering Algorithms.

The results revealed that Hierarchical Clustering with the Spearman distance metric was the only combination to align with existing literature. This alignment indicates the suitability of the algorithm for analyzing nuanced relationships between variables, thereby making it an ideal choice for our CPDA model.

Evaluation Criteria and Implications: We adopted the Silhouette coefficient as a standardized evaluation metric, given its reliability in indicating the quality of cohesion and separation between clusters. Specifically, we chose to use the Cosine distance metric in the calculation of the Silhouette coefficient (Figure 6) for its ability to handle high-dimensional spaces and for its emphasis on direction rather than the magnitude of variables.

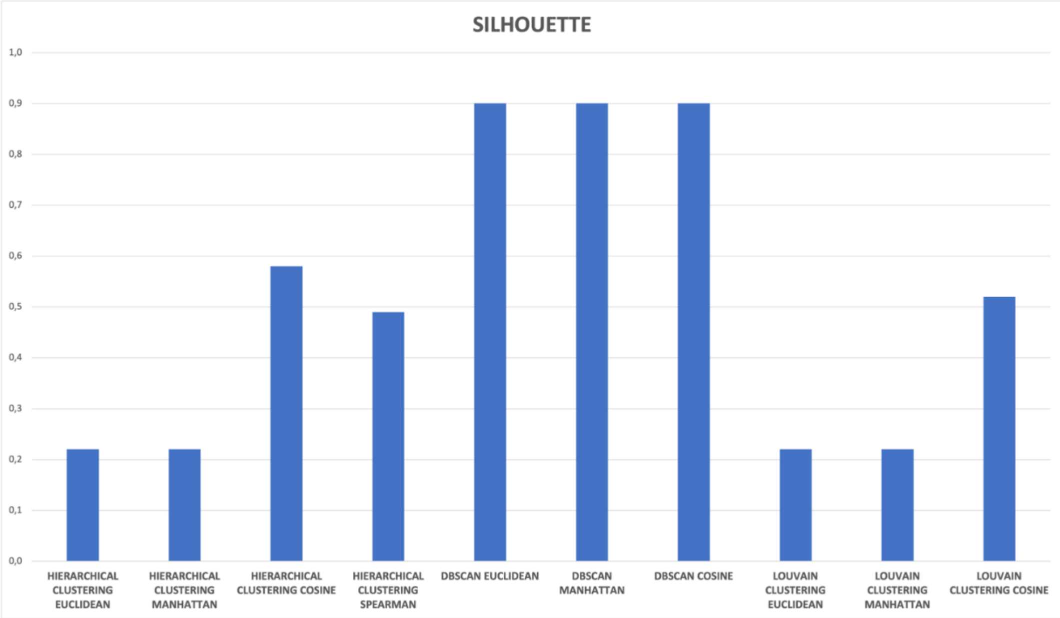


Figure 6 Barplot of Silhouette Coefficients

These characteristics are particularly relevant in the context of hospital data, which are often high-dimensional and require consideration of the correlation between variables. While DBSCAN showed the highest Silhouette coefficients (0.9) for all distance metrics, only Hierarchical Clustering with the Spearman metric satisfied all the specific constraints of our CPDA model, effectively separating the input and output variables into clusters. Despite its slightly lower Silhouette coefficient (0.5), this result highlights the importance of balancing metric-based optimization with the practical needs and constraints of the model. In summary, Hierarchical Clustering with the Spearman metric emerges as the most suitable approach to be integrated into the CPDA model, thereby enhancing its robustness and providing a methodological framework for future research in the field of hospital efficiency evaluation.

In conclusion, the choice of hierarchical clustering was driven by its ability to provide a hierarchical visual representation of the data, its efficiency in identifying clearly separated clusters, and its adaptability to the requirements of our dataset. This hierarchical clustering played a pivotal role as an initial filtration system in our analytical process, allowing us to identify and remove variables that did not meet our predetermined criteria. This streamlined the dataset, retaining only the most relevant variables for further analysis and ensuring that input and output variables were grouped in a coherent and meaningful manner. This approach provided valuable insights for further analysis, affirming the validity of our variable selection based on the scientific literature and seamlessly integrating it into our CPDA workflow.

3.10 HIERARCHICAL CLUSTER ANALYSIS

Hierarchical cluster analysis is a technique that allows for the exploration of hidden patterns within data by forming homogeneous groups of similar observations. In this study, we conducted a hierarchical cluster analysis on the set of 15 variables identified in the literature (para. 3.9.), pertaining to 59 hospitals in the Apulia region, with the aim of revealing patterns of similarity among observations and identifying potential meaningful groupings.

The employed algorithm for hierarchical clustering is the Agglomerative Clustering algorithm. This approach starts by treating each individual observation as a separate cluster and then iteratively merges the most similar clusters until a single cluster containing all observations is formed. The output of the algorithm is presented in the form of a dendrogram, which visualizes the hierarchy of clusters established during the merging process.
The dendrogram is a hierarchical structure diagram that illustrates the clustering process based on the distance between observations (Fig. 7). Within the dendrogram, two main clusters have emerged, each exhibiting distinct characteristics.

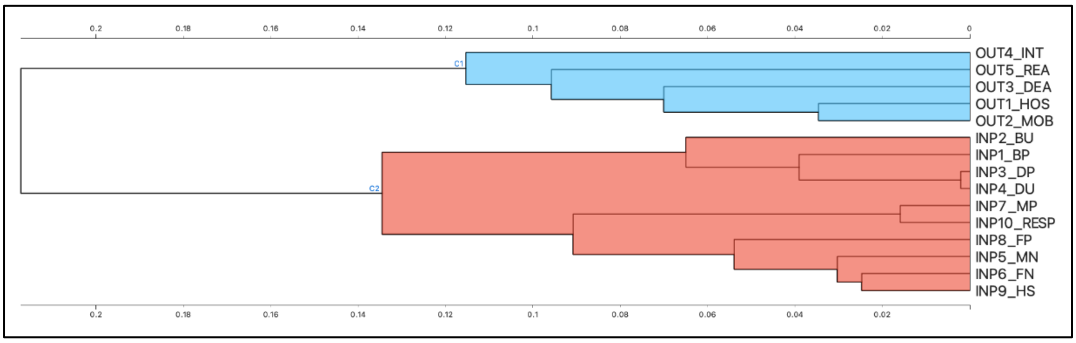


Figure 7 Dendrogram of Identified Healthcare Input and Output Variables.

The first cluster, denoted as C1, consists of the 5 output variables, while the second cluster, denoted as C2, is formed by the 10 input variables. In this clustering approach, we set the distance between columns and Spearman's metric using the "Distance" widget in Orange software, allowing us to gain a better understanding of how variables group together based on their similarities.
The use of Spearman's correlation metric in hierarchical clustering was necessary due to the non-normal distribution of the 15 examined variables. The Spearman correlation metric assesses the strength of relationships between variables based on their ranks rather than their actual values. By utilizing the "Hierarchical clustering" widget in Orange software, we set the Ward linkage as the method for calculating cluster distances, minimizing variance within combined clusters.
We consider an "height ratio" of 80%, indicating that clusters will be merged until the linkage height between them reaches 80% of the maximum dendrogram height. These approaches facilitated the grouping of variables based on their ordinal resemblance, identifying two distinct clusters.
In the context of hospital data analysis, hierarchical cluster analysis proves to be a powerful tool for understanding complex relationships between performance variables and organizational resources:
Cluster 1 - Performance Variables: This cluster aggregates hospital performance variables, including the number of admissions, mortality rates, hospital readmissions, and other quality-of-care metrics. This enables a comprehensive analysis of the overall effectiveness of provided healthcare. Cluster 1 provides an overview of medical performance and patient satisfaction – critical factors that can influence the hospital's reputation and patients' inclination to seek care from the facility.
Cluster 2 - Resource and Personnel Variables: This cluster includes variables related to resource management and personnel. Proper management of hospital departments, beds, and resources is essential for resource utilization optimization

and cost reduction. Ensuring an adequate number of available beds and departments can impact operational efficiency and the facility's capacity to accommodate a sufficient number of patients. Furthermore, the distribution of hospital staff, such as doctors and nurses, can influence wait times, care quality, and patient satisfaction.

3.11 HIERARCHICAL CLUSTERING ALGORITHM FOR INPUT AND OUTPUT VARIABLES

The analysis of the input and output variables from the data of the Apulian hospital structures is carried out using a hierarchical clustering algorithm. This method starts by treating each variable as an individual cluster and then repeatedly merges the most similar variables into broader clusters. The procedure follows these key steps:

1. Calculation of Spearman Correlation:
For each pair of variables, we compute the Spearman correlation, a non-parametric measure of the statistical correlation between two datasets. The formula is:

$$\rho_{XY} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad for\ i = 1, 2, \dots, n. \tag{3}$$

Where:

X and Y refer to two generic variables (columns of the dataset) for which the Spearman correlation is calculated;
d_i^2 represents the difference between the ranks of the corresponding observations in the two variables, and n is the total number of observations.
2. Creation of the Dissimilarity Matrix:
After calculating the correlation, the dissimilarity between two variables is determined as:

$$Dissimilarity_{XY} = 1 - \rho_{XY} \tag{4}$$

This dissimilarity matrix represents the distance between the variables and serves as the foundation for subsequent grouping.
3. Agglomerative Clustering using Ward's Method:
At each iteration, the two variables (or clusters of variables) with the minimum dissimilarity are merged into a single cluster. The Ward's method for the combined distance between two clusters, p and q, when they are merged, using a third and fourth cluster, r and s, as reference is:

$$D_{pq} = \sqrt{\frac{|p|+|r|}{T}\, x\, d(p,r)^2 + \frac{|p|+|s|}{T}\, x\, d(p,s)^2 + \frac{|r|+|s|}{T}\, x\, d(r,s)^2 - \frac{|r|}{T}\, x\, d(r)^2 + \frac{|s|}{T}\, x\, d(s)^2} \tag{5}$$

Where D_{pq} is the combined distance between clusters p and q, and $d(p,r)$, $d(p,s)$ and $d(r,s)$ are the original distances between the clusters.
$|p|, |r|\ and\ |s|$ represent the sizes of the clusters, whereas T is the total size of the dataset.
4**. Selection of Final Clusters based on the Hight Ratio:**

Once the complete hierarchy of clusters has been constructed, the final two clusters are selected to capture 80% of the overall information. This can be determined by examining the dendrogram and cutting the tree at a level where 80% of the total information (or variance) is covered.

5. Visualization with a Dendrogram:

The dendrogram will display the hierarchical nature of the clustering process, but the main focus will be on the final two clusters that represent the groups of input and output variables respectively.

Based on the results obtained from the hierarchical clustering analysis and considering that these results align with expectations based on previous literature, we can conclude that the 15 variables previously identified in the literature for the DEA analysis have been correctly assigned as inputs and outputs. This confirms the validity of the initial choices and provides a solid foundation for subsequent analyses and interpretations.

3.12 STANDARDIZATION

Standardizing variables helps improve the stability of PCA and avoid variables with high variances dominating the analysis.

The original variables are transformed into new variables that have a mean of zero and a standard deviation of one. Standardizing variables before PCA is an important step to ensure that the analysis is accurate and that variables are evaluated correctly based on their importance in the variation of the data.

The general formula for standardizing a data set with n input variables and m output variables can be written as (1):

$$Z_i = \frac{(X_i - \mu_i)}{\sigma_i} \; for \; i = 1, 2, \ldots, n. \tag{6}$$

where:

Z_i is the standardized value of the i-th variable (input or output);

X_i is the respective non-standardized value of the i-th variable (input or output);

μ_i is the mean of all observations in the i-th variable (input or output);

σ_i is the standard deviation of all observations in the i-th variable (input or output);

n is the number of variables (=15).

3.13 EXPLORATORY FACTOR ANALYSIS

Applying an Exploratory Factor Analysis (EFA) separately for each cluster will allow examining the data structure within each cluster and identifying any underlying factors that explain the relationships between the variables. The task of exploratory factor analysis is to group variables together based on their higher correlations. The so-called "factor" is seen as a latent variable that influences the observed variables. The correlation matrices for the considered variables demonstrate how within each cluster, the variables are strongly correlated with each other (Figure 8).

Cluster 2

Correlation Matrix

		INP1_BP	INP2_BU	INP3_DP	INP4_DU	INP5_MN	INP6_FN	INP7_MP	INP8_FP	INP9_HS	INP10_RESP
INP1_BP	Spearman's rho	—									
	p-value	—									
INP2_BU	Spearman's rho	0.920 ***	—								
	p-value	< .001	—								
INP3_DP	Spearman's rho	0.938 ***	0.879 ***	—							
	p-value	< .001	< .001	—							
INP4_DU	Spearman's rho	0.927 ***	0.874 ***	0.996 ***	—						
	p-value	< .001	< .001	< .001	—						
INP5_MN	Spearman's rho	0.862 ***	0.853 ***	0.907 ***	0.902 ***	—					
	p-value	< .001	< .001	< .001	< .001	—					
INP6_FN	Spearman's rho	0.847 ***	0.848 ***	0.893 ***	0.890 ***	0.948 ***	—				
	p-value	< .001	< .001	< .001	< .001	< .001	—				
INP7_MP	Spearman's rho	0.770 ***	0.771 ***	0.883 ***	0.887 ***	0.865 ***	0.846 ***	—			
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	—			
INP8_FP	Spearman's rho	0.854 ***	0.792 ***	0.893 ***	0.892 ***	0.902 ***	0.904 ***	0.875 ***	—		
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—		
INP9_HS	Spearman's rho	0.863 ***	0.847 ***	0.901 ***	0.901 ***	0.936 ***	0.951 ***	0.860 ***	0.914 ***	—	
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—	
INP10_RESP	Spearman's rho	0.795 ***	0.764 ***	0.885 ***	0.886 ***	0.881 ***	0.886 ***	0.968 ***	0.934 ***	0.899 ***	—
	p-value	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	—

Note. * p < .05, ** p < .01, *** p < .001

Cluster 1

Correlation Matrix

		OUT1_HOS	OUT2_MOB	OUT3_DEA	OUT4_INT	OUT5_REA
OUT1_HOS	Spearman's rho	—				
	p-value	—				
OUT2_MOB	Spearman's rho	0.931 ***	—			
	p-value	< .001	—			
OUT3_DEA	Spearman's rho	0.877 ***	0.871 ***	—		
	p-value	< .001	< .001	—		
OUT4_INT	Spearman's rho	0.804 ***	0.842 ***	0.765 ***	—	
	p-value	< .001	< .001	< .001	—	
OUT5_REA	Spearman's rho	0.854 ***	0.843 ***	0.801 ***	0.789 ***	—
	p-value	< .001	< .001	< .001	< .001	—

Note. * p < .05, ** p < .01, *** p < .001

Figure 8 Correlation matrixes for the two clusters.

The results of the exploratory factor analyses applied separately to the two identified clusters are presented in Figure 9.

Exploratory Factor Analysis - Cluster 2

Factor Loadings

	Factor 1	Uniqueness
INP1_BP	0.980	0.0405
INP2_BU	0.943	0.1102
INP3_DP	0.986	0.0271
INP4_DU	0.978	0.0430
INP5_MN	0.977	0.0461
INP6_FN	0.973	0.0540
INP7_MP	0.985	0.0301
INP8_FP	0.965	0.0685
INP9_HS	0.987	0.0252
INP10_RESP	0.968	0.0633

Note. 'Minimum residual' extraction method was used in combination with a 'oblimin' rotation

Factor Statistics

Summary

Factor	SS Loadings	% of Variance	Cumulative %
1	9.49	94.9	94.9

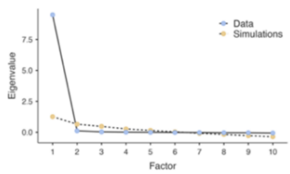
Assumption Checks

Bartlett's Test of Sphericity

χ^2	df	p
1614	45	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.887
INP1_BP	0.940
INP2_BU	0.922
INP3_DP	0.885
INP4_DU	0.873
INP5_MN	0.947
INP6_FN	0.856
INP7_MP	0.895
INP8_FP	0.879
INP9_HS	0.844
INP10_RESP	0.846



Exploratory Factor Analysis - Cluster 1

Factor Loadings

	Factor 1	Uniqueness
OUT1_HOS	0.954	0.0907
OUT2_MOB	0.904	0.1832
OUT3_DEA	0.903	0.1844
OUT4_INT	0.928	0.1384
OUT5_REA	0.948	0.1020

Note. 'Minimum residual' extraction method was used in combination with a 'oblimin' rotation

Factor Statistics

Summary

Factor	SS Loadings	% of Variance	Cumulative %
1	4.30	86.0	86.0

Assumption Checks

Bartlett's Test of Sphericity

χ^2	df	p
363	10	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.917
OUT1_HOS	0.888
OUT2_MOB	0.937
OUT3_DEA	0.943
OUT4_INT	0.930
OUT5_REA	0.895

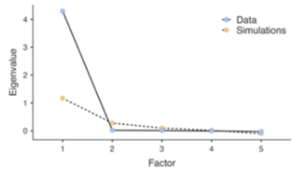


Figure 9 Factor Analysis results for the 2 clusters.

The exploratory factor analysis of the two clusters revealed significant findings. In Cluster 2, consisting of 10 variables, a single latent factor emerged that explains 94.9% of the total variance within the cluster. All variables within the cluster exhibit correlations above 0.96 with the factor, indicating a strong association among them. The Bartlett's test of sphericity confirmed the presence of a significant factor structure in the cluster, with a p-value below 0.001. Additionally, the KMO MSA indicates adequate data suitability for factor analysis in the cluster (0.887).

In Cluster 1, consisting of 5 variables, a single latent factor was identified that explains 86.0% of the total variance within the cluster. All variables within the cluster exhibit correlations above 0.90 with the factor, highlighting a strong association among them. The Bartlett's test of sphericity confirmed the presence of

a significant factor structure in the cluster, with a p-value below 0.001. Additionally, the KMO MSA indicates adequate data suitability for factor analysis in the cluster (0.917).
The screening test based on the parallel analysis confirmed the importance of one factor in both clusters.
In conclusion, both clusters demonstrate significant factor structures and strong associations among variables. These findings indicate the presence of a latent factor in each cluster that can consistently explain the observed variations in their respective variables.

3.14 RELIABILITY ANALYSIS

In this phase of the study, we will focus on the reliability analysis of the two previously identified clusters. The aim is to evaluate the internal consistency of the measures within each cluster and determine whether the selected variables reliably represent the latent factor identified for each cluster.
To make the latent variables "measurable," a scale is used. A scale is a group of variables, in our case, that are collectively used to measure a latent factor. If these variables are highly correlated, it is referred to as high internal consistency. Cronbach's alpha is a measure of the internal consistency of a scale. By calculating Cronbach's alpha for the two clusters, we obtain the results shown in Figure 10.

Cluster 2

Scale Reliability Statistics

	Cronbach's α
scale	0.995

Cluster 1

Scale Reliability Statistics

	Cronbach's α
scale	0.968

Figure 10 Cronbach's alpha for the two clusters.

The calculated Cronbach's alpha is 0.995 for cluster 2 and 0.968 for cluster 1, indicating that the selected variables within the two clusters are consistent with each other and reliably measure the respective single latent factor identified for each cluster.
These results further confirm the choice of a single factor for each cluster, as the data suggest a strong internal consistency of the measures within each cluster.
Based on the results obtained from the cluster analysis, factor analysis, and reliability analysis, both for cluster 2 and cluster 1, the subsequent PCA can be conducted considering a single principal component, in line with the data structure and the coherence of the measures within each cluster.

3.15 EXPLORATORY FACTOR ANALYSIS AND RELIABILITY IN HOSPITAL EFFICIENCY EVALUATION: METHODOLOGICAL CHOICES, ADVANTAGES, AND LIMITATIONS

In the context of evaluating hospital efficiency in the Apulia region, this study adopted the Exploratory Factor Analysis (EFA) and Reliability Analysis to explore and validate latent structures in the data.
Methodological Limitations:

- **Sample Size:** A limited sampling of hospitals in Apulia could compromise the robustness of the EFA, influencing the proper identification of latent variables.
- **Sample Dependence:** Being sensitive to the nature of the sample, EFA might produce diverging results with data from different regions or timeframes.

- **Subjective Interpretation:** Given the novelty of the CPDA approach, interpreting emerging factors might require further scrutiny, with the risk of biases.

Advantages of the Chosen Approach:

- **Data Exploration:** EFA, with its capacity to unveil latent structures, is apt for examining hospital data without predefined assumptions.
- **Robustness and Reliability:** While various metrics exist for assessing internal consistency, such as McDonald's Omega, Cronbach's Alpha was chosen due to its familiarity and widespread use in research. Its established interpretation and capability to provide a consistent and reliable measurement of variables made it the preferred choice in this study's context.
- **Foundation for Subsequent Analysis:** After determining the factorial structure with EFA, subsequent analysis can proceed on solid ground.

Relation to Alternatives:

- **Depth of Analysis:** Compared to alternatives like Confirmatory Factor Analysis (CFA), EFA offers an exploratory view, fitting for the CPDA approach.

Considering the specific context and needs of the present study, the combined adoption of EFA and Reliability Analysis appears well-motivated and well-calibrated, providing a solid foundation for further inquiries in the realm of hospital efficiency evaluation.

3.16 PRINCIPAL COMPONENT ANALYSIS

The problem of insufficient discriminating power is often overlooked in DEA studies, which can occur when the number of Decision-Making Units (DMU) rated as efficient exceeds the number of inputs and outputs. The number of efficient DMUs is dependent on the number of variables considered, with larger numbers of variables resulting in less demanding analysis. To address this issue, it's necessary to reduce the number of variables in the DEA structure. One technique used for this purpose is principal component analysis (PCA), which involves simplifying source data to maximize variance by calculating the weight to be given to each source variable. This results in one or more new variables (called principal components) that are linear combinations of the source variables, representing the characteristics of the starting phenomenon.

PCA was applied separately to the inputs and outputs, with the respective principal components determined. The analyses were conducted using Orange software. For both inputs and outputs, only one main factor was identified.

Applying PCA to the first group of input variables (Cluster 2), we obtained a principal component ($Input_{PC1}$) preserving almost 95% of the total variance with minimal loss of information.

Applying PCA to the second group of output variables (Cluster 1), one main components ($Output_{PC1}$) was identified, preserving almost 89% of the total variance. The graphical representation produced by Orange Software's 'Pca' widget is shown in Figure 11. The parameters of the PCA analyzes are represented in figure 13.

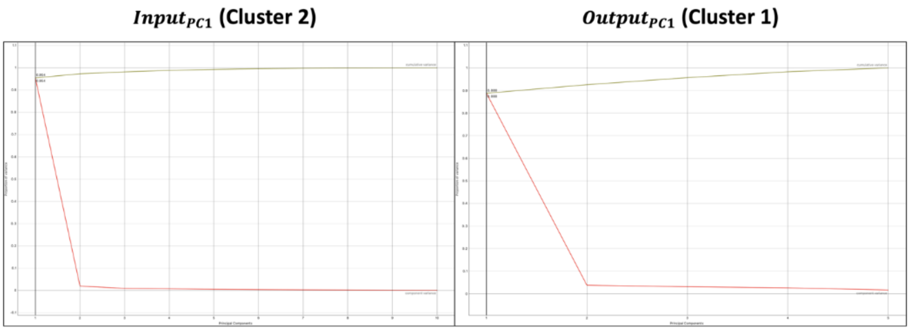


Figure 11 Principal Components variance representation.

The impact of each individual original variable on the main components can be easily visualized in Figure 12 of the report.

Data Table

Data instances: 1
Features: 10
Meta attributes: 2

	components	variance	INP7_MP	INP10_RESP	INP8_FP	INP5_MN	INP6_FN	INP9_HS	INP2_BU	INP3_DP	INP4_DU	INP1_BP
1	PC1	0.954227	0.318983	0.314576	0.313861	0.316871	0.315813	0.319617	0.308111	0.319368	0.317284	0.317624

Data Table

Data instances: 1
Features: 5
Meta attributes: 2

	components	variance	OUT4_INT	OUT5_REA	OUT3_DEA	OUT1_HOS	OUT2_MOB
1	PC1	0.88791	0.447566	0.453067	0.44015	0.454725	0.44035

Figure 12 The impact of each individual original variable on the main components.

3.17 PRINCIPAL COMPONENT ALGORITHM FOR INPUT AND OUTPUT CLUSTERS

Principal Component Analysis (PCA) is a dimensionality reduction technique. The PCA algorithm used is described in detail below:

1. Covariance Matrix Calculation:
Once the data is standardized, we compute the covariance matrix, denoted as C or "Covariance matrix" from the standardized data.

2. Eigenvalues and Eigenvectors Calculation:
From the covariance matrix C, eigenvalues λ and eigenvectors υ are derived. These are computed by solving:

$$C\upsilon = \lambda\upsilon \tag{7}$$

3. Selection of Principal Components:
The eigenvalues are arranged in descending order. We select the k principal components that account for the majority of the variance.

4. Transformation of Original Data:
Using the eigenvectors of the selected principal components, the original data is transformed into:

$$Y = Z * V \tag{8}$$

In the context of the two analyzed clusters:
The results for the two main components identified indicate a fairly uniform distribution of incidence across all variables.
The formulas to calculate the principal components ($Input_{PC1}$ e $Output_{PC1}$) of the input variables Z_1, Z_2, ..., Z_{17} are:

$$Input_{PC1} = \sum_{i=1}^{10} a_i * Z_i \tag{9}$$

$$Output_{PC1} = \sum_{i=1}^{5} b_i * Z_i \tag{10}$$

where:
$Input_{PC1}$ is the value of the first principal component of the input variables;
a_i is the coefficient of the i-th input variable Z_i in the principal component;
$Output_{PC1}$ is the value of the first principal component of the output variables;
b_i is the coefficient of the i-th output variable Z_i in the principal component.
The coefficients a_i e b_i are calculated as optimal weights that maximize the variance of the principal component subject to the constraint that the sum of their squares equals 1. In other words, the coefficients are determined as the eigenvectors associated with the maximum eigenvalue of the covariance matrix, respectively of the input and output variables.
As a result, the first principal component of the input variables ($Input_{PC1}$) was renamed Hospital Organization, while the first principal component of the output variables ($Output_{PC1}$) was renamed Propension Hospitalization.

Cluster 2 (Input Variables):

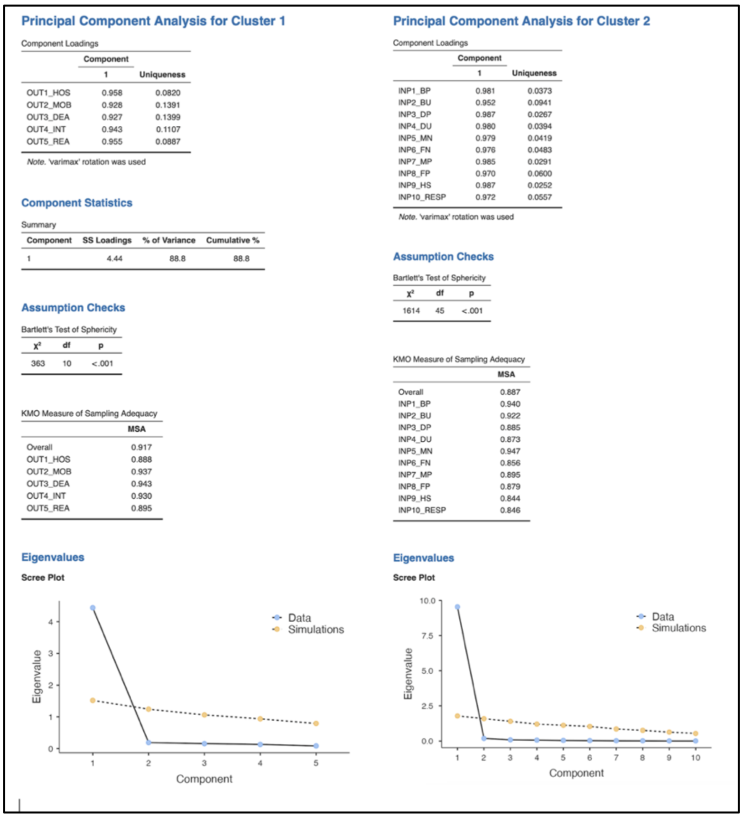
The first principal component is a linear combination of the standardized input variables:

$$\boldsymbol{Hospital\ Organization} = Input_{PC1} = 0.317624 * Z_{INP1_{BP}} + 0.308111 * Z_{INP2_{BU}} + 0.319368 * Z_{INP3_{DP}} + 0.317284 * Z_{INP4_{DU}} + 0.316871 * Z_{INP5_{MN}} + 0.315813 * Z_{INP6_{FN}} + 0.318983 * Z_{INP7_{MP}} + 0.313861 * Z_{INP8_{FP}} + 0.319617 * Z_{INP9_{HS}} + 0.314576 * Z_{INP10_{RESP}} \tag{11}$$

Cluster 1 (Output Variables):

Similarly, the first principal component for the output variables is:

$$\boldsymbol{Propension\ Hospitalization} = Output_{PC1} = 0.454725 * Z_{OUT1_{HOS}} + 0.440350 * Z_{OUT2_{MOB}} + 0.440150 * Z_{OUT3_{DEA}} + 0.447566 * Z_{OUT4_{INT}} + 0.453067 * Z_{OUT5_{REA}} \tag{12}$$



Principal Component Analysis for Cluster 1

Component Loadings

	Component 1	Uniqueness
OUT1_HOS	0.958	0.0820
OUT2_MOB	0.928	0.1391
OUT3_DEA	0.927	0.1399
OUT4_INT	0.943	0.1107
OUT5_REA	0.955	0.0887

Note. 'varimax' rotation was used

Component Statistics

Summary

Component	SS Loadings	% of Variance	Cumulative %
1	4.44	88.8	88.8

Assumption Checks

Bartlett's Test of Sphericity

χ^2	df	p
363	10	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.917
OUT1_HOS	0.888
OUT2_MOB	0.937
OUT3_DEA	0.943
OUT4_INT	0.930
OUT5_REA	0.895

Eigenvalues

Scree Plot

Principal Component Analysis for Cluster 2

Component Loadings

	Component 1	Uniqueness
INP1_BP	0.981	0.0373
INP2_BU	0.952	0.0941
INP3_DP	0.987	0.0267
INP4_DU	0.980	0.0394
INP5_MN	0.979	0.0419
INP6_FN	0.976	0.0483
INP7_MP	0.985	0.0291
INP8_FP	0.970	0.0600
INP9_HS	0.987	0.0252
INP10_RESP	0.972	0.0557

Note. 'varimax' rotation was used

Assumption Checks

Bartlett's Test of Sphericity

χ^2	df	p
1614	45	< .001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.887
INP1_BP	0.940
INP2_BU	0.922
INP3_DP	0.885
INP4_DU	0.873
INP5_MN	0.947
INP6_FN	0.856
INP7_MP	0.895
INP8_FP	0.879
INP9_HS	0.844
INP10_RESP	0.846

Eigenvalues

Scree Plot

Figure 13 Principal components Analysis parameters.

3.18 POSITIVE SHIFT

The "positivization" of principal components identified through PCA refers to a transformation process applied to the components to obtain only positive values. This can be useful in some situations where the principal components are used as input for further analysis, such as DEA analysis. Positivization can help eliminate any negative effects of variables with negative values, which could negatively impact efficiency evaluations.

The method of adding a constant involves adding a sufficiently large positive constant to the principal components so that all values become positive. In this case, all values have been increased by the minimum value plus one (Hajiagha et al., 2023). Since the minimum value is equal to -2.43053, identified with the feature statistics widget, all individual values comprising the two variables will be increased by 3.43053. We will obtain the two new principal components, which are continuous and positive, not normally distributed, and have the same mean.

If the two principal components are $Input_{PC1}$ and $Output_{PC1}$, the formulas to make the components positive can be expressed as:

$$Input_{PC1_{pos}} = Input_{PC1} + |\min(\mathrm{Input_{PC1}}, \mathrm{Output_{PC1}})| + k \qquad (13)$$

$$Output_{PC1_{pos}} = Output_{PC1} + |\min(\mathrm{Input_{PC1}}, \mathrm{Output_{PC1}})| + k \qquad (14)$$

Where:

$Input_{PC1_{pos}}$ represents the positive value of the first principal component of the input variables;

$Output_{PC1_{pos}}$ represents the positive value of the first principal component of the output variables;
$Input_{PC1}$ is the value of the first principal component of the input variables;
$Output_{PC1}$ is the value of the first principal component of the output variables;
$\min(\mathrm{Input}_{\mathrm{PC1}}, \mathrm{Output}_{\mathrm{PC1}})$ represents the minimum value between the first principal component of the Input variables and the first principal component among the Output variables;
k = 1, represents a sufficiently large positive constant added to ensure that all resulting values are positive.

3.19 DATA ENVELOPMENT ANALYSIS

The methodology used in this study is Data Envelopment Analysis (DEA), which is appropriate for providing an efficiency score to each observation (hospital facility), as suggested by Charnes et al. (1978). In this study, efficiency was defined as the ability of each region to maximize the number of patients attracted to its health facilities to receive appropriate medical treatment, given the "technical" resources at its disposal (medical staff, beds, etc.).
In the first phase of the study approach, principal component analysis (PCA) was applied to the input and output variables described in Table 4. Two new variables were derived from this analysis, which were then used in the second phase of the study to determine efficiency scores with the application of DEA (Guede-Cid et al., 2021; Hajiagha et al., 2023).
The chosen approach is output-oriented, maximizing outputs while holding inputs constant. Variable returns to scale (VRS) and unrestricted weights were assumed, as the hospital facilities analyzed differ based on characteristics peculiar to each territorial area. The output-oriented model is therefore the most appropriate tool to represent the ability of each region to maximize the number of patients attracted to its health facilities to receive appropriate medical treatment, given the "technical" resources at its disposal.
Weight restrictions were not introduced in this study, allowing for complete weight flexibility to take advantage of the freedom of unit behavior offered by DEA. Without any restrictions on weights, each hospital can be efficient by operating in its way (Guede-Cid et al., 2021).
In addition, the 15 iidentified variables, grouped into two clusters identified through cluster analysis, underwent Data Envelopment Analysis using an output-oriented VRS (variable returns to scale) model, without the application of Cluster and Principal Component Analysis.
The efficiency scores derived from the DEA analysis for the first model (DEA model), without the application of Cluster-PCA analysis, and for the second model, with the application of Cluster-PCA analysis, were respectively recorded in the PTE_{DEA} column and the PTE_{CPDA}A column in Table 4.
The differences in discriminatory power between the two models for the considered decision-making units will be analyzed and discussed in the upcoming Section 4.3, and graphically represented in Figure 22.

3.20 DATA ENVELOPMENT ANALYSIS ALGORITHM FOR ASSESSING EFFICIENCY USING ORIGINAL VARIABLES AS INPUTS AND OUTPUTS: DEA MODEL

Objective: Assess the efficiency of Decision-Making Units (DMUs) using the DEA model with variable returns to scale (VRS), output-oriented, without weights, based on the original 10 input and 5 output variables.
The DEA algorithm used is described in detail below:

1. Initialization:
For each DMU i:
- Define $x_{i1}, x_{i2}, \dots, x_{i10}$ as the value of 10 input variables.
- Define $y_{i1}, y_{i2}, \dots, y_{i5}$ as the value of the 5 output variables.

2. Optimization:
- $Maximize$ s, for each DMU, subject to:

$$s \leq \frac{\sum_{K=1}^{5} y_{ik}}{\sum_{J=1}^{10} x_{IJ}}, for\ i = 1, \dots, n \tag{15}$$

Where n is the total number of decision-making units (DMU).

3. Output:
- Return the efficiency score s for each DMU. A score of 1 indicates efficiency, while scores below 1 indicate relative inefficiency.

The problem seeks to maximize relative efficiency subject to the constraint of not exceeding the sum of the five original output variables divided by the sum of the ten original input variables for each decision-making unit i.
The resulting efficiency scores for the 59 hospitals in Apulia are shown in table 4, column DEA.

3.21 DATA ENVELOPMENT ANALYSIS ALGORITHM FOR ASSESSING EFFICIENCY USING PRINCIPAL COMPONENTS AS INPUTS AND OUTPUTS: CPDA MODEL

Objective: Assess the efficiency of Decision-Making Units (DMUs) using the DEA model with variable returns to scale (VRS), output-oriented, without weights.
The DEA algorithm used is described in detail below:

1. Initialization:
- Define x_i as the value of $Input_{PC1_{pos}}$ for the i-th DMU.
- Define y_i as the value of $Output_{PC1_{pos}}$ for the i-th DMU.

2. Optimization:
- Max s, subject to:

$$s \leq \frac{x_i}{y_i}, for\ i = 1, \dots, n. \tag{16}$$

Where n is the total number of decision-making units (DMU).

3. Output:
- Return the efficiency score s for each DMU. A score of 1 indicates efficiency, while scores below 1 indicate relative inefficiency.

The problem seeks to maximize relative efficiency subject to the constraint of not exceeding the value of $Output_{PC1_{pos}}$ divided by the value of $Input_{PC1_{pos}}$ for each decision-making unit i.

The resulting efficiency scores for the 59 hospitals in Apulia are shown in table 9, column CPDA.

Table 9. Efficiency Scores Expressed for 59 Hospitals in Apulia using DEA and Cluster-PCA-DEA Analysis.

ASL	NETWORK	HOSPITAL LEVEL	DMU	PTE_DEA	PTE_CPDA
ASL BA	PRIVATE	FIRST LEVEL	Ospedale Regionale EE 'Miulli' Acquaviva Delle Fonti (BA)	1	1
ASL BA	PUBLIC	FIRST LEVEL	Ospedale Della Murgia - Perinei Altamura (BA)	0,748337	0,487939
ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Anthea Bari (BA)	1	0,759454
ASL BA	PRIVATE	FIRST LEVEL	Casa Di Cura C.B.H. Mater Dei Hospital Bari (BA)	1	0,734477
ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Santa Maria Bari (BA)	1	0,974464
ASL BA	PUBLIC	SECOND LEVEL	Consorziale Policlinico Bari Bari (BA)	1	1
ASL BA	PRIVATE	IRCCS	Ics Maugeri SPA Societa' Benefit Bari (BA)	0,713011	0,526283
ASL BA	PUBLIC	IRCCS	Istituto Tumori Giovanni Paolo II Bari (BA)	0,978384	0,533106
ASL BA	PUBLIC	FIRST LEVEL	Ospedale Di Venere Bari (BA)	1	0,610928
ASL BA	PUBLIC	FIRST LEVEL	Ospedale San Paolo Bari (BA)	1	0,60401
ASL BA	PUBLIC	IRCCS	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	0,915301	0,646663
ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura - Villa Lucia Hospital Conversano (BA)	1	0,739994
ASL BA	PUBLIC	BASE LEVEL	Ospedale Monopoli Monopoli (BA)	0,864865	0,527683
ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura ' Monte Imperatore' - Noci (BA)	0,561814	0,632582
ASL BA	PUBLIC	BASE LEVEL	Ospedale Putignano Putignano (BA)	0,694008	0,491971
ASL BR	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura 'Salus' Brindisi (BR)	0,696644	0,701617
ASL BR	PRIVATE	IRCCS	IRCCS 'E.Medea' - Brindisi (BR)	1	0,776651
ASL BR	PUBLIC	SECOND LEVEL	Ospedale Perrino Brindisi (BR)	0,773561	0,661974
ASL BR	PUBLIC	FIRST LEVEL	Ospedale Francavilla Fontana Francavilla Fontana (BR)	1	0,56429
ASL BR	PUBLIC	BASE LEVEL	Ospedale Ostuni Ostuni (BR)	1	0,750455
ASL BT	PUBLIC	FIRST LEVEL	Ospedale Andria Andria (BT)	1	0,557563
ASL BT	PUBLIC	FIRST LEVEL	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	1	0,509324
ASL BT	PUBLIC	BASE LEVEL	Ospedale Bisceglie Bisceglie (BT)	0,851196	0,407984
ASL BT	PRIVATE	PRIVATE NURSING HOMES	Ospedale Opera Don Uva Bisceglie (BT)	0,69266	0,539597
ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Leonardo De Luca Castelnuovo Della Daunia (FG)	1	0,7431
ASL FG	PUBLIC	FIRST LEVEL	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	0,70621	0,464968
ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Prof. Brodetti Foggia (FG)	1	0,71789
ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Universo Salute - Don Uva Foggia (FG)	1	0,613089
ASL FG	PRIVATE	PRIVATE NURSING HOMES	Case Cura Riunite Villa Serena-S. Francesco Foggia (FG)	1	0,936337
ASL FG	PUBLIC	SECOND LEVEL	Ospedali Riuniti Di Foggia Foggia (FG)	1	0,771869
ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura 'S.Michele' Gest. Brodetti Manfredonia (FG)	1	1
ASL FG	PUBLIC	BASE LEVEL	Ospedale Manfredonia Manfredonia (FG)	0,417031	0,450104
ASL FG	PRIVATE	FIRST LEVEL	Ospedale Casa Sollievo Della Sofferenza San Giovanni Rotondo (FG)	1	0,644767
ASL FG	PUBLIC	FIRST LEVEL	Ospedale San Severo - Teresa Masselli San Severo (FG)	0,704223	0,520972
ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Riabilitativa Euroitalia - Casarano (LE)	0,740515	0,700656
ASL LE	PUBLIC	BASE LEVEL	Ospedale Casarano Casarano (LE)	0,821128	0,576929
ASL LE	PUBLIC	BASE LEVEL	Ospedale Copertino Copertino (LE)	0,837988	0,580115
ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura San Francesco Galatina (LE)	1	0,60302
ASL LE	PUBLIC	FIRST LEVEL	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	0,710385	0,602163
ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura 'Prof. Petrucciani' SRL Lecce (LE)	0,891143	0,609719
ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Citta' Di Lecce Lecce (LE)	1	1
ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Villa Bianca Lecce (LE)	1	1
ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Villa Verde - Lecce (LE)	1	0,65533
ASL LE	PUBLIC	SECOND LEVEL	Ospedale Lecce 'V. Fazzi' Lecce (LE)	1	0,867505
ASL LE	PUBLIC	FIRST LEVEL	Ospedale Scorrano Scorrano (LE)	0,996396	0,696432
ASL LE	PRIVATE	FIRST LEVEL	Ospedale Regionale EE 'G. Panico' Tricase (LE)	1	1
ASL LE	PUBLIC	BASE LEVEL	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	0,544061	0,391888
ASL TA	PUBLIC	FIRST LEVEL	Ospedale Castellaneta Castellaneta (TA)	0,792742	0,536464
ASL TA	PRIVATE	PRIVATE NURSING HOMES	Centro Medico Riabilitazione Ics Maugeri Ginosa (TA)	0,760352	0,771804
ASL TA	PUBLIC	BASE LEVEL	Ospedale Manduria 'Giannuzzi' Manduria (TA)	0,813625	0,538073
ASL TA	PRIVATE	FIRST LEVEL	Casa Di Cura Villa Bianca SRL - Martina Franca (TA)	0,755585	0,709735
ASL TA	PUBLIC	FIRST LEVEL	Ospedale Civile Martina Franca (TA)	1	0,679362
ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Bernardini Taranto (TA)	1	0,739977
ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura D'Amore SRL Taranto (TA)	1	0,841161
ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura San Camillo Taranto (TA)	1	0,754517
ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Santa Rita SRL Taranto (TA)	1	0,709031
ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Villa Verde SRL Taranto (TA)	1	0,755949
ASL TA	PRIVATE	PRIVATE NURSING HOMES	Fondazione Cittadella Della Carita` Taranto (TA)	1	0,728615
ASL TA	PUBLIC	SECOND LEVEL	Presidio Ospedaliero centrale Taranto (TA)	0,905823	0,691367

3.22 ANOVA ANALYSIS

The distribution of efficiency scores can influence the choice of statistical tools used for data analysis.
Generally, the distribution of efficiency scores produced by the DEA analysis should follow a normal distribution. This is vital because it allows the use of standard statistical techniques such as the analysis of variance (ANOVA) to evaluate efficiency differences among different hospital groups.
To ensure a more normal distribution of the data, a base-10 logarithmic transformation was applied to the PTE variable using the "Feature Constructor" widget in Orange. The formula used for this transformation is:

$$\mathrm{PTE}_{\mathrm{normalized}} = LOG_{10}(PTE) \tag{17}$$

This transformation allowed the necessary assumptions for the ANOVA analysis to be met, making the data more suitable for this kind of analysis.
The distribution of efficiency scores is graphically represented through the "distribution" widget in Figure 14.

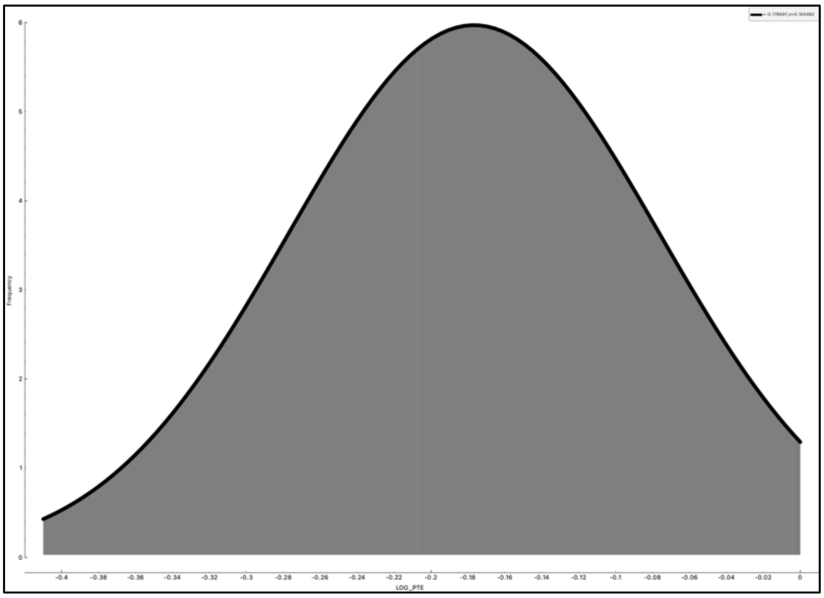


Figure 14 Normal distribution of Pure Technical Efficiency (PTE).

The objective of ANOVA analysis is to identify which factors have a significant impact on hospital efficiency, in order to implement targeted interventions to improve the performance of different levels of hospitals. The ANOVA model was used to assess the difference in efficiency among the hospitals level (figure 15). Level was assigned according to type management.

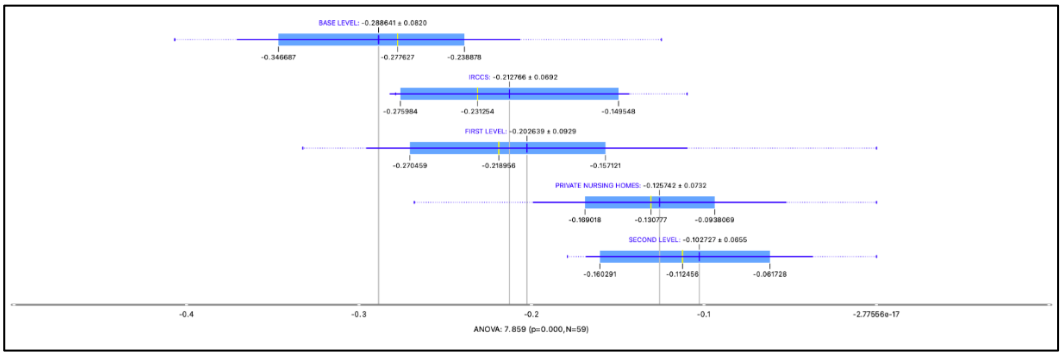


Figure 15 ANOVA analysis by hospital level.

According to the resolution of the Apulian Regional Council on September 23, 2019, No. 1726, the hospital network is divided as follows: 5 second-level hospitals, 17 first-level hospitals, and 9 basic hospitals. The hospital network is further complemented by 4 Scientific Research and Treatment Institutes and 24 accredited private healthcare facilities.

The results show that there is a significant difference in efficiency between hospitals levels (ANOVA=7.859, p=0.000, N=59).

In particular, the group of second-level hospitals has the highest efficiency score (-0.102727 +/- 0.0655), followed by private nursing homes (-0.125742 +/- 0.0732 and first level (-0.202639 +/- 0.0929). The IRCCS has an intermediate efficiency score (-0.212766 +/- 0.0692), while the base level has the lowest score (-0.288641 +/- 0.0820).

These results suggest that second-level hospitals and private nursing homes are the most efficient, while base level hospitals are the least efficient. Additionally, IRCCS have an intermediate efficiency score. The graphical representation of the PTE scores by hospital level is provided by the scatter plot widget of Orange and shown in figure 16. The hospital units depicted in the scatter plot have a size proportional to the PTE efficiency score; units represented by larger spheres will have a higher PTE efficiency score, while those with smaller spheres will have a lower score.

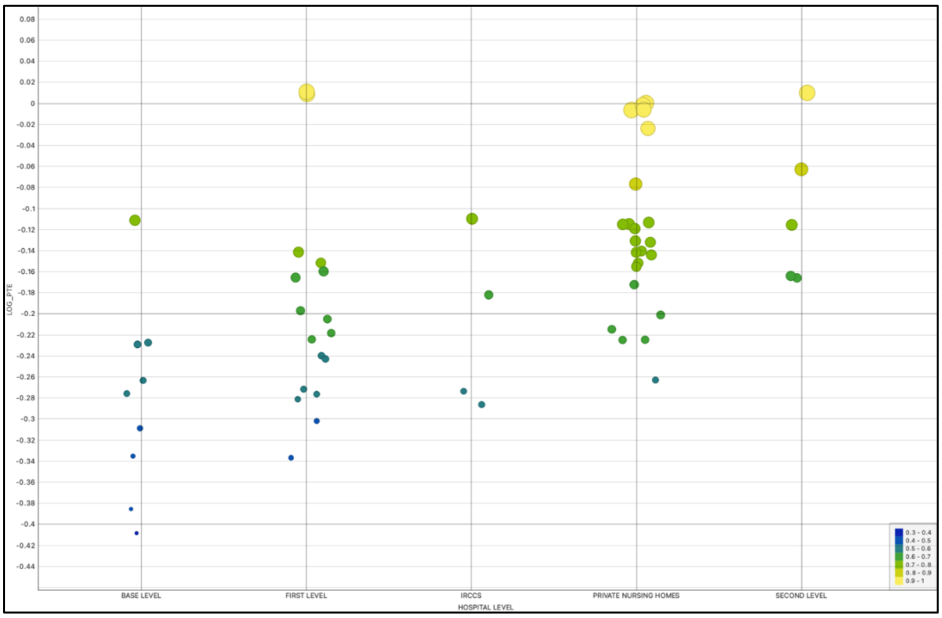


Figure 16 Efficiency scores scatter plot for hospital level.

In general, these analyses have been useful in providing an objective evaluation of hospital efficiency in the Apulia region and in identifying the causes of any differences in efficiency between different hospitals levels. This information can be useful for making informed decisions about the management and allocation of resources within hospital networks and for improving the quality of hospital care in the region.

3.23 ANOVA ANALYSIS ALGORITHM: CPDA MODEL

Objective: Assess the differences in efficiency scores among different hospital levels using the ANOVA technique.

1. Initialization:

- Identify the dependent variable, which in this case is $PTE_{normalized}$;
- Identify the grouping variable, which is the hospital level.

2. Assumption Checks:

2.1. Homogeneity of Variances:

- Perform Levene's Test: If $p > 0.005$, variances are assumed to be equal across groups.
- Perform Bartlett's Test: If $p > 0.005$, variances are assumed to be equal across groups.

2.2. Normality:

For each group, assess the distribution of the dependent variable using:

- Shapiro-Wilk Test
- Kolmogorov-Smirnov Test
- Anderson-Darling Test

If $p > 0.005$ for these tests, the distribution is assumed to be normal.

3. ANOVA:

- Perform one-way ANOVA using the formula:

$$F = \frac{Between - group\ variability}{Within - group\ variability} \tag{18}$$

If $p < 0.005$, there is a statistically significant difference in the mean efficiency scores among the hospital levels.

4. Post-hoc Test:

- If there are significant differences from the ANOVA, perform post-hoc tests (like Tukey's HSD) to pinpoint which groups differ from each other.

The various steps of this algorithm are illustrated in Figure 17 and were performed using the statistical software Jamovi.

The ANOVA analysis was conducted to investigate the differences in normalized PTE efficiency scores across different hospital levels. The assumptions of homogeneity of variances and normality were met, as indicated by the Levene's test ($p = 0.869$), Bartlett's test ($p = 0.869$), and the Shapiro-Wilk ($p = 0.080$), Kolmogorov-Smirnov ($p = 0.652$), and Anderson-Darling ($p = 0.113$) tests, respectively. By examining the Q-Q Plot as part of the assumption's verification process, a more robust and visual validation of the residuals' normality is ensured.

If the points on the QQ-Plot follow a straight diagonal line, it indicates that the residuals are normally distributed.
The overall ANOVA was significant, indicating differences in efficiency scores among hospital levels. Post-hoc comparisons using Tukey's HSD test indicated that the base level differed significantly from private nursing homes ($p_{TUKEY} < 0.01$) and second level ($p = 0.02$). Additionally, the first level was significantly different from private nursing homes ($p_{TUKEY} = 0.041$).

ANOVA

ANOVA - PTE_normalized

	Sum of Squares	df	Mean Square	F	p
HOSPITAL LEVEL	0.219	4	0.05477	7.86	<.001
Residuals	0.376	54	0.00697		

Assumption Checks

Homogeneity of Variances Tests

	Statistic	df	df2	p
Levene's	0.311	4	54	0.869
Bartlett's	1.25	4		0.869

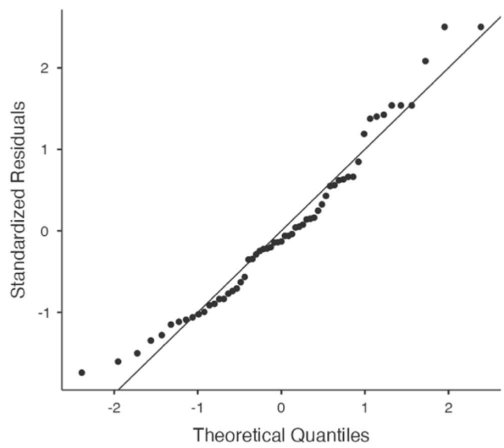
Note. Additional results provided by *moretests*

Normality tests

	statistic	p
Shapiro-Wilk	0.964	0.080
Kolmogorov-Smirnov	0.0957	0.652
Anderson-Darling	0.601	0.113

Note. Additional results provided by *moretests*

Q-Q Plot



Post Hoc Tests

Post Hoc Comparisons - HOSPITAL LEVEL

Comparison							
HOSPITAL LEVEL		HOSPITAL LEVEL	Mean Difference	SE	df	t	p_{tukey}
BASE LEVEL	-	FIRST LEVEL	-0.0860	0.0344	54.0	-2.499	0.106
	-	IRCCS	-0.0759	0.0502	54.0	-1.512	0.559
	-	PRIVATE NURSING HOMES	-0.1629	0.0326	54.0	-4.992	<.001
	-	SECOND LEVEL	-0.1859	0.0466	54.0	-3.992	0.002
FIRST LEVEL	-	IRCCS	0.0101	0.0464	54.0	0.218	0.999
	-	PRIVATE NURSING HOMES	-0.0769	0.0265	54.0	-2.906	0.041
	-	SECOND LEVEL	-0.0999	0.0425	54.0	-2.352	0.145
IRCCS	-	PRIVATE NURSING HOMES	-0.0870	0.0451	54.0	-1.930	0.314
	-	SECOND LEVEL	-0.1100	0.0560	54.0	-1.965	0.297
PRIVATE NURSING HOMES	-	SECOND LEVEL	-0.0230	0.0410	54.0	-0.561	0.980

Note. Comparisons are based on estimated marginal means

Figure 17 Evaluation of efficiency differences for different hospital levels.

4. DISCUSSION AND RESULTS

The discussion of the results obtained from the CPDA methodology applied to the 59 hospitals in Apulia is divided into three sections corresponding to the proposed research questions and one section containing the limitations.

4.1 RESEARCH QUESTION ONE

The CLUSTER-PCA-DEA analysis utilized one input and one output, and the graphical representation can be found in Figure 18. The efficiency frontier encloses the inefficient units and displays the relative efficiency of each hospital with a red square. The yellow squares on the frontier indicate better performance than the DMUs located below it. DMUs that are on the frontier are deemed 100% efficient, whereas those below it are relatively less efficient, as evident from the efficiency score expressed in Table 2.

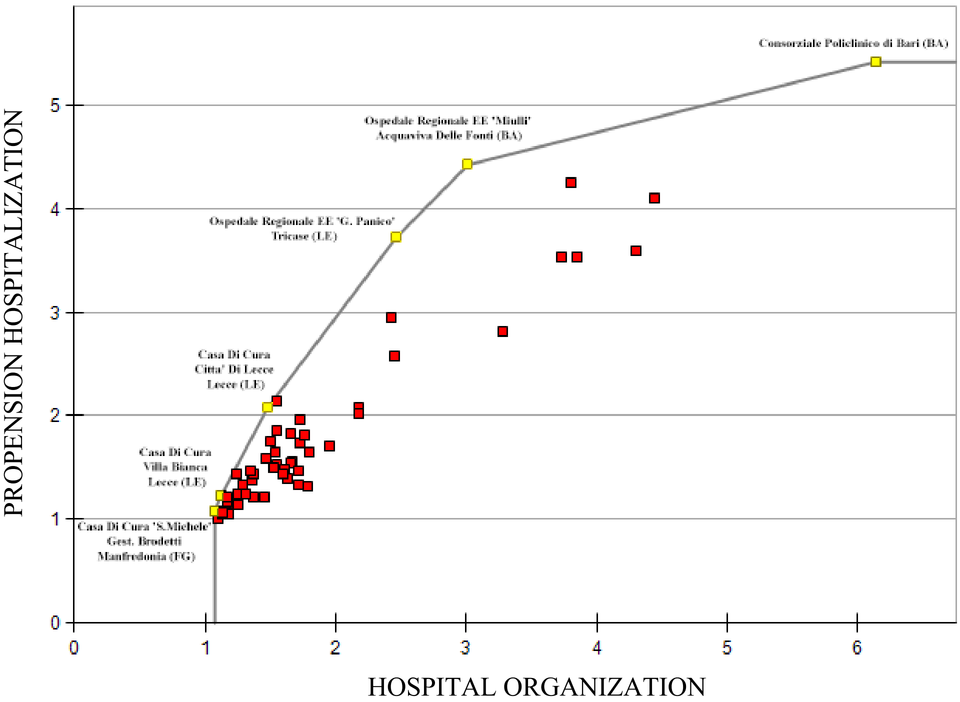


Figure 18 Efficiency frontier of the PCA-DEA model, output oriented with variable returns to scale (VRS).

The graph in Figure 18 was created using the PIM-DEA software (Emrouznejad & Thanassoulis, 2013). The model optimizes the output and therefore the propensity for patient hospitalization, who will pursue the best care based on perceived quality. The results of the efficiency scores assigned to the 59 hospital facilities in Apulia, using the output-oriented DEA model with variable returns to scale, indicate that six hospitals are efficient in terms of perceived quality by residents of Apulia. The scale of values assigned to all hospitals is presented in Table 4, column C-PCA-DEA (CPD).

The efficient facilities are as follows:

- 3 private nursing home: Casa Di Cura Citta' di Lecce (LE), Villa Bianca Nursing Home in Lecce (LE), and Casa Di Cura 'S.Michele' Gest. Brodetti Manfredonia (FG);
- 2 first level hospital: Regional EE Hospital 'Miulli' in Acquaviva delle Fonti (BA), EE Regional Hospital 'G. Panico' of Tricase (LE);

● 1 second level hospital: Bari Polyclinic Consortium Hospital (BA).
The use of cluster analysis followed by PCA and the utilization of principal components as inputs and outputs in the DEA model contribute to the overall robustness of the efficiency results. The obtained efficiency scores can be considered robust and reliable. This approach provides a solid assessment of hospital efficiency in relation to organizational structure and admission propensity.
After evaluating and quantifying the efficiency of Apulian hospitals, we wanted to investigate the causes of inefficiencies. To do so, we used Cooper et al.'s formula (2007), recently applied by Hajiagha et al. (2023), to measure the performance of public hospitals in Iran. Specifically, we decomposed technical efficiency into pure technical efficiency and scale efficiency to understand the inefficiency resulting from the environment.
The formula used:

$$TE = PTE \times SE \tag{19}$$

represents a decomposition of a firm's total efficiency (TE) into two components: productive efficiency (PTE) and allocative efficiency (SE).
Productive efficiency (PTE) measures the firm's ability to use its inputs efficiently to produce the desired outputs, i.e., how close the firm is to the efficient frontier. The PTE values are the efficiency scores assigned to the 59 Apulian hospital facilities using output-oriented DEA analysis with variable returns to scale, expressed as percentages in Table 4.
Allocative efficiency (SE), on the other hand, measures the firm's ability to allocate its inputs efficiently among different productive activities, i.e., how close the firm is to the optimal input combination for producing the desired outputs. Scale Efficiency measures the degree of optimality in the size of decision-making units in relation to production. In the context of DEA, it can be calculated as the ratio of Output-Oriented CRS (Constant Returns to Scale) Efficiency to Output-Oriented VRS (Variable Returns to Scale) Efficiency.
The trend graph and the related comparison of efficiency components are shown in Figure 19.

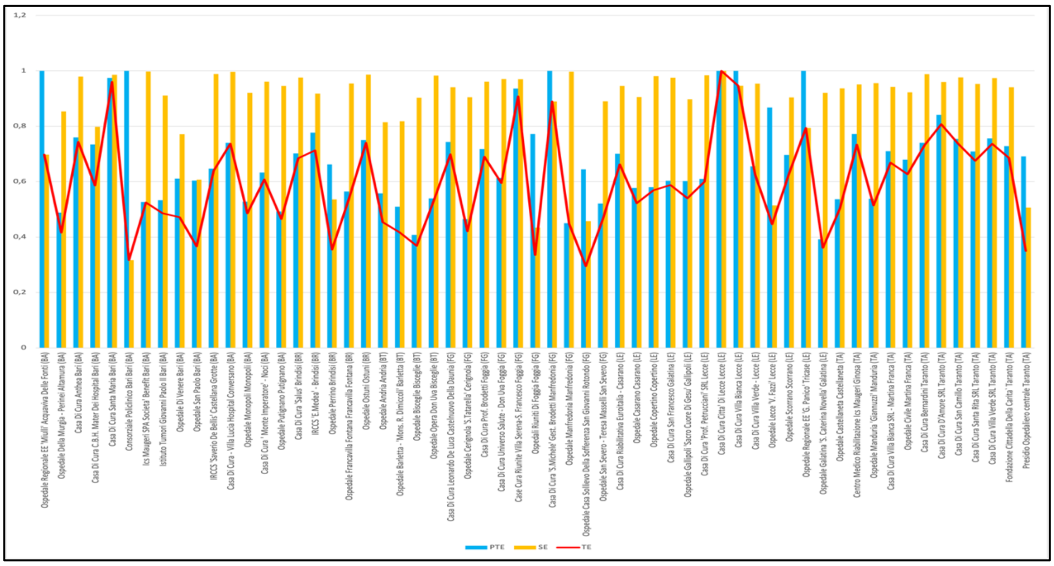


Figure 19 Technical efficiency decomposition.

The Città di Lecce (LE) Clinic has been identified as a highly efficient Data Envelopment Analysis (DEA) unit, in terms of technical efficiency (TE).
The hospitals, the Regional EE 'Miulli' of Acquaviva delle Fonti (BA), the Policlinico Consortium of Bari (BA), the 'S.Michele' Gest. Brodetti Manfredonia (FG) Clinic, the 'G. Panico' Hospital of Tricase (LE), and the Villa Bianca Clinic of Lecce (LE) have been identified as efficient in terms of productive output and thus pure technical efficiency (PTE), but they exhibit technical inefficiencies related to the surrounding environment (SE).

4.1.1 PERCEIVED QUALITY INFLUENCED BY HOSPITAL EFFICIENCY

The relationship between the perceived quality of patients, as measured by their propensity to be admitted, and the efficiency of a hospital is a crucial factor to consider when assessing the effectiveness of a hospital's organizational structure and inpatient practices.
This correlation can provide insights into the efficiency with which a hospital is able to meet patients' needs and expectations, as well as potential disparities between the perceived quality of care and the actual efficiency in providing it. Understanding this relationship can help identify potential areas for improvement in hospital management and resource allocation, ultimately leading to better outcomes and patient satisfaction. Furthermore, understanding the link between hospital efficiency and perceived quality of care can also have significant policy implications.
This study aimed to investigate the influence of the identified Apulian hospital efficiency on the propensity to hospitalization of resident patients.
The linear regression algorithm in a machine learning environment was used to analyze this influence.
The propensity to hospitalize was identified as the target variable of the regression model and the hospital efficiency of scale (SE) and pure hospital technical efficiency (PTE) as features.
Before applying the linear regression, the variables were normalized in the interval [0,1], using the "continuize" widget of the orange software. By normalizing, or scaling, the variables to a similar range, to eliminate potential biases that can arise from differences in measurement units or scales. This process enhances the model's ability to accurately capture relationships between variables, as the regression algorithm can more effectively compare and weigh their impacts. Subsequently, outliers were removed. We assessed the Spearman correlation coefficient between the target variable and the two features, which are not normally distributed, both for the entire Apulian hospital network, for private and public hospital networks and for private nursing home, using the "correlation" widget of the orange software. The results are presented in the table 10.

Table 10. Spearman correlation coefficient between the target variable and the two features.

NETWORK	TARGET VARIABLE	PTE	SE
APULIAN PUBLIC HOSPITAL NETWORK	PROPENSION HOSPITALIZATION	+ 0.709	- 0.851
APULIAN PRIVATE HOSPITAL NETWORK	PROPENSION HOSPITALIZATION	+ 0.369	+ 0.180
PRIVATE NURSING HOME	PROPENSION HOSPITALIZATION	+ 0.507	+ 0.743

Linear regression is a statistical model that attempts to establish a linear relationship between a dependent variable (target) and one or more independent variables (features).

The linear regression model produces a linear function that attempts to predict the value of the dependent variable based on the values of the independent variables. The model in multiple linear regression consists of more than one predictor variable:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P + \varepsilon \quad (20)$$

where Y is the response variable, $X_1; X_2; \dots X_P$ is the predictor variables with p as the number of variables, $\beta_0;\ \beta_1;\ \beta_2 \cdots \beta_P$ are the regression coefficients, and ε is an error to account for the discrepancy between predicted data and the observed data. The Linear regression widget is used to provide the prediction algorithm with the dataset containing the variables to be analyzed, and the performance of the model is evaluated using the Test and scores widget with a cross-validation of 10 folds (Santamato et al. 2023). The results of the evaluation are described in table 11.

Table 11. Performances of Linear regression models.

NETWORK	MSE	RMSE	MAE	R2
APULIAN PUBLIC HOSPITAL NETWORK	0.096	0.310	0.209	0.941
APULIAN PRIVATE HOSPITAL NETWORK	3.406	1.845	1.026	-2.064
PRIVATE NURSING HOME	0.045	0.212	0.175	0.930

FOR APULIAN PUBLIC HOSPITAL NETWORK:

The regression coefficients indicate how both the scale efficiency (SE) and the pure technical efficiency (PTE) influence the likelihood of hospitalization. Specifically, a negative coefficient for scale efficiency (-5.77018) suggests that an increase in scale efficiency is associated with a decrease in the likelihood of hospitalization, while a positive coefficient for pure technical efficiency (2.46398) indicates that an improvement in technical efficiency is correlated with an increase in the likelihood of hospitalization.
The model's performance metrics, calculated using 10-fold cross-validation, provide information about the goodness of fit of the model to the data:
The intercept term of 7.34619 represents the expected value of the target variable when all predictor variables (SE and PTE) are zero.
In Figure 20, the relationships between hospitalization propensity and efficiency measures, PTE and SE, for the public hospital network are depicted. The regression line highlights the overall trend between these two variables. The size of each point represents the hospitalization propensity of the corresponding hospital, with larger points indicating a higher propensity.
The chart shows that as the hospitalization propensity in the public network increases, there is an increase in pure technical efficiency (PTE) and a decrease in scale allocative efficiency (SE).

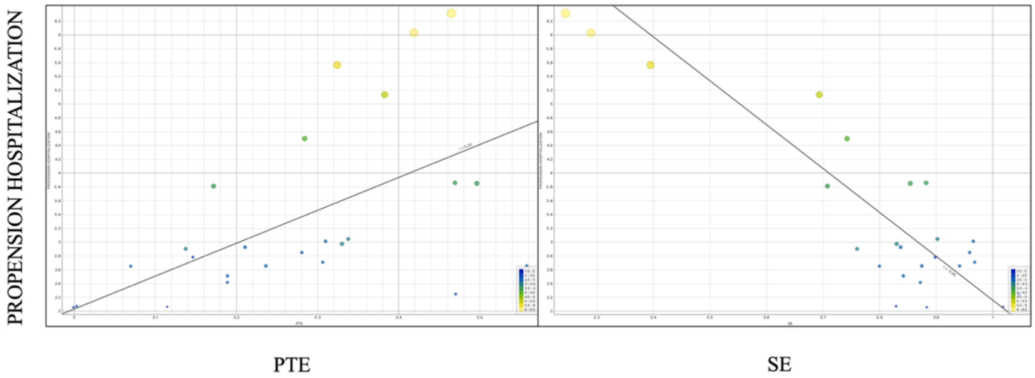


Figure 20 Relationship between Hospitalization Propensity and Efficiency Measures (PTE and SE) for the Public Hospital Network.

Mean Squared Error (MSE): 0.209 - This value represents the average of the squared differences between the values predicted by the model and the actual values of the hospitalization likelihood. A lower MSE indicates that the model has good predictive accuracy.
Root Mean Squared Error (RMSE): 0.310 - The RMSE is the square root of the MSE and provides an estimate of the average error between the model's predictions and the actual values. A smaller RMSE indicates greater model accuracy.
Mean Absolute Error (MAE): 0.096 - This value represents the average of the absolute differences between the model's predictions and the actual values. The MAE measures the average deviation between predictions and actual data.
R-squared (R2): 0.941 - The R2 represents the proportion of the variance in the target data that the model can explain. In this case, a value very close to 1 (0.94) indicates that the model effectively explains the variation in hospitalization likelihood using the predictive variables.
In general, the utilization of 10-fold cross-validation along with the very low error metrics (MSE, RMSE, and MAE) and the high R2 suggests that the model fits the data well and accurately explains the variation in hospitalization likelihood.

FOR APULIAN PRIVATE HOSPITAL NETWORK:
The model associated with the Apulian Private Hospital Network demonstrates poorer performance. The higher values of MSE, RMSE, and MAE (3.406, 1.845, and 1.026 respectively) indicate larger prediction errors compared to the other networks. Additionally, the negative R2 value of -2.064 suggests that this model does not fit the data well and might not effectively capture the underlying relationships.
FOR PRIVATE NURSING HOME:
The fourth linear regression model, focusing on Private Nursing Homes within the Private Hospital Network, stands out for its excellent performance. The notably low values of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) — recorded respectively as 0.045, 0.212, and 0.175 — highlight the model's high predictive accuracy regarding hospitalization propensity. Furthermore, a remarkable R-squared (R2) value of 0.930 suggests that the model effectively explains 93% of the variation in hospitalization propensity using the predictor variables.
The model's coefficients provide further insights: the intercept is -19.8463, representing the baseline hospitalization propensity when all other variable values

are zero. Scale Efficiency (SE) has a coefficient of 21.5713, suggesting its influence on hospitalization propensity, while the coefficient for Pure Technical Efficiency (PTE) is 2.85371. These coefficients suggest that both types of efficiency significantly impact the hospitalization propensity within Private Nursing Homes.
In Figure 21, the relationships between hospitalization propensity and efficiency measures, PTE and SE, for private nursing home are depicted. The size of each point represents the hospitalization propensity of the corresponding hospital, with larger points indicating a higher propensity.
The chart shows that as the hospitalization propensity in the public network increases, there is an increase in pure technical efficiency (PTE) and in scale allocative efficiency (SE).

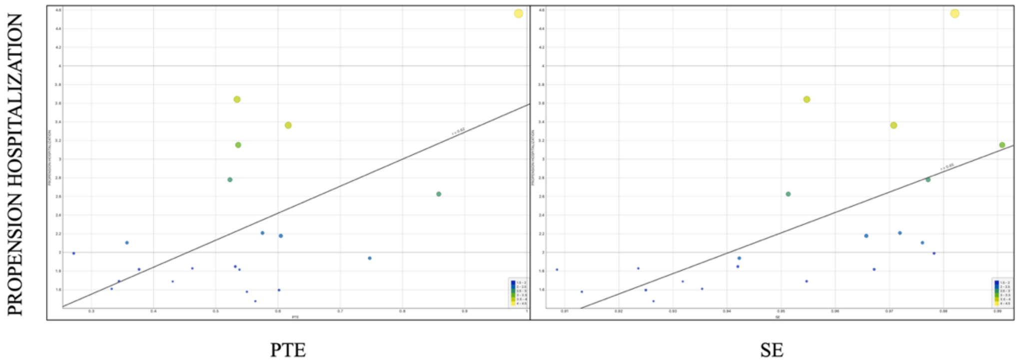


Figure 21 Relationship between Hospitalization Propensity and Efficiency Measures (PTE and SE) for the Private Nursing Homes.

Drawing upon the data provided in the contingency table (Tab. 2), we observe that the Apulian public hospital network, combined with the private nursing homes, constitutes 88.2% of the total health system in the region. This percentage is substantial and accentuates the paramount role of efficiency in determining hospitalization propensity within the Apulian healthcare sector.
The correlation witnessed between efficiency metrics and hospitalization propensity indicates a direct interplay between the operational performances of the institutions and patient choice. In an increasingly resource-optimized and quality-driven healthcare landscape, efficiency stands out as a pivotal factor. Its capability to influence hospitalization propensity suggests that policies and practices aimed at bolstering efficiency could translate into tangible benefits, not only in terms of hospital management but also in the perceptions and decisions of patients.

4.2 RESEARCH QUESTION TWO

The results of the analysis, obtained using the Data Envelopment Analysis (DEA) methodology with variable returns to scale, assessed the efficiency of hospitals considering organizational structure and workforce as inputs, and patient admission rate and perceived quality as outputs. The results reveal a significant difference in efficiency among different hospital levels, indicating that the level of specialization and types of services provided by hospitals impact overall efficiency.
Hospitals at the second level, offering more complex services such as specialized surgery and intensive care, achieved the highest efficiency scores. This suggests that these hospitals effectively utilize available resources to provide high-quality care to patients. Accredited private healthcare facilities, operating in the private sector and accredited by the National Health Service, also demonstrated high

efficiency scores. This implies that they efficiently deliver care and rehabilitation services while maintaining high quality standards.

Institutes of Scientific Research and Healthcare (IRCCS), specialized in scientific research and highly specialized healthcare, obtained intermediate efficiency scores. This highlights their significant role in medical research and the development of new therapies, effectively combining research and patient care.

First-level hospitals achieved intermediate efficiency scores, while basic-level hospitals showed the lowest efficiency scores. This indicates that first-level hospitals provide basic care relatively efficiently, while basic-level hospitals, primarily performing primary care functions, could benefit from improving their efficiency to enhance service quality.

The decomposition of technical efficiency (TE) into pure technical efficiency (PTE) and allocative efficiency (SE) allows identifying the sources of inefficiency within a firm. For example, if total efficiency is low but productive efficiency is high, it means that the firm is producing desired outputs efficiently but not utilizing inputs optimally. In this case, the firm should focus on allocative efficiency to improve its overall efficiency. Conversely, if productive efficiency is low but allocative efficiency is high, it means that the firm is using inputs efficiently but not producing desired outputs efficiently. In this case, the firm should focus on productive efficiency to improve its overall efficiency.

The results of pure technical efficiency (PTE) reveal significant differences among hospital groups. Second-level hospitals are the most efficient, followed by private RSAs and IRCCS, while basic-level hospitals show the lowest score. These results indicate that the specialization of services influences the overall efficiency of hospitals. Accredited private healthcare facilities demonstrate high efficiency in providing care and rehabilitation services, while IRCCS obtain intermediate scores, highlighting their role in medical research. Improving hospital efficiency requires strategies such as service specialization, resource optimization, and efficient allocation.

These considerations underscore the importance of adopting strategies to enhance hospital efficiency and ensure better quality of care for patients.

The ANOVA analysis revealed that second-level hospitals have a relatively lower level of scale efficiency (SE) compared to other categories. These findings indicate the need to focus more on resource optimization and implement specific interventions to improve the performance of these hospitals. It may be beneficial to carefully examine operational practices, human resource management, and the adoption of advanced technologies to identify the underlying causes of this inefficiency. Through accurate assessment and continuous monitoring, opportunities for improvement can be identified, and targeted strategies can be developed to increase the scale efficiency of second-level hospitals. This will optimize the use of resources and ensure the delivery of effective and efficient care at the hospital level. Regarding allocative efficiency (SE), the ANOVA analysis reveals significant differences among different hospital levels (Figure 22).

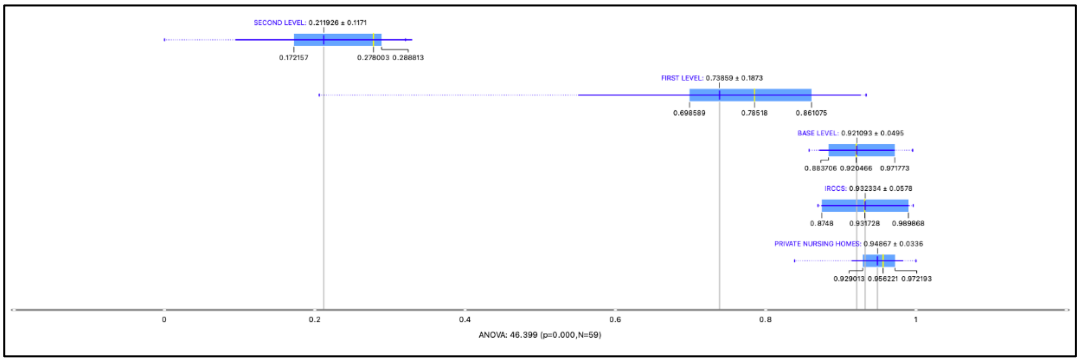


Figure 22 ANOVA analysis by hospital level (SE efficiency).

The results indicate that basic-level hospitals, IRCCS, and private nursing homes have higher scale efficiency (SE) scores, while second-level hospitals show relatively lower scores. This suggests that basic-level hospitals, IRCCS, and private nursing homes are achieving cost and productivity advantages through appropriate resource size, while second-level hospitals may have opportunities for improvement to optimize resource utilization and increase scale efficiency. The implications of this may include reviewing resource allocation strategies and implementing targeted interventions to improve the scale efficiency of second-level hospitals.

Considering the results of pure technical efficiency (PTE) and scale efficiency (SE), some deductions and implications for hospital management can be drawn:

Second-level hospitals may be characterized by relatively low scale efficiency but have shown good efficient productivity of output. This suggests that despite suboptimal resource size, they are able to provide a high perceived quality of care. Consequently, it is possible that second-level hospitals have implemented management processes and targeted care strategies to ensure superior quality in healthcare delivery.

Basic-level hospitals, IRCCS, and private nursing homes have demonstrated both higher scale efficiency and efficient productivity of output. This implies that these facilities have an optimal resource size and are able to provide a high perceived quality of care. These results suggest that further investment in resources for these facilities could lead to additional improvements in efficiency and the quality of healthcare services provided.

Based on these deductions, some implications for hospital management could be:

Consider specific interventions to improve the scale efficiency of second-level hospitals in order to optimize resource utilization and improve overall efficiency.

Allocate additional resources to basic-level hospitals, IRCCS, and private nursing homes to support their already high scale efficiency and promote further improvements in the quality of healthcare services. Identify best practices adopted by second-level hospitals to ensure high perceived quality and consider implementing such strategies in other hospital categories.

In summary, the combined analysis of PTE and SE provides valuable insights for hospital management, enabling the optimization of resource utilization, improvement in efficiency, and the delivery of high perceived quality of care to patients.

4.3 RESEARCH QUESTION THREE

In the ongoing evolution of the field of hospital efficiency analysis, the capability to discern accurately and differentiate performances among various healthcare institutions is paramount. This section aims to methodically assess the discriminatory capacity of the CPDA approach in comparison to traditional DEA models in determining hospital efficiency. Moving forward, we will examine how each individual component of the CPDA framework interacts and contributes to this discriminatory capability, juxtaposing it with the performance of the DEA model. The goal is to delve deeply into understanding how the integration of techniques such as clustering, principal component analysis, and variance analysis might influence and potentially enhance the precision and differentiation capacity in evaluating efficiency. Through a detailed analysis, we intend to outline the nuances and specifics that make CPDA a promising contender for offering a more nuanced and detailed view of hospital efficiency, compared to traditional DEA approaches.

4.3.1 CLUSTER COMPONENT

In the realm of data analysis, clustering techniques serve as powerful tools for grouping data points based on inherent similarities and distinctions. When applied to hospital efficiency analysis, the cluster component can play a pivotal role in enhancing the discriminatory power of the model. It allows for a more nuanced grouping of hospitals, potentially revealing underlying patterns and structures that might be obscured in a holistic analysis. In the context of the CPDA framework, the integration of clustering presents both opportunities and challenges. In this subsection, we delve into a comparative evaluation of the cluster component's strengths and weaknesses in enhancing the discriminatory capacity of both the CPDA and traditional DEA models. The following table outlines the key points of comparison, offering insights into the implications of incorporating clustering in efficiency analysis models (Table 12).

Table 12. Comparison of Cluster Analysis: Implications in the CPDA Model versus Traditional DEA.

ASPECT	DEA	CPDA
STRENGTHS		
Enhanced Discrimination	Uses all variables as given, potentially leading to a wider spread of efficiency scores.	The grouping of correlated variables might lead to a more distinct separation of DMUs based on latent factors or underlying concepts, thereby potentially enhancing the discriminative power.
Reduction of Noise	Any noise or small variations in individual variables directly affect the efficiency scores.	By grouping correlated variables, minor variations or noise in individual variables might be smoothed out, leading to clearer efficiency scores.
Focus on Major Contributors	Efficiency scores are influenced by all variables, even those that might have minor contributions.	Efficiency scores are based on clusters of variables, emphasizing major contributing factors and potentially de-emphasizing minor ones, leading to sharper distinctions between efficient and inefficient DMUs.
WEAKNESSES		
Over-discrimination	With many variables, there's a risk of over-segmenting DMUs based on minor differences.	Clustering reduces this risk by focusing on major variable groups.
Misidentification	Inefficiencies in one or a few variables can heavily influence the overall efficiency score.	By using clusters, the model might mitigate the impact of individual variable inefficiencies, focusing on broader inefficiencies across grouped variables. However, there's a risk of missing nuances related to specific variables.
Interpretation Complexity	Direct interpretation based on individual variables can be simpler.	Interpreting results based on clusters of variables might introduce complexity, as one needs to understand what each cluster represents. However, this can also be seen as a strength as it brings forth the underlying patterns in the data.

While traditional DEA could offer greater granularity using all variables, the use of cluster analysis in CPDA could provide a more targeted and meaningful discrimination based on the main contributing factors.

4.3.2 PCA-DEA COMPONENT

The DEA analysis was conducted on ten input variables and five output variables (table 2), and the numeric results (table 4, column DEA) were represented by a blue line in the graph. Subsequently, DEA analysis was applied to the two principal components identified in the PCA analysis, and the numeric results (table 4, column C-PCA-DEA) were represented by a red line in the graph (figure 23). From the results obtained, it emerged that the CPDA model proposed in this study significantly improved the discriminant ability between the analyzed DMUs (Decision Making Units) (Guede-Cid et al., 2021; Hajiagha et al., 2023). In particular, in DEA analysis, more hospitals were considered efficient with a score of 1, while in the CPDA analysis, only six hospitals out of a total of 59 were deemed efficient.

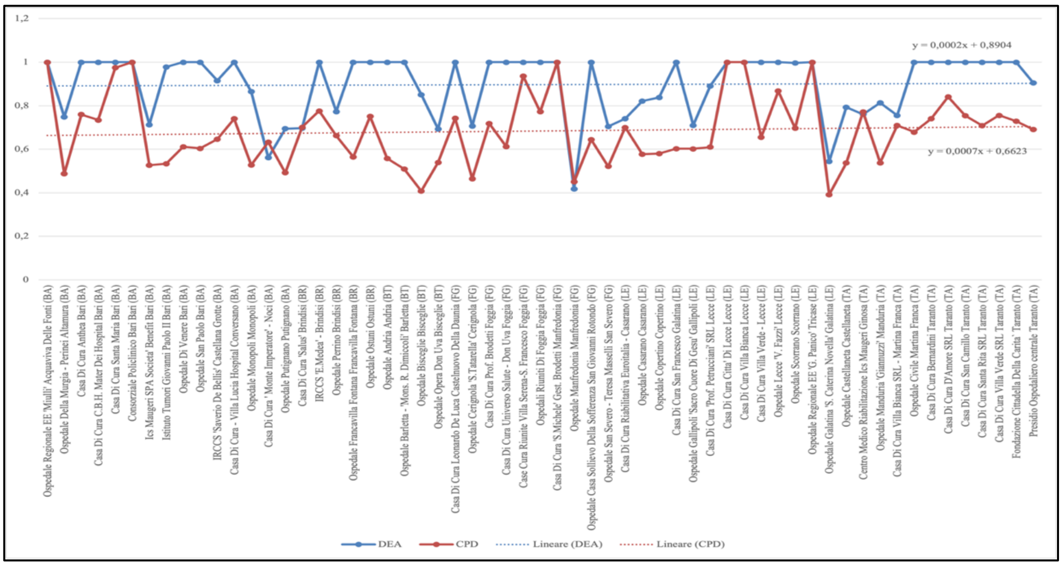


Figure 23 Graphic comparison of DEA and CPDA analysis.

By drawing trend lines for the two curves, represented by blue dashed lines for the DEA analysis on the 15 original variables and red lines for the C-PCA-DEA, we have identified the equations of the lines and their corresponding definite integrals. By calculating the areas under the two lines for values [1;59], we have identified a variation of about 29% in surface units, and therefore in discriminating power in terms of quantitative interpretation (Table 13).

Table 13. Efficiency Score Area Analysis: DEA and CPDA Comparison.

MODEL	TRENDLINE EQUATION	DEFINITE INTEGRAL	∫ [1,59] (y dx) = F(59) - F(1)	PERCENTAGE DIFFERENCE (%)
DEA	y = 0,0002x + 0,8904	$F(x) = 0{,}0001x^2+0{,}8904x+C$	51,9912	23,77%
CPDA	y = 0,0007x + 0,6623	$F(x) = 0{,}00035x^2 + 0{,}6623x + C$	39,6314	

These results demonstrate that applying CLUSTER-PCA to DEA analysis can improve the discriminant capacity among DMUs and more accurately identify efficient structures. Furthermore, this approach can be useful in identifying the most

relevant variables in hospital efficiency analysis and providing useful information for improving efficiency and quality of healthcare services.
We can state that the CPDA methodology has a discriminant power compared to the classical DEA methodology, which is more accurate by about 24%.

4.3.3 ANOVA COMPONENT

In the intricate realm of hospital efficiency analysis, understanding the nuances that differentiate various performance metrics is imperative. Analysis of Variance (ANOVA) stands as a potent statistical tool, specifically designed to analyze differences between group means within a sample. For the scope of our study, ANOVA will be employed to scrutinize efficiency scores against different hospital levels, aiming to unveil specific trends or anomalies.
In addition to ANOVA, a non-parametric analysis was conducted using the Kruskal-Wallis test to compare the PTE scores of the traditional DEA model with the CPDA model. This test was chosen due to its ability to compare means from samples that aren't normally distributed.
It's essential to emphasize that to compare the discriminative power between the two models, the non-parametric ANOVA approach, Kruskal-Wallis, was employed due to the non-normal distribution of PTE for the traditional DEA model. Specifically, the challenge in normalizing a variable that presents values within the interval [0;1], with many values equal to 1, led to this methodological choice. Conversely, the PTE of the CPDA model was previously normalized using a logarithmic transformation. Applying Kruskal-Wallis to the two non-normalized PTEs, we found, in the subsequent post hoc for the PTE of the CPDA model, results consistent with the post hoc of the ANOVA.
The results from the Kruskal-Wallis test (figure 24) are as follows: for PTE of the DEA model, chi-square=9.84, df=4, p=0.043, effect size=0.170. For the PTE of the CPDA model, chi-square=21.66, df=4, p<0.001, effect size=0.374. Further pairwise group comparisons Dwass-Steel-Critchlow-Flinger (DSCF) revealed significant differences between hospital levels for both models. These analyses and subsequent visualizations were meticulously executed using the jamovi statistical software.

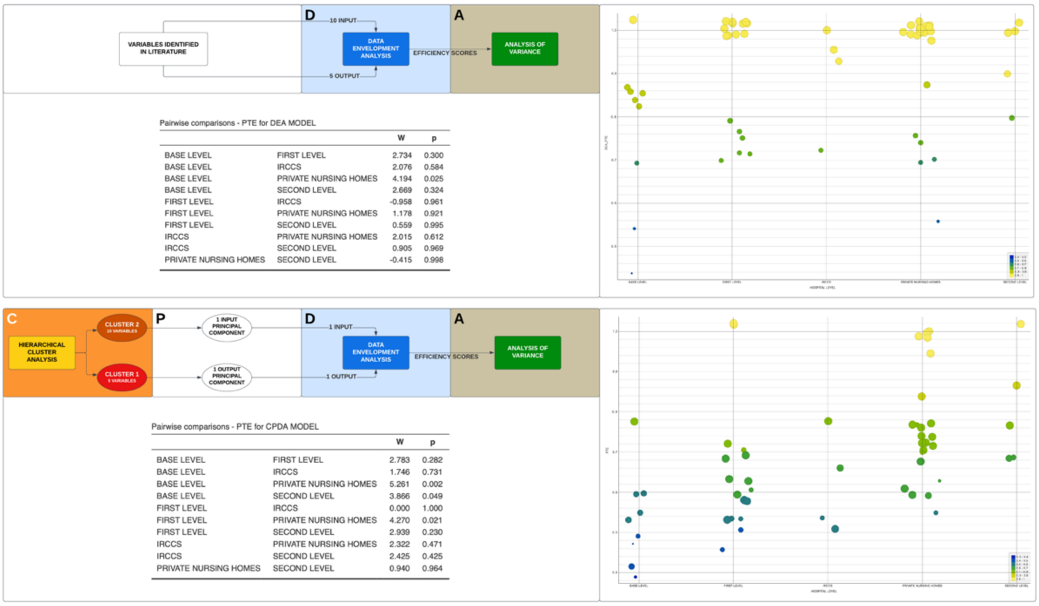
One-Way ANOVA (Non-parametric)

Kruskal-Wallis

	χ^2	df	p	ε^2
PTE for DEA MODEL	9.84	4	0.043	0.170
PTE for CPDA MODEL	21.66	4	<.001	0.374

Figure 24 Kruskal – Wallis analysis.

In Figure 25, the results of the ANOVA analyses for both the CPDA and traditional DEA models are presented. Accompanying scatterplots vividly delineate the efficiency score differences across various hospital levels for each model. These visual representations offer a clear depiction of these discrepancies, underscoring the enhanced discriminatory prowess of the CPDA model in comparison to the traditional DEA. Additionally, within the same figure, post-hoc analyses for both models following the Kruskal-Wallis tests are displayed, further emphasizing the distinctions in hospital efficiencies as gauged by the two models.



Pairwise comparisons - PTE for DEA MODEL

		W	p
BASE LEVEL	FIRST LEVEL	2.734	0.300
BASE LEVEL	IRCCS	2.076	0.584
BASE LEVEL	PRIVATE NURSING HOMES	4.194	0.025
BASE LEVEL	SECOND LEVEL	2.669	0.324
FIRST LEVEL	IRCCS	-0.958	0.961
FIRST LEVEL	PRIVATE NURSING HOMES	1.178	0.921
FIRST LEVEL	SECOND LEVEL	0.559	0.995
IRCCS	PRIVATE NURSING HOMES	2.015	0.612
IRCCS	SECOND LEVEL	0.905	0.969
PRIVATE NURSING HOMES	SECOND LEVEL	-0.415	0.998

Pairwise comparisons - PTE for CPDA MODEL

		W	p
BASE LEVEL	FIRST LEVEL	2.783	0.282
BASE LEVEL	IRCCS	1.746	0.731
BASE LEVEL	PRIVATE NURSING HOMES	5.261	0.002
BASE LEVEL	SECOND LEVEL	3.866	0.049
FIRST LEVEL	IRCCS	0.000	1.000
FIRST LEVEL	PRIVATE NURSING HOMES	4.270	0.021
FIRST LEVEL	SECOND LEVEL	2.939	0.230
IRCCS	PRIVATE NURSING HOMES	2.322	0.471
IRCCS	SECOND LEVEL	2.425	0.425
PRIVATE NURSING HOMES	SECOND LEVEL	0.940	0.964

Figure 25 Comparison of PTE scores between the traditional DEA model and CPDA model using the Kruskal-Wallis test.

These findings, combined with those from the ANOVA, offer a clear view of efficiency differences across different hospital levels and between the two models. The variation in efficiency scores, as determined by the CPDA model and validated by the Kruskal-Wallis test, provides a more nuanced understanding of relative efficiencies across different hospital levels compared to the traditional DEA model, underscoring the superior discriminative capacity of the CPDA model.
The integration of ANOVA and Kruskal-Wallis analyses within the CPDA model reveals enhanced discriminatory capabilities. The combined analysis offers a more realistic perspective on hospital efficiency and can guide interventions and improvement strategies, further emphasizing the superiority of the CPDA model over the traditional DEA in capturing and reflecting nuances in hospital efficiency.

4.3.4.BENCHMARKING ANALYSIS: NEURAL NETWORK PERFORMANCE COMPARISON

In the realm of hospital efficiency analysis, the rise of machine learning techniques has offered new avenues for data interpretation and modeling. Neural networks, in particular, have become an indispensable tool given their prowess in capturing intricate and non-linear relationships within data. We have already seen how the CPDA model exhibits superior discriminatory capability compared to the traditional DEA model when considering all its components.
However, to gain a comprehensive understanding of the CPDA model's potential, it's essential to juxtapose it with advanced predictive techniques, especially neural networks.
Neural networks, inspired by the neural processes of the human brain, stand as one of the cutting-edge frontiers in machine learning. These models can capture and model intricate relationships in data, offering unique insights into the data's nature and underlying structure. Within the scope of our analysis, we employed a neural network to compare the efficiency scores derived from DEA and CPDA models. The target variable chosen for this analysis was the classification of hospitals as "Private" or "Public", while the input variables, or features, were the pure technical

efficiency (PTE) scores and scale efficiency (SE), as derived from the CPDA and DEA models.
To embark on the analysis, we utilized the Orange software. Feature values were first normalized in the [0;1] range using the "Continuize" widget. This normalization ensures that all variables have the same weight and contribution in the neural network model. Instead of omitting outliers, we chose to incorporate them into the analysis, embedding them in the dataset. This approach was adopted to guarantee that the model mirrored the entire range of variations in the data and permitted a fair comparison between models based on the same number of structures.
With the data duly preprocessed, we configured the "Neural Network" widget to set up our neural network:

1. **Structure with 100 neurons in the hidden layer:** This indicates that the neural network has a "hidden layer" composed of 100 neurons (or nodes). A neural network can have multiple layers, and each layer can have a varying number of neurons. These neurons are responsible for capturing and modeling features in the data.
2. **Activation function ReLu:** The ReLu (Rectified Linear Unit) activation function is a function used to determine the output of each neuron. The ReLu function returns the input value if it's positive; otherwise, it returns zero. It's popular in neural networks because it helps prevent some common issues during training, such as the vanishing gradient problem.
3. **Optimizer "Adam":** Adam is an optimization algorithm used to update the neural network weights during training. It's a variant of the stochastic gradient descent algorithm that calculates adaptive estimates of the first and second-order moments. Adam is known for its efficiency and its ability to converge quickly.
4. **Maximum of 200 iterations for training:** This indicates that the neural network will be trained for a maximum of 200 cycles (or "epochs"). In each cycle, the entire dataset is presented to the network, the weights are updated, and the error is computed. Training may terminate before the 200 epochs if a certain convergence threshold is reached or if other early stopping techniques are employed.

Finally, we connected the neural network model to the "Test and Scores" widget configured for stratified ten-fold cross-validation. This approach allowed for an accurate and robust evaluation of the model's performance.

Table 14. Performance Benchmarking Comparison between CPDA and Traditional DEA Models using Neural Networks.

METRIC	DEA	CPDA
AUC	0.547	0.894
CA	0.610	0.847
F1 Score	0.608	0.848
Precision	0.609	0.848
Recall	0.610	0.847

The results presented in the comparison table (Table 14) clearly highlight the superior performance of the CPDA model compared to the traditional DEA model:

1. **AUC (Area Under the ROC Curve):** The AUC is an important indicator to assess the overall performance of a classification model. An AUC value close to 1 indicates an excellent ability of the model to distinguish between classes. The CPDA model has an AUC of 0.901, very close to optimal, whereas the traditional

DEA model has an AUC of only 0.598, which is mediocre. This suggests that the CPDA model has a significantly better ability to correctly classify hospital facilities compared to traditional DEA. A graphical representation of the AUC for the two models is shown in figure 26.

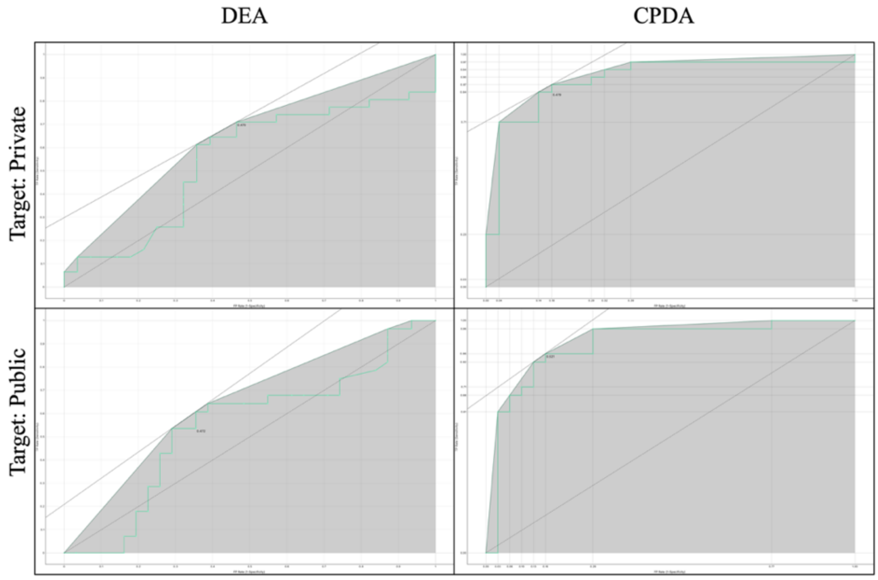
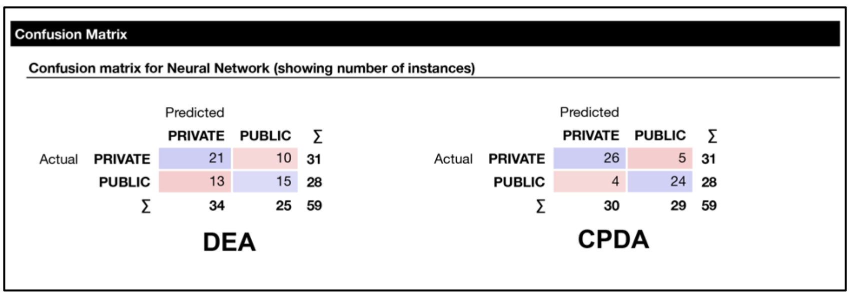


Figure 26 AUC representation for DEA and CPDA.

2. **CA (Accuracy):** The CA represents the proportion of correct predictions relative to the total. The CPDA model has an accuracy of 84.7%, which means it correctly predicted the efficiency of about 85% of the hospital facilities. In contrast, the traditional DEA has an accuracy of 61%, which is significantly lower.
3. **F1 Score:** The F1 score is a measure that combines both precision and recall. A higher F1 score indicates a better balance between precision and recall. The CPDA model has an F1 score of 84.8%, while the traditional DEA stands at 60.8%. This indicates that the CPDA model is much more balanced in its predictions.
4. **Precision:** Precision indicates the proportion of positive identifications that were actually correct. The CPDA model has a precision of 84.8%, compared to 60.9% of the traditional DEA model. This suggests that the CPDA model has a much higher ability to avoid false positives.
5. **Recall:** Recall indicates the proportion of actual positives that were correctly identified. Both the CPDA model and the traditional DEA have similar values for this metric, 84.7% and 61% respectively, but again CPDA prevails.

Confusion matrices were generated for both models, CPDA and DEA (Fig.27):


Confusion Matrix

Confusion matrix for Neural Network (showing number of instances)

DEA

		Predicted PRIVATE	Predicted PUBLIC	∑
Actual	PRIVATE	21	10	31
	PUBLIC	13	15	28
	∑	34	25	59

CPDA

		Predicted PRIVATE	Predicted PUBLIC	∑
Actual	PRIVATE	26	5	31
	PUBLIC	4	24	28
	∑	30	29	59

Figure 27 Confusion Matrixes for DEA and CPDA models.

1. True Positives (private facilities classified as private): 29 for DEA and 26 for CPDA;
2. False Positives (private facilities classified as public): 18 for DEA and 4 for CPDA;
3. False Negatives (public facilities classified as private): 2 for DEA and 5 for CPDA;
4. True Negatives (public facilities classified as public): 10 for DEA and 24 for CPDA.

Overall, the CPDA model makes fewer mistakes compared to the traditional DEA model as highlighted by the confusion matrices.
In summary, when comparing the discriminatory capacity between the CPDA model and the traditional DEA model through the use of neural networks, the CPDA model clearly stands out as the more effective one. While the traditional DEA model offers limited discriminatory capability, as evidenced by the lower benchmark metrics, the CPDA model, which integrates advanced techniques such as cluster analysis, ANOVA, and PCA, exhibits a considerably superior discernment capability. This distinction is particularly evident in the CPDA model's ability to accurately identify the network affiliation of hospital facilities (both private and public). In conclusion, the integrated approach adopted by the CPDA model makes it a much more powerful tool for assessing hospital efficiency compared to the traditional DEA model.

4.3.5. CHOICE OF TARGET VARIABLE AND FEATURES FOR THE NEURAL NETWORK MODEL

The decision to select the "network" as the target variable and "PTE" (Pure Technical Efficiency) and "SE" (Scale Efficiency) as features for the neural network model is rooted in both the intrinsic nature of hospital operations and the objectives of the study.

1. **Relevance to Hospital Operations:**

- Network (Private/Public): The classification of hospitals into private or public networks is fundamental in healthcare research. This distinction often brings with it inherent differences in operational strategies, funding mechanisms, patient demographics, and service delivery models. By analyzing the efficiency measures in the context of this distinction, one can gain insights into the relative performance of the two sectors.
- PTE and SE: These efficiency metrics are fundamental to DEA analysis, providing a holistic view of a hospital's operational performance. PTE measures how well a hospital converts its inputs into outputs, reflecting operational prowess. SE, on the other hand, gauges how optimally a hospital utilizes its size and resources in its operations.

2. **Objective Alignment:** The study aimed to evaluate and compare the efficiency of hospitals in Apulia. The efficiency of a hospital can influence patient choices, policy decisions, and management strategies. By analyzing how PTE and SE influence the classification of hospitals into public or private networks, the study

can provide nuanced insights into the operational strengths and challenges inherent to each sector.

3. **Influence on Neural Network Results:**

- Feature Interplay: Neural networks excel at capturing intricate relationships and interactions between features. By feeding the model with PTE and SE, the network can discern patterns that might be less apparent in traditional statistical models.
- Predictive Power: The combination of PTE and SE as features offers a comprehensive view of a hospital's efficiency. This comprehensive perspective bolsters the model's predictive capabilities, allowing it to more accurately classify hospitals into their respective networks based on efficiency measures.
- Model Interpretability: While neural networks are often considered "black-box" models, the choice of meaningful features like PTE and SE can aid in drawing qualitative insights from the model's results. For instance, if the model consistently misclassifies certain types of hospitals, it might indicate unique operational strategies or external factors influencing those hospitals' efficiency metrics.

The choice of the "network" as the target variable and "PTE" and "SE" as features is both strategic and purposeful. It ensures alignment with the study's objectives while maximizing the neural network's potential to provide meaningful and actionable insights. By leveraging the power of neural networks and the significance of chosen variables, the study can offer robust recommendations for enhancing hospital efficiency in Apulia.

4.3.6. STRATEGIC PARAMETER SELECTION IN ADVANCED HOSPITAL EFFICIENCY ANALYSIS

While the CPDA model remains a solid methodology for hospital efficiency analysis, it could benefit from integrating recent innovations in the field of soft computing. Emerging literature offers a series of advanced techniques that could further strengthen the discriminatory capability and precision of CPDA.

Devi et al. (2022) introduced IRKO (Improved Runge-Kutta Optimization) as a cutting-edge solution for global optimization. Although CPDA mainly focuses on efficiency analysis, integrating optimization techniques like IRKO could improve variable selection and weight determination, ensuring a more accurate and robust efficiency estimation.

On another front, the approach proposed by Gupta et al. (2021) combines the firefly algorithm with genetic techniques, offering an optimized solution for nonlinear optimization problems. This fusion could be used within the CPDA model to further refine the clustering phase, ensuring that hospital structures with similar characteristics are optimally grouped.

Finally, Ghasemi et al. (2022) highlighted the potential of the Circulatory System Based Optimization (CSBO), a biologically-inspired algorithm. Algorithms like this could provide new avenues to model complex interactions between variables within the CPDA framework, offering a more holistic and nuanced view of hospital efficiency.

In the context of advanced analysis of hospital efficiency through the CPDA model, we have conducted a phase of parameter optimization used in various components of the process. This phase was implemented with the aim of improving the accuracy

and reliability of the analysis, while also providing a comprehensive and exhaustive evaluation of hospital efficiency.
The parameter selection began with the Cluster Analysis phase, where we considered different clustering algorithms like DBSCAN, Louvain Clustering, and Hierarchical Clustering. The choice of clustering method and the number of clusters were carefully evaluated in relation to the characteristics of the hospital dataset. The goal was to correctly segment the data to identify patterns of similarity among hospital structures (paragraph 3.6).
In the PCA (Principal Component Analysis) phase, we tackled the question of selecting the optimal number of principal components to retain. This step is crucial for balancing the goal of capturing maximum variance in the data with the need to reduce dimensionality. The objective was to maintain an optimal number of principal components that preserve the essence of the data without introducing noise or redundancy.
In the DEA (Data Envelopment Analysis) phase, a conscious decision was made not to assign weights to the variables. We adopted an output-oriented approach and the VRS (Variable Returns to Scale) model to calculate the efficiency of hospital structures. This choice allowed for the evaluation of efficiency without introducing subjective weights to the variables.
In the ANOVA (Analysis of Variance) phase, we selected the significance level p value to use for evaluating differences between the groups identified in previous phases. The choice of p value was critical in determining whether the observed differences between the groups were statistically significant and thus representative of true disparities among hospital structures.
We will focus on the phase following the Cluster Analysis. In this phase, the focus shifts towards optimizing the weights of the variables identified through the Cluster Analysis. We start from the initial weight assignment, where all variables have an equal weight of 1. This step marks the beginning of the optimization process, where we seek to find the optimal combination of weights that maximizes the efficiency of the CPDA model.
The goal is to maximize the difference between the efficiency values calculated via the CPDA model and those calculated via the traditional DEA model. This not only increases the discriminatory power of the model but also contributes to optimizing the CPDA model as a whole.
We begin with standardizing the variables, an essential step to ensure that all variables are on the same scale and are comparable. This process allows us initially to assign a uniform weight to each variable, as all are now standardized and comparable among themselves. However, assigning uniform weights may not optimally reflect their actual impact on hospital efficiency. Therefore, the goal is to adjust these weights so that they accurately reflect the importance of each variable in the scope of efficiency analysis.
To achieve this goal, we turn to optimization algorithms. We have chosen the Particle Swarm Optimization (PSO) algorithm suitable for optimizing our objective function. The workflow for this optimization phase is illustrated in Figure 28.
During the execution of the Particle Swarm Optimization (PSO) optimization algorithm, several key parameters were recorded that provide insights into the optimization process. Analyzing these parameters can help in better understanding

the course of optimization and the effectiveness of the algorithm in achieving the set objective.
The iterations field of the output object indicates that 21 iterations were performed during the optimization. This value represents the number of times the algorithm updated the particle positions in an attempt to improve the objective function value.
The fun count field provides the total number of objective function evaluations carried out during the optimization. In this case, the algorithm evaluated the objective function 2200 times, exploring different weight combinations to determine which combination minimizes the difference between DEA and CPDA.
The message field contains a termination message for the optimization. In the current execution, the message indicates that the optimization ended because the relative change in the objective value over the last iterations is below a specified threshold. This suggests that the algorithm has reached a stable solution or a stagnation situation.
Finally, the hybrid flag field is empty in this execution, indicating that no form of hybrid optimization was used in addition to PSO.
Overall, the analysis of these parameters provides an overview of the progress of the PSO optimization and its ability to converge to an optimal solution for the given problem.

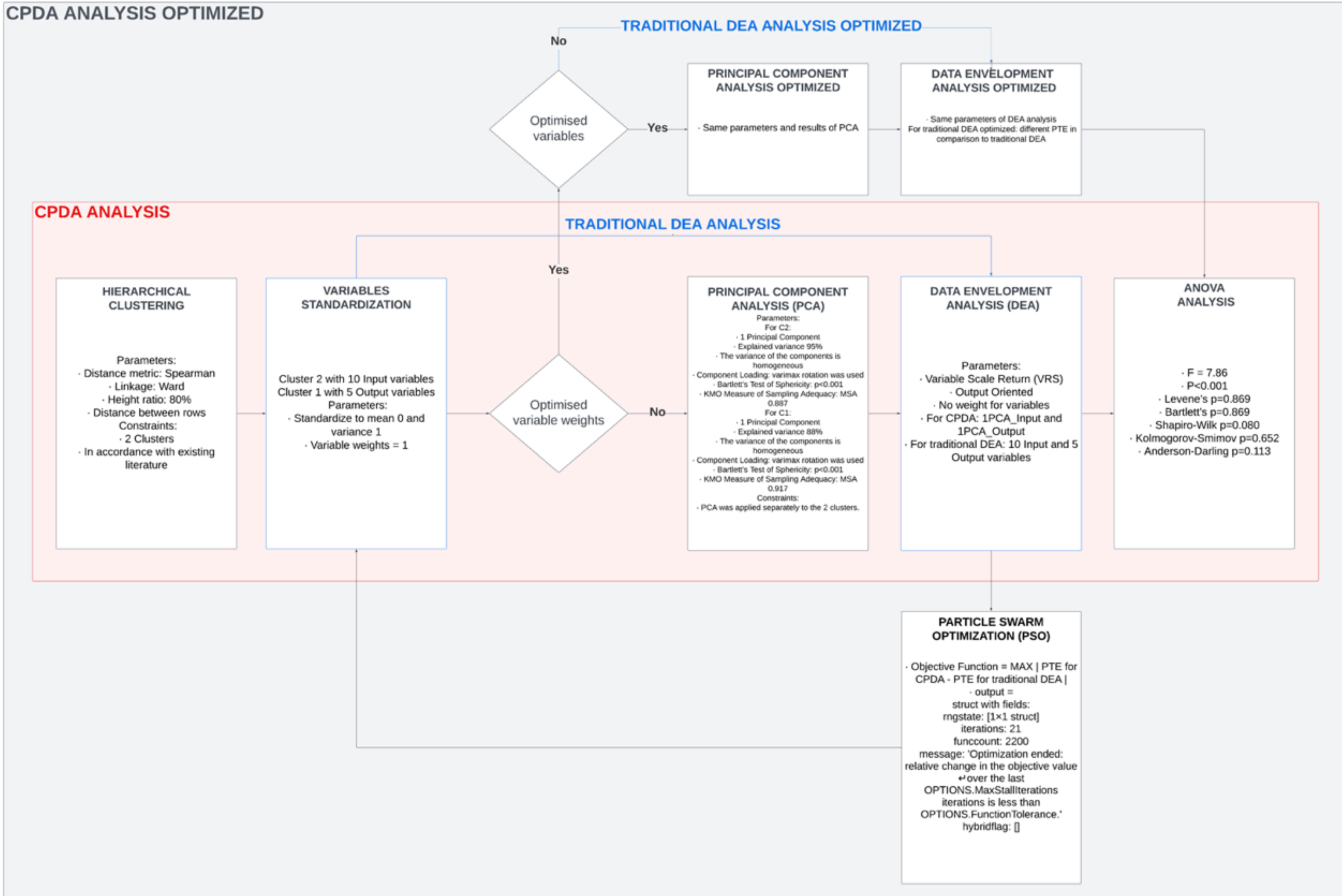


Figure 28. Workflow of CPDA analysis optimized.

The parameters used in our model are as follows:

HIERARCHICAL CLUSTERING:
Parameters:
- Distance metric: Spearman
- Linkage: Ward
- Height ratio: 80%
- Distance between rows

Constraints:
- 2 Clusters
- In accordance with existing literature

VARIABLES STANDARDIZATION
C2: Cluster 2 with 10 Input variables
C1: Cluster 1 with 5 Output variables
Parameters:
- Standardize to $\mu = 0$ e $\sigma^2 = 1$
- Variable weights = 1.
-

PRINCIPAL COMPONENTS ANALYSIS
Parameters:
For C2:
- 1 Principal Component
- Explained variance 95%
- The variance of the components is homogeneous
- Component Loading: varimax rotation was used
- Bartlett's Test of Sphericity: $p<0.001$
- KMO Measure of Sampling Adequacy: MSA 0.887

For C1:
- 1 Principal Component
- Explained variance 88%
- The variance of the components is homogeneous
- Component Loading: varimax rotation was used
- Bartlett's Test of Sphericity: $p<0.001$
- KMO Measure of Sampling Adequacy: MSA 0.917

Constraints:
- PCA was applied separately to the 2 clusters.
-

DATA ENVELOPMENT ANALYSIS
Parameters:
- Variable Scale Return (VRS)
- Output Oriented
- No weight for variables
- For CPDA: 1PCA_Input and 1PCA_Output
- For traditional DEA: 10 Input and 5 Output variables

ANOVA ANALYSIS

- F = 7.86
- P<0.001
- Levene's p=0.869
- Bartlett's p=0.869
- Shapiro-Wilk p=0.080
- Kolmogorov-Smimov p=0.652
- Anderson-Darling p=0.113

PARTICLE SWARM OPTIMIZATION (PSO)

- $FO = \max\ |PTE\ for\ CPDA - PTE\ for\ traditional\ DEA|$
- output =

struct with fields:
rngstate: [1×1 struct]

iterations: 21
funccount: 2200
message: 'Optimization ended: relative change in the objective value ↵over the last OPTIONS.MaxStallIterations iterations is less than OPTIONS.FunctionTolerance.'
hybridflag: []

PRINCIPAL COMPONENTS ANALYSIS OPTIMIZED

- Same parameters and results of PCA

DATA ENVELOPMENT ANALYSIS OPTIMIZED

- Same parameters of DEA analysis
- For traditional DEA optimized: different PTE in comparison to traditional DEA

Objective and Algorithmic Approach

To maximize the difference in efficiency scores between our CPDA model and the traditional DEA model, we explored algorithmic options, specifically the Particle Swarm Optimization (PSO) algorithm. PSO simulates the behavior of a particle swarm, where each particle represents a potential solution. The algorithm is well-suited for complex problems and can converge rapidly to promising solutions.

Mathematical Formulation

The optimization objective can be mathematically expressed as:

$$OF = \max|PTE\ for\ CPDA - PTE\ for\ traditional\ DEA| \tag{21}$$

Where:
PTE for CPDA = Efficiency as expressed by the CPDA model (CPDA value in Table 4).
PTE for traditional DEA = Efficiency as expressed by the traditional DEA model (DEA value in Table 4).
This optimization process aims to identify weight combinations that maximize discrimination between the two methodologies, thereby identifying key variables

for hospital efficiency. The weights assigned to the 10 input and 5 output variables by the optimization algorithms are depicted in Figure 29.

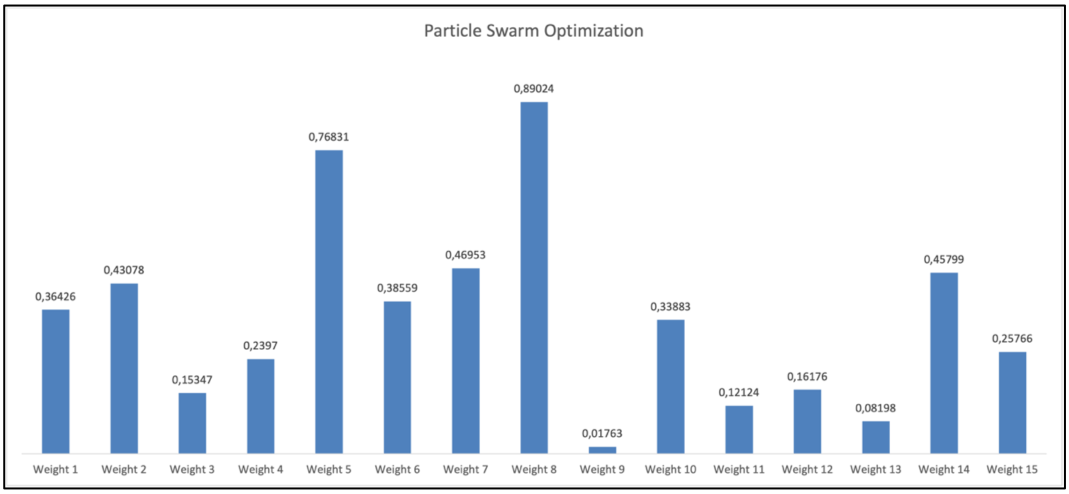


Figure 29 Optimized weights generated by Particle Swarm Optimization algorithm.

Variable Selection and Weighting

We then identified 15 optimized variables by multiplying the standardized variables post-clustering phase by their respective optimized weights. These optimized weights were then applied to the input (1-10) and output variables (11-15).

Data Analysis

PCA was applied to the 10 optimized input variables and the 5 optimized output variables, yielding one principal component for each. Subsequently, we used output-oriented VRS DEA to compute optimized efficiency scores. We also applied traditional DEA analysis on these optimized variables to obtain efficiency scores via the conventional DEA method.

Results and Insights

The average efficiency score difference between the CPDA and traditional DEA models was 0.21, while the average difference using optimized variables with the PSO algorithm was 0.28. This demonstrates that variable optimization leads to more discriminative or sensitive efficiency estimations. The percentage difference between the efficiency scores from the optimized CPDA and traditional DEA models was 29.17%, surpassing the 23.77% identified without algorithmic optimization. From the results obtained, we calculated the CPDA-DEA difference; the results of the average differences are illustrated in Figure 30.

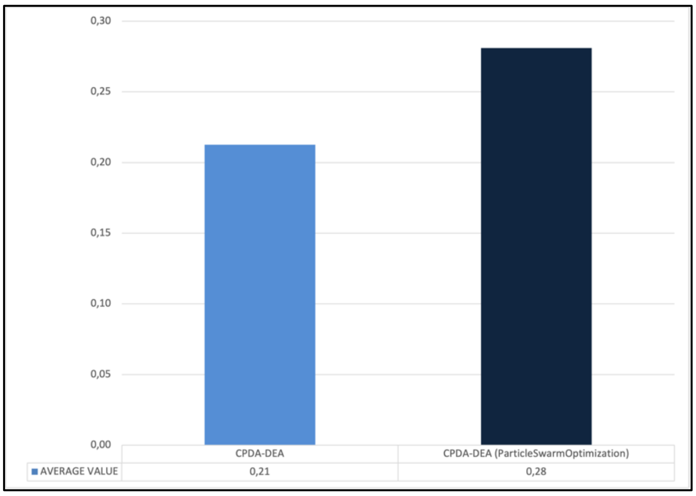


Figure 30 Average differences between CPDA and optimized DEA.

Results and Insights

The average efficiency score difference between the CPDA and traditional DEA models was 0.21, while the average difference using optimized variables with the PSO algorithm was 0.28. This demonstrates that variable optimization leads to more discriminative or sensitive efficiency estimations. The percentage difference between the efficiency scores from the optimized CPDA and traditional DEA models was 29.17%, surpassing the 23.77% identified without algorithmic optimization. The graphical results are illustrated in figure 31 and the numerical results in table 15.

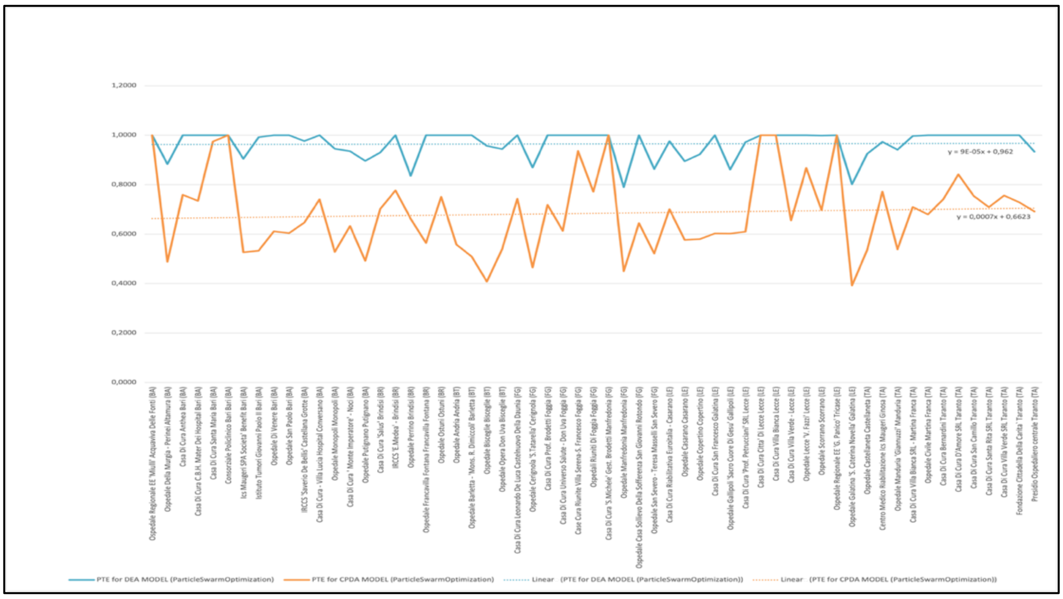


Figure 31 Graphic comparison of DEA and CPDA analysis optimized.

Table 15. Efficiency Score optimized Area Analysis: DEA and CPDA Comparison.

MODEL	TRENDLINE EQUATION	DEFINITE INTEGRAL	∫ [1,59] (y dx) = F(59) - F(1)	PERCENTAGE DIFFERENCE (%)
DEA	y = 9E-05x + 0,962	F(x) = 4,5E-05x^2 + 0,962x + C	55,9526	29,17%
CPDA	y = 0,0007x + 0,6623	F(x) = 0,00035x^2 + 0,6623x + C	39,6314	

Statistical Tests

Non-parametric Kruskal-Wallis tests were conducted both for the optimized traditional DEA and CPDA models. The results, consistent with those from non-optimized models, further confirm the superior discriminatory power of the CPDA model.

The use of Particle Swarm Optimization (PSO) is not limited to simply enhancing the discriminative power of the CPDA model. In addition to optimizing this specific metric, PSO is also used to confirm and validate the efficiency scores generated by the CPDA model itself. Furthermore, the PSO algorithm contributes to optimizing the efficiency scores in the traditional DEA model.
In sum, while the CPDA model already represents a significant advancement in hospital efficiency analysis, integrating soft computing innovations like PSO could offer even more substantial improvements. This positions the CPDA model at the cutting edge of advancements in the field.

4.4 RESPONSES TO I SESSION RESEARCH QUESTIONS

The answers to the research questions proposed in I Session are as follows:
A_1: Using the proposed CPDA methodology in this study, we successfully assessed the efficiency of hospitals in the Apulian region. Standardized efficiency scores ranging from 0 to 1 were identified. Within the public hospital network of Apulia, both pure technical efficiency (PTE) and scale efficiency (SE) proved pivotal in determining hospitalization propensity.
A_2 : The CPDA methodology facilitated the evaluation of hospital efficiency in the Apulian region based on their affiliation and level. The results highlighted superior efficiency in higher-level hospitals.
A_3 : The findings underscored that the CPDA methodology, anchored in machine learning techniques, boasts superior discriminant power compared to traditional DEA models in evaluating hospital efficiency. This implies that the proposed methodology can arm healthcare organizations with invaluable insights to enhance the efficiency and quality of healthcare services.

4.5 LIMITATIONS

Despite the value and innovation brought forth by this study in the realm of hospital efficiency through the adoption of the CPDA methodology, it's imperative to underscore certain intrinsic limitations that might influence the interpretation and generalization of the findings:

- **Sample Size:** The research focused on a specific sample of hospitals in 2020. This temporal and geographic circumscription might not fully reflect the variability and complexity of the entire hospital ecosystem, potentially limiting the generalizability of the outcomes to a broader context.

- **Intraregional Mobility:** The active intraregional mobility variable was gauged using an interpolated path between the patient's city of origin and the hospital city of destination. While pragmatic, this methodology might introduce inaccuracies and fail to capture all patient movement dynamics.
- **Geographical Scope:** Even though the analysis zeroes in on the Apulian region, it's crucial to acknowledge that each region, or country, has its unique peculiarities and challenges. Hence, directly applying these findings to other settings might necessitate methodological adjustments.
- **CPDA Methodology:** While the CPDA model demonstrated promising discriminative capabilities, its intricate nature and the melding of various methods could present challenges in interpretation and practical deployment, especially in settings with constrained resources or expertise.
- **Variables and Data:** As with any study, the quality and completeness of the collected data can influence the outcomes. Even though measures were taken to ensure data accuracy, the absence of certain relevant variables or potential inaccuracies in the data might impact the precision of the analysis.

These limitations provide invaluable directions for future research. It's paramount that subsequent studies consider these challenges, expanding and deepening the CPDA methodology, and exploring its applicability across diverse contexts to ensure a more robust and generalizable analysis of hospital efficiency.
This study presents several pivotal limitations:

- **Scope and Complexity of Investigation:** The primary aim was to introduce and validate the CPDA methodology within the Apulian context. An expansive investigation, incorporating every recommendation, would have broadened the study's scope excessively, potentially diluting the core objective and making the analysis overly intricate.
- **Data Availability:** Access to specific data, or granularity levels required for some of the recommended investigations, might not have been available or easily accessible during the research timeframe.
- **Preliminary Nature of Investigation:** This study serves as an exploratory inquiry introducing the CPDA method. In research, it's strategic to establish a foundational footing before branching into more detailed or nuanced investigations.
- **Depth of Analysis:** To ensure a rigorous understanding and validation of the CPDA methodology, it was crucial to delve deeply into its application in the given context. Expanding the focus to multiple investigations could have sacrificed the depth and rigor of the analysis.

Though these limitations outline the scope and boundaries of our study, they also underscore the potential for further research.

5. CONCLUSION OF I SESSION

This study introduces an innovative methodology, CPDA, which combines Cluster Analysis, Principal Component Analysis (PCA), Data Envelopment Analysis (DEA), and Analysis of Variance (ANOVA) to evaluate hospital efficiency. The machine learning environment is integrated through the use of linear regression and neural network algorithms, while optimization using Particle Swarm Optimization (PSO) further enhances the CPDA model.

Through the application of CPDA to hospital efficiency in the Apulia region, Italy, original findings emerge. Relationships between hospital efficiency and hospitalization propensity are identified, underscoring the importance of efficient resource allocation to enhance care quality. Significant variations in hospital efficiency based on hospital network affiliation and level are highlighted.
The utilization of machine learning algorithms, particularly neural networks, demonstrates CPDA's superior discriminatory power compared to traditional Data Envelopment Analysis (DEA). Neural networks also provide a benchmark for evaluating the CPDA model against DEA, confirming its effectiveness.
Optimization through Particle Swarm Optimization (PSO) enhances the CPDA model in terms of discriminatory power and confirmation of efficiency values.

Further Analyses:

In addition to the conducted analyses, the integration of PSO, optimizing the CPDA model, is noteworthy for its enhanced discriminatory capability and validation of identified efficiency results. Moreover, the incorporation of various clustering algorithms within CPDA and the selection of the best-performing among them represent further methodological improvements.

Future Perspectives:

This study lays the foundation for future research in the realm of hospital efficiency assessment. An intriguing direction is the development of a Decision Support System (DSS) based on CPDA, enabling the practical implementation of this methodology in hospital management decisions.

In conclusion, CPDA proves to be an advanced and promising approach to address challenges in hospital efficiency identification and evaluation. This study paves the way for further research and practical application of CPDA in the field of hospital management, promoting progress in this domain.

SESSION II

COMPARATIVE ANALYSIS OF HOSPITAL SYSTEMS: APULIA AND EMILIA-ROMAGNA

1. INTRODUCTION

Italy, through its National Health Service, is constantly committed to providing high-quality medical care to all its citizens. However, in a country with a regionalized health structure like Italy, there are inevitably variations in performance between different regions. However, the regionalized health structure introduces performance disparities across regions, prompting experts to scrutinize and benchmark regional health systems to discern models of excellence, areas needing improvement, and best practices (Chisari & Lega, 2023).

In the current landscape, our focus is on the hospital systems of two particularly relevant regions: Apulia and Emilia-Romagna. But what are the reasons behind this choice?

Emilia-Romagna, as highlighted by the GIMBE Report 2023, stands out as one of the leading regions in Italy regarding the provision of Essential Levels of Assistance (LEA). Its excellence in offering essential services to its citizens places it as a shining example in the Italian health context. The GIMBE Foundation report underscores the importance of the Guarantee System, an essential tool to ensure quality, appropriateness, and uniformity in the delivery of health services, an area in which Emilia-Romagna excels (GIMBE Foundation, n.d.).

On the other hand, Apulia, despite having a history, demography, and geography that differentiate it, has shown remarkable adaptability and innovation in the healthcare sector, facing specific challenges with determination. The comparison between Emilia-Romagna and Apulia does not simply aim to establish which region "provides better services," but rather to unveil how different contexts, resources, and strategies can shape the efficiency and effectiveness of healthcare services. While Emilia-Romagna can offer solutions based on its consolidated experience, Apulia can present innovative approaches to overcome particular challenges, which could be replicable in other contexts.

A fundamental aspect of our study is the analysis of hospital efficiency based on the quality perceived by the resident patient. For this reason, Apulia and Emilia-Romagna were considered as a single territory, analyzing health mobility within the borders of this ideal macro-region exclusively by residents. This approach allows for a more accurate assessment of the efficiency and quality of healthcare services, taking into account the direct experiences of patients and their choices in terms of healthcare facilities.

The perception of the quality of hospital service by patients emerges as a cornerstone for raising health standards. The analysis of reviews on Facebook, conducted through machine learning techniques, revealed a significant link between hospital accreditation and emotions expressed online, underscoring the importance of careful evaluation of patients' opinions (A Rahim et al., 2021). Simultaneously, the implementation of a targeted conceptual framework has provided hospital administrators with an effective method for examining and enhancing service quality in different hospital environments (Pai et al., 2018). These approaches emphasize the imperative need to integrate patients' perceptions in the path of optimizing healthcare services, ensuring cutting-edge and patient-centered care.

Hospital efficiency in Italy represents a fundamental pillar to ensure a quality health service. In the national panorama, the Hub & Spoke network model emerges as a significant example, highlighting the importance of overcoming organizational barriers to favor effective and constructive change (Rosa, 2018). The issue of transient efficiency is equally central, underlining how short-term strategies can have a considerable impact on the overall improvement of the health system (Colombi et al., 2017). The careful and prudent management of hospital and intensive care beds further contributes to avoiding overcrowding situations, ensuring timely and adequate assistance to all patients (Pecoraro et al., 2020). Finally, the adoption of a Prospective Payment System stands out as a key element to promote hospital efficiency, offering a more agile and sustainable funding model (Cavalieri et al., 2014). In this context, Italy is moving towards a profound renewal, aimed at consolidating and enhancing the efficiency of its health system, to better respond to the needs of all citizens.

In the context of our in-depth examination of perceived quality, the importance of considering the patient's propensity for hospitalization as a revealing indicator clearly emerges. This propensity not only reflects the outcomes and the number of hospitalizations but is closely intertwined with the active kilometric mobility of resident patients. Patient mobility in health systems represents a crucial aspect that directly affects the perceived quality of hospital service. This mobility, an expression of freedom of choice and the search for quality care, manifests itself through the movement of patients between different health facilities, both within and outside regional borders.

Moreover, the challenge of elevating the quality of hospital service is enriched with complex nuances. The importance of considering a multitude of factors, including patient mobility, has been emphasized, outlining a more holistic framework for the continuous improvement of hospital service quality (Rose et al., 2004). The integration of these elements in our examination offers a magnifying lens through which to observe with greater clarity and depth the perceived quality in the context of the hospital systems of Apulia and Emilia-Romagna.

In an era characterized by increasing global interconnection and unprecedented knowledge exchange, it is crucial to analyze the differences and similarities between the hospital systems of different regions. This is not only for improvement at the national level, but also because the lessons learned in Italy could provide valuable insights for other countries and vice versa. Moreover, in light of the COVID-19 pandemic, it is essential to understand how different regions have responded and adapted to a health crisis of this magnitude.

A key element in our comparison between the hospital systems of Apulia and Emilia-Romagna is the adoption of the CPDA methodology. This innovative methodology combines Cluster Analysis, Principal Component Analysis (PCA), Data Envelopment Analysis (DEA), and Analysis of Variance (ANOVA) to evaluate hospital efficiency. This methodology has allowed the identification of relationships between hospital efficiency and the propensity for hospitalization, underscoring the importance of proper resource allocation to improve the quality of care. Additionally, the CPDA highlighted significant variations in hospital efficiency based on affiliation and the level of the hospital network.

The comparison between the hospital systems of Apulia and Emilia-Romagna is not limited to a simple analysis of performance. It offers a holistic view of the dynamics, processes, and strategies that can elevate the quality of care and improve the health of citizens. Through critical analysis and mutual learning, the goal is to build a health system that responds equitably, resiliently, and efficiently to the needs of all.

2. BACKGROUND

2.1 NETWORK AND HOSPITAL FACILITIES IN APULIA AND EMILIA-ROMAGNA: A DETAILED ANALYSIS.

Hospital complexes, entities that can be composed of multiple hospitals, represent a fundamental pillar of the healthcare system. In Emilia-Romagna, the structuring of hospital complexes is an integral part of a healthcare organization aimed at ensuring complete patient care, ensuring continuity of care and socio-health integration. The presence of multiple hospitals within a single hospital complex allows for greater specialization and a more effective distribution of resources and skills, guaranteeing patients access to high-quality care for a wide range of medical and surgical conditions.

The region is divided into various Local Health Units (Aziende USL), each covering specific provincial areas, and there are also four Hospital-University Companies located in Parma, Modena, Bologna, and Ferrara. These companies are further divided into Districts and Territorial and Hospital Departments, ensuring the provision of essential assistance services to the reference population.

The Scientific Hospitalization and Care Institutes (IRCCS) in the region are facilities that offer health services of hospitalization and care together with specific biomedical research activities. The IRCCS are considered multi-zonal hospital complexes of the local health companies, fully integrated into the Regional Health Service and function as reference and excellence centers for assistance, research, and training (Aziende sanitarie, Irccs, Asp, n.d.).

To make an effective comparison between Apulia and Emilia-Romagna in terms of health services, a hypothetical macro-region was conceived. Within this context, the focus was exclusively on the mobility of patients residing in the two regions, considering their movements in hospitals and hospital complexes within the territory of the macro-region itself. This approach allowed excluding from the study the flows of patients coming from other Italian regions, thus offering a clearer and more precise picture of the health situation and patient mobility exclusively between Emilia-Romagna and Apulia. The analysis aims to evaluate the efficiency, accessibility, and quality of the health facilities present, as well as to understand the dynamics of choice and preference of patients in relation to the health services offered by the two regions.

Table 16 details the distribution of these structures in the two regions. In Emilia Romagna, a marked predominance of private hospitals emerges, with 27 base hospitals and 15 second-level hospitals. In contrast, Apulia shows a balanced distribution, with a significant presence of 5 public second-level hospitals.

Table 16. Contingency Table: Relationship between Hospital Network Nature and Hospital Level in the Apulia and Emilia Romagna Region.

			LEVEL					
REGION	**NETWORK**		**BASE LEVEL**	**FIRST LEVEL**	**IRCCS**	**PRIVATE NURSING HOMES**	**SECOND LEVEL**	**Total**
EMILIA ROMAGNA	PRIVATE	Observed	27	0	1	15	3	46
		% of total	39.7 %	0.0 %	1.5 %	22.1 %	4.4 %	67.6 %
	PUBLIC	Observed	1	6	3	0	12	22
		% of total	1.5 %	8.8 %	4.4 %	0.0 %	17.6 %	32.4 %
	Total	Observed	28	6	4	15	15	68
		% of total	41.2 %	8.8 %	5.9 %	22.1 %	22.1 %	100.0 %
APULIA	PRIVATE	Observed	0	4	2	24	0	30
		% of total	0.0 %	6.8 %	3.4 %	40.7 %	0.0 %	50.8 %
	PUBLIC	Observed	9	13	2	0	5	29
		% of total	15.3 %	22.0 %	3.4 %	0.0 %	8.5 %	49.2 %
	Total	Observed	9	17	4	24	5	59
		% of total	15.3 %	28.8 %	6.8 %	40.7 %	8.5 %	100.0 %
Total	PRIVATE	Observed	27	4	3	39	3	76
		% of total	21.3 %	3.1 %	2.4 %	30.7 %	2.4 %	59.8 %
	PUBLIC	Observed	10	19	5	0	17	51
		% of total	7.9 %	15.0 %	3.9 %	0.0 %	13.4 %	40.2 %
	Total	Observed	37	23	8	39	20	127
		% of total	29.1 %	18.1 %	6.3 %	30.7 %	15.7 %	100.0 %

Table 17 presents the results of the χ 2 test, performed to evaluate the differences in the distribution of hospitals and hospital complexes in the two regions and overall. The very low p-value (<.001) indicates that we can reject the null hypothesis, suggesting that there is a significant association between the region and the distribution of hospitals by type and sector. The distribution of hospitals by type and sector is not independent from the region in which they are located.

Table 17. χ 2 test

REGION		**Value**	**df**	**p**
EMILIA ROMAGNA	χ^2	49.2	4	< .001
	N	68		
APULIA	χ^2	42.8	4	< .001
	N	59		
Total	χ^2	64.5	4	< .001
	N	127		

The chi-square test analysis clearly shows that the distribution of hospitals in Emilia-Romagna and Apulia is not uniform among the different categories of type and sector. This suggests that there are significant differences in the distribution of

hospitals between the two regions, with a different distribution by type (First Level, Second Level, Base Level) and sector (Public and Private) in each region.
In Emilia-Romagna, the private hospital network, accredited to the regional health system, is strongly oriented towards basic services, with 27 base-level hospitals out of a total of 46. The presence of four IRCCS (one in the private sector and three in the public sector) and 12 second-level hospitals in the public sector highlights a substantial commitment towards research and specialized care.
In Apulia, the private hospital network, also accredited, is mainly composed of first-level structures (4 out of 30), with no presence of base-level hospitals. This distribution suggests a focus on specialized care in the private sector. However, in the public sector, the presence of five second-level structures underlines a parallel commitment to provide a broad spectrum of health services.
Considering Emilia-Romagna and Apulia as a macro-region, the difference in the distribution of hospital levels becomes evident. While Emilia-Romagna focuses on basic and specialized services, Apulia shows a greater emphasis on second-level structures in the public sector, compensating for the absence of base-level hospitals in the private sector. This inter-regional balance could reflect strategic complementarity, with each region covering different aspects of the population's health needs.
The configuration of the hospital network in Apulia, with a strong presence of second-level structures in the public sector, highlights a commitment to ensuring specialized care, even in the absence of base-level hospitals in the private sector. This may indicate a strategy of focusing resources on specialized and advanced care. However, it is essential to ensure that access to basic care is not compromised, and that there is an adequate geographical distribution of facilities to ensure accessibility for all residents.

2.2 APULIA AND EMILIA-ROMAGNA: ANALYSIS OF THE RELATIONSHIP BETWEEN DOCTORS AND RESIDENTS

The integrated analysis of health resources between Emilia-Romagna and Apulia reveals complex and multifaceted dynamics that affect the distribution of medical staff and access to care for residents. The residents/doctor contingency table is a key tool for analyzing these dynamics, offering a detailed view of the relationship between the number of doctors and the resident population in the different health facilities of the two regions (Table 18).

Table 18. Contingency table: Percentage relationship between the nature of the hospital network and hospital level in the Apulia and Emilia Romagna regions expressed by physician/resident.

			LEVEL					
REGION	NETWORK		BASE LEVEL	FIRST LEVEL	IRCCS	PRIVATE NURSING HOMES	SECOND LEVEL	Total
EMILIA ROMAGNA	PRIVATE	% of total	20.4 %	0.0 %	0.9 %	2.8 %	3.3 %	27.5 %
	PUBLIC	% of total	0.5 %	6.8 %	7.6 %	0.0 %	57.6 %	72.5 %
	Total	% of total	21.0 %	6.8 %	8.5 %	2.8 %	60.9 %	100.0 %
APULIA	PRIVATE	% of total	0.0 %	8.3 %	0.8 %	11.5 %	0.0 %	20.5 %
	PUBLIC	% of total	9.5 %	32.5 %	2.5 %	0.0 %	34.9 %	79.5 %
	Total	% of total	9.5 %	40.8 %	3.3 %	11.5 %	34.9 %	100.0 %
Total	PRIVATE	% of total	10.8 %	3.9 %	0.8 %	6.9 %	1.7 %	24.2 %
	PUBLIC	% of total	4.8 %	18.9 %	5.2 %	0.0 %	46.9 %	75.8 %
	Total	% of total	15.6 %	22.8 %	6.0 %	6.9 %	48.7 %	100.0 %

The analysis of the contingency table shows specific distributions of the residents/doctor ratio in the two regions. In Emilia-Romagna, 21.0% of private facilities and 72.5% of public facilities make up the total health facilities, with a high percentage (60.9%) dedicated to second-level facilities. This data highlights a marked focus on specialized and advanced care in the region. In Apulia, on the contrary, the distribution is more homogeneous. Private facilities represent 20.5% of the total, while public facilities constitute 79.5%. The distribution among the various levels of facilities is more balanced compared to Emilia-Romagna, with 40.8% of basic facilities, 3.3% of first-level facilities, and 11.5% of private facilities.

Table 19. χ^2 Tests

χ^2 Tests

REGION		Value	df	p
EMILIA ROMAGNA	χ^2	3.44e+6	4	< .001
	N	4.43e+6		
APULIA	χ^2	2.20e+6	4	< .001
	N	3.92e+6		
Total	χ^2	4.28e+6	4	< .001
	N	8.35e+6		

The analysis of the chi-square test (Table 19) clearly shows that the distribution of hospitals in the two regions is not uniform among different types and sectors. The very high $\chi 2$ values and very low p-values (<.001) for both regions and the total indicate that we can reject the null hypothesis. The results of the chi-square test

confirm the significant association between the region and the residents/doctor ratio for each level of hospital, further highlighting the need for careful analysis and planning of health resources in the two regions.
In Emilia-Romagna, a marked concentration of the residents/doctor ratio in second-level facilities highlights a pronounced focus on specialized and advanced care, suggesting the need for a redistribution of resources. On the contrary, Apulia shows a more harmonious distribution of the residents/doctor ratio among the different levels of hospital structure.
This study contributes to a deeper understanding of the regional dynamics of the residents/doctor ratio in Emilia-Romagna and Apulia, providing a solid foundation for the development of effective and sustainable health policies. As the two regions continue to evolve in response to growing health needs, the research underscores the importance of a holistic and data-driven approach to ensure that every citizen has access to timely and adequate medical care, thereby promoting the overall health and well-being of the population.

2.3 ANALYSIS OF ACTIVE KILOMETRIC MOBILITY IN APULIA AND EMILIA-ROMAGNA

Active kilometric mobility is an important indicator for understanding the dynamics of access to health services by the population. The analysis of active kilometric mobility in Emilia-Romagna and Apulia provides an in-depth look at the characteristics and health mobility needs of residents within the two regions, highlighting possible areas of intervention for the improvement of services and the optimization of available resources (Table 20).

Table 20. Contingency table: Percentage relationship between the nature of the hospital network and hospital level in the Apulia and Emilia Romagna regions expressed by active kilometric mobility.

			LEVEL					
REGION	NETWORK		BASE LEVEL	FIRST LEVEL	IRCCS	PRIVATE NURSING HOMES	SECOND LEVEL	Total
EMILIA ROMAGNA	PRIVATE	% of total	8.1 %	0.0 %	0.0 %	0.1 %	3.0 %	11.1 %
	PUBLIC	% of total	0.0 %	12.1 %	9.5 %	0.0 %	67.3 %	88.9 %
	Total	% of total	8.1 %	12.1 %	9.5 %	0.1 %	70.2 %	100.0 %
APULIA	PRIVATE	% of total	0.0 %	23.4 %	0.6 %	14.1 %	0.0 %	38.1 %
	PUBLIC	% of total	5.7 %	25.3 %	2.0 %	0.0 %	28.9 %	61.9 %
	Total	% of total	5.7 %	48.7 %	2.6 %	14.1 %	28.9 %	100.0 %
Total	PRIVATE	% of total	4.4 %	10.6 %	0.3 %	6.4 %	1.6 %	23.3 %
	PUBLIC	% of total	2.6 %	18.1 %	6.1 %	0.0 %	49.9 %	76.7 %
	Total	% of total	7.0 %	28.7 %	6.4 %	6.4 %	51.5 %	100.0 %

The analysis of active kilometric mobility of residents in Emilia-Romagna and Apulia highlights significant dynamics related to the type of facility and level of care. In Emilia-Romagna, the public sector dominates, absorbing 88.9% of total mobility, with a greater incidence at the second level of care (70.2%). This underlines the crucial importance of second-level public facilities in managing regional health mobility.
In contrast, Apulia shows a more balanced distribution between the public (61.9%) and private (38.1%) sectors. The first level of care plays a predominant role, accounting for 48.7% of total mobility. This highlights a greater distribution of mobility towards first-level facilities, both public and private, demonstrating a different organizational and functional setup compared to Emilia-Romagna.
At the aggregate level, observing both regions, the public sector maintains a predominant role (76.7%), with a marked inclination towards the second level of care (51.5%). This data, exploring active kilometric mobility within regional borders, offers valuable insights for the planning and improvement of health services, highlighting the specificities and needs of each territorial context.
The analysis of the χ2 test (Table 21) shows a significant level of association between the examined variables, with a p-value of < .001 in both regions and overall. This indicates that the distribution of active kilometric mobility among different levels and types of facilities is significantly different in the two regions. The high χ2 value in both regions and overall highlights an important discrepancy between expectations and actual observations, emphasizing the importance of considering regional specificities in the planning and management of health services.

Table 21. χ^2 Tests

REGION		Value	df	p
EMILIA ROMAGNA	χ^2	4.52e+6	4	< .001
	N	6.34e+6		
APULIA	χ^2	2.43e+6	4	< .001
	N	5.24e+6		
Total	χ^2	5.01e+6	4	< .001
	N	1.16e+7		

The analysis has highlighted a clear disparity between the two regions in terms of active kilometric mobility and use of care facilities, both public and private. In Emilia-Romagna, the predominance of the public sector and the second level of care highlights a health system strongly oriented towards high-level facilities, potentially capable of providing specialized and complex assistance. This could also reflect a greater propensity for admission to second-level facilities in the region.
On the contrary, Apulia shows greater balance between sectors and levels, signaling a different organizational model, potentially more decentralized and closer to the needs of the local population. The greater distribution of mobility towards first-level facilities in Apulia could indicate a lower propensity for hospitalization, with a possible preference for outpatient or short-term treatments.
The results of the analysis of active kilometric mobility in Emilia-Romagna and Apulia underscore the importance of careful and targeted health planning, able to respond to the specific territorial and population needs of each region. The

significant differences observed require diversified approaches for the continuous improvement of health services and to ensure fair and quality access to health care in both regions.
Consideration of the propensity for hospitalization, highlighted by mobility data, offers further insights for reflection and action. In Emilia-Romagna, attention could be focused on ensuring that the propensity for admission to second-level facilities is not excessive, avoiding overloading the more specialized facilities and ensuring efficient and timely healthcare. In Apulia, attention could be directed towards strengthening first-level facilities, ensuring that they are able to effectively respond to the health needs of the population and act as an effective filter for access to second-level facilities, thus ensuring optimal use of available resources.
In both cases, the analysis provides a valuable framework for directing future intervention strategies, contributing to the achievement of an increasingly efficient, effective, and responsive health system to the needs of citizens.

2.4 IMPLEMENTING THE CPDA METHODOLOGY FOR ENHANCED HEALTHCARE PERFORMANCE IN THE APULIA-EMILIA ROMAGNA MACROREGION

In the context of the ongoing evolution of health systems, the application of the CPDA methodology to the Apulia-Emilia-Romagna macroregion stands out as a crucial tool to promote health efficiency and excellence. The need for a thorough and objective analysis of health performances in these regions is made evident by the diversity and complexity of their health contexts. The objective is twofold: on one hand, it aims to provide a detailed and data-based evaluation of health performances, highlighting areas of strength and opportunities for improvement. On the other hand, it aims to develop and implement effective strategies to enhance access, quality, and efficiency of health care. The CPDA methodology, with its systematic and data-based approach, emerges as the key to deciphering the dynamics of regional health systems, offering valuable insights and supporting the decision-making process. In applying the CPDA methodology to the Apulia-Emilia-Romagna macroregion, it is essential to note that the variables under examination remain consistent with those discussed in paragraph 3.1 of the first section, although they are updated to 2021 data. This update allows for a more current and relevant evaluation of health performances in the regions in question. In particular, the OUT1_HOS variable has been replaced to ensure greater consistency with other output variables. The new data source for OUT1_HOS is the National Outcomes Plan, which provides detailed information on hospitalizations related to the indicators provided by the same plan.
Additionally, the OUT2_MOB variable, representing active intra-regional mobility by territorial scope, is now specifically referred to the Apulia-Emilia-Romagna macroregion. This change reflects the goal of exclusively analyzing kilometric mobility within the borders of these two regions, ensuring that the analysis is as focused as it is accurate. These changes and updates in the variables will ensure that the analysis of health performance in the Apulia and Emilia-Romagna regions is not only up-to-date but also meticulously aligned with relevant standards and metrics, thus providing high-quality and practically relevant results and insights.
To ensure continuity and consistency in the analysis in the second section of the document, the methodological workflow adopted will follow that outlined in paragraph 3.4 of the first section. The process of extracting and organizing data will follow the same path, using robust data mining tools like Knime for generating the

dataset with original input and output variables. The cluster analysis to identify input and output groups, the standardization of variables, the application of Principal Component Analysis, and ANOVA analysis will be performed using the Orange software. This methodological consistency will ensure that the results of the analysis are comparable and reliable, providing a solid foundation for any conclusions and recommendations.
The need for a deep and objective analysis of health performance in these regions is made evident by the diversity and complexity of their healthcare contexts. The goals are manifold:
Detailed Evaluation of Health Performance: provide a precise and data-based analysis of health performance, highlighting areas of strength and opportunities for improvement.
Evaluation of Hospital Efficiency for Perceived Quality: CPDA will be used to calculate hospital efficiency scores based on perceived quality, analyzing its influence on hospital performance.
Improvement of Access and Quality of Healthcare: develop and implement effective strategies to enhance access, quality, and efficiency of healthcare.
Optimization of Healthcare Resources: promote optimal allocation of resources to maximize benefits for patients and the overall healthcare system.
Support to the Decision-Making Process: provide valuable insights to support the decision-making process at all levels.

2.5 DEFINING AND ANALYZING RESEARCH QUESTIONS FOR HOSPITAL EFFICIENCY EVALUATION IN APULIA-EMILIA ROMAGNA

In the healthcare context of the Apulia-Emilia Romagna macroregion, the evaluation of hospital efficiency and understanding its impact on the perceived quality of healthcare are of paramount importance. The region, facing various challenges in terms of healthcare performance, requires in-depth analysis to identify potential areas for improvement and implement effective strategies to optimize both efficiency and the quality of care provided.
The proposed research question:
Q_4: "What is the current state of hospital efficiency in the Apulia-Emilia Romagna macroregion and how does it influence the perceived quality of healthcare by resident patients?",
aims to provide a clear and detailed picture of the current state, analyzing how hospital efficiency affects patients' perception of healthcare quality. This analysis allows for the outlining of targeted strategies for continuous improvement, guiding collective efforts towards the optimization of healthcare resources and the enhancement of patient satisfaction in the Apulia-Emilia Romagna macroregion.
Advantages of the Adopted Research Questions
The adopted research question allows for a comprehensive analysis of hospital efficiency in the Apulia-Emilia Romagna macroregion, providing insight into its impact on the perceived quality of healthcare. This approach offers the following advantages:

- Holistic Understanding: Offers a complete view of both hospital efficiency and its correlation with patients' perception of care quality, allowing for a deeper understanding of the current healthcare context in the region.

- Identifies Areas for Improvement: Helps in pinpointing specific areas where enhancements in efficiency and care quality can be made.
- Guides Policy and Decision Making: Provides valuable data and insights that can inform policy and decision-making for healthcare improvement in the region.

Impact on Results

The adopted research question's comprehensive nature ensures that the results obtained will be holistic and actionable. The insights gained will be:

- Actionable Insights: Deliver clear and actionable insights for healthcare administrators and policymakers to make informed decisions.
- Enhanced Healthcare Quality and Efficiency: Contribute to the enhancement of healthcare quality and efficiency in the macroregion.
- Improved Patient Satisfaction: Potentially lead to increased patient satisfaction by addressing the areas of concern identified through the research.

Alternatives to the Proposed Research Questions

While the proposed research question offers a comprehensive approach, alternatives could include:

- Focusing on Specific Aspects of Hospital Efficiency or Specific Medical Areas or Disciplines: Research questions could focus more narrowly on specific elements of hospital efficiency, such as staffing, resource allocation, patient throughput, or specific medical areas or disciplines.
- Exploring Perceived Healthcare Quality Independently: Separate research questions could explore patients' perceived quality of healthcare without considering hospital efficiency.
- Analyzing Other Geographical Regions: Research questions could focus on different geographical areas to understand the variability and similarities in hospital efficiency and perceived healthcare quality in various regions.

3. METHODOLOGY

3.1 APPLICATION OF THE HIERARCHICAL CLUSTERING ALGORITHM

In expanding our investigation on hospital efficiency (Paragraph 1.13.), we extended the analysis to a combined dataset that includes both Apulia and Emilia-Romagna, conducting a macro-regional analysis. In the initial phase focused solely on the Apulia region (Paragraph 1.13.), a hierarchical analysis was performed on non-standardized data (Paragraph 1.12.). This choice was due to the desire to maintain the natural distribution of the data and use the Spearman metric, suitable for non-normalized data, to explore the relationships between variables without the influence of standardization. However, with the expansion of the analysis to the combined dataset and the increased complexity and volume of data, we recognized the need for preliminary standardization of the variables to ensure uniformity and comparability across the two regions (Paragraph 1.14.).
The standardization was carried out using the formula (1) of paragraph 1.14.; where Z_i is the standardized value of the i-th variable (input or output), X_i is the respective non-standardized value of the i-th variable (input or output), μ_i is the mean of all observations in the i-th variable (input or output), is the standard deviation of all

observations in the i-th variable (input or output) and n is the number of variables (=15).
After standardization, hierarchical analysis was applied to the normalized data, adopting the cosine metric instead of the previous Spearman metric. This change was motivated by the need to adapt the analysis to the new distribution of normalized data, ensuring precise and effective segmentation. The Ward linkage method was retained for its proven effectiveness in minimizing intra-cluster variance.
The results of the extended analysis reinforced our initial conclusions, confirming the validity of the hierarchical approach even when applied to standardized and normalized data. The use of the cosine metric offered additional flexibility in the analysis, allowing precise and effective segmentation of data across different regions, and contributing to a deeper understanding of the dynamics of hospital efficiency in diversified regional contexts.
Despite the change in metric and the standardization of data, the results of the hierarchical analysis showed substantial consistency with our initial analyses (Paragraph 1.13.). Figure 32 show the scatterplots of the hierarchical algorithm with Spearman and Cosine metrics respectively, while figures 33 display the corresponding dendrograms.

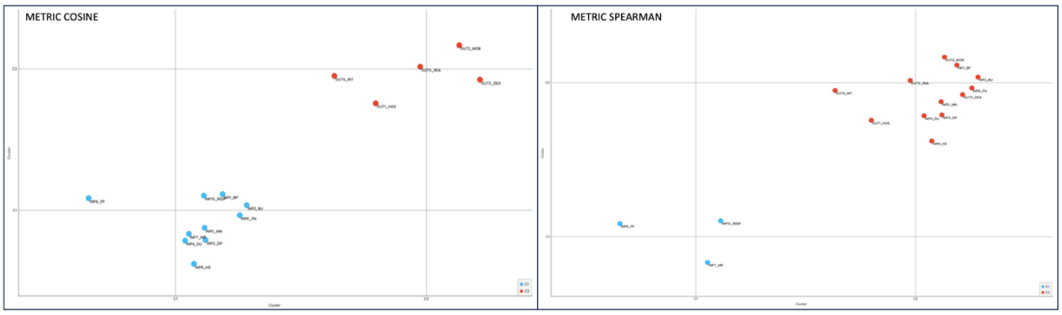


Figure 32 Scatterplots of the hierarchical algorithm with Spearman and Cosine metrics.

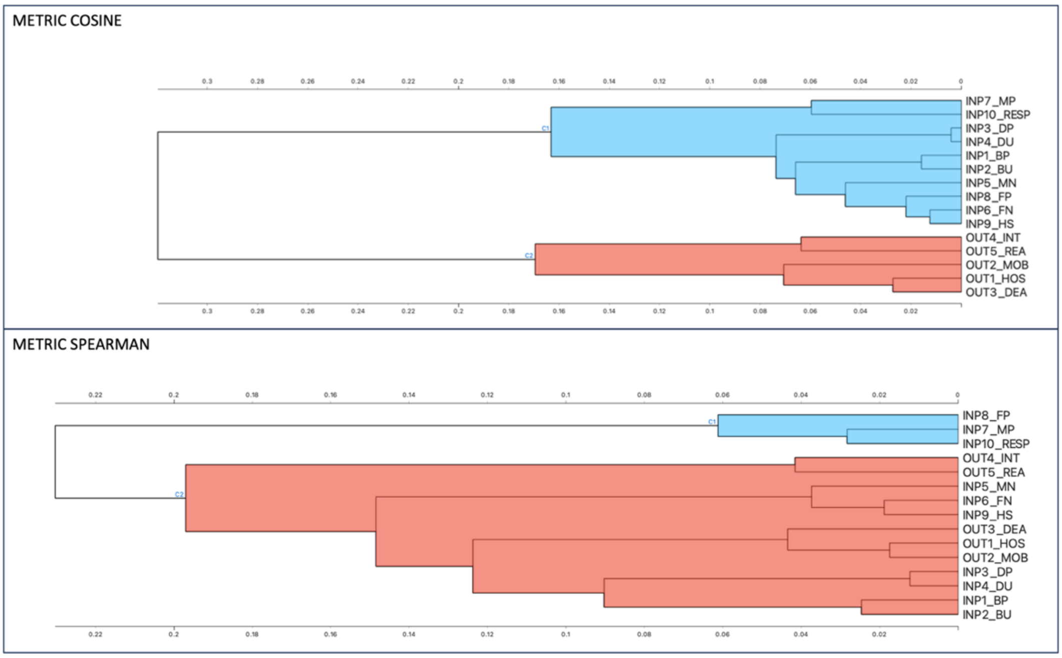


Figure 33 Dendrograms of the hierarchical algorithm with Spearman and Cosine metrics.

This reinforces our confidence in the application of the hierarchical clustering algorithm for hospital efficiency analysis, demonstrating its robustness and reliability even in diverse data contexts. Additionally, the obtained silhouette index, approximately 0.5, substantially confirms the value from the initial analysis, providing further validation of our results. Adapting to the use of standardized data and the cosine metric has contributed to a deeper understanding of the dynamics of hospital efficiency in diversified regional contexts. The experience gained in this process will provide a solid foundation for future research and analysis in this field, allowing for further improvements and refinements in the methodologies of hospital efficiency analysis.
The application of the hierarchical clustering algorithm with standardized data and the use of the cosine metric have proven to be valid tools for the analysis of hospital efficiency. The insights gained from this extended analysis will further enrich our CPDA model, enhancing its robustness and providing a strengthened methodological framework for future research in the field of hospital efficiency evaluation.
In the context of the hierarchical analysis conducted, it's crucial to highlight the composition of the identified clusters. In accordance with the referenced literature (as outlined in Table 1), Cluster C1 is composed of the 10 input variables, while Cluster C2 encompasses the 5 output variables.
This grouping underscores the importance of separately considering input and output in the analysis of hospital efficiency, ensuring that each aspect is thoroughly examined for a deeper understanding of the dynamics at play. The distinction between input and output variables within the context of the identified clusters contributes to a more accurate interpretation of the results, providing additional insight into the relationships and interactions among the various variables involved in the analysis of hospital efficiency.
Moreover, observing such a division can offer significant insights for further research, contributing to the delineation of targeted strategies for enhancing efficiency in the analyzed hospital contexts. Through a careful analysis of the input

and output variables within their respective clusters, specific areas of intervention can be identified, facilitating the planning and implementation of effective actions to bolster hospital efficiency in the various regions examined.

3.2 RELIABILITY AND EXPLORATORY FACTOR ANALYSIS

The reliability analysis, conducted coherently as outlined in Paragraph 1.16, was performed on the two distinct clusters, C1 and C2, playing a crucial role in the comprehensive analysis of the selected variables. Utilizing the Cronbach's alpha test, we conducted an accurate investigation into the internal consistency of the variables in each cluster. For C1, composed of input variables, an alpha of 0.994 was obtained, a value that underlines excellent internal consistency and demonstrates the reliability of the analyzed input variables. Similarly, for C2, which includes output variables, an alpha of 0.981 attests to solid internal consistency, affirming the reliability of the output variables in the conducted analysis. The detailed results of this analysis are illustrated in Figure 34.

CLUSTER 1

Scale Reliability Statistics

	Cronbach's α
scale	0.994

CLUSTER 2

Scale Reliability Statistics

	Cronbach's α
scale	0.981

Figure 34 Cronbach's alpha for the two clusters.

The heatmaps, depicted in Figure 35, clearly display the internal correlations within each cluster. Within C1, correlation values range between 0.87 and 0.99, signifying high cohesion among the input variables. This elevated degree of correlation highlights the tight interconnection among variables, revealing a common trend of coordinated movement. Similarly, C2 showcases correlation values between 0.85 and 0.97, signaling a similar synchronization among the output variables.

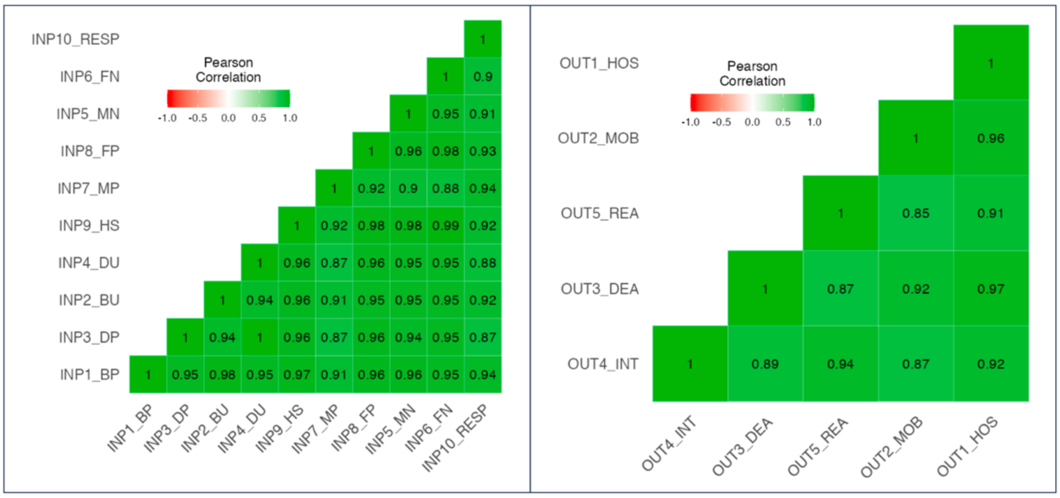


Figure 35 Heatmap for clusters C1 and C2.

The high internal consistency observed for both clusters further reinforce confidence in the comprehensive analysis conducted, highlighting the ability of each group of variables to provide significant insights and reliable measurements.

This meticulous approach to reliability analysis is essential to ensure the integrity and robustness of the entire study, providing a solid foundation for further investigations and analysis in this crucial area.
In alignment with the methodologies outlined in Paragraph 1.15, an Exploratory Factor Analysis (EFA) is meticulously applied to each cluster, C1 and C2, to delve deeper into the data structure within each cluster and unearth any underlying factors elucidating the relationships between the variables. This analysis aims to group together variables based on their heightened correlations, identifying latent factors that influence the observed variables. The results of the exploratory factor analyses applied separately to the two identified clusters are presented in Figure 36.

Exploratory Factor Analysis - Cluster 1

Factor Loadings

	Factor	
	1	Uniqueness
INP1_BP	0.985	0.0307
INP2_BU	0.977	0.0463
INP3_DP	0.967	0.0643
INP4_DU	0.971	0.0570
INP5_MN	0.974	0.0516
INP6_FN	0.975	0.0500
INP7_MP	0.926	0.1430
INP8_FP	0.987	0.0252
INP9_HS	0.993	0.0147
INP10_RESP	0.937	0.1226

Note. 'Minimum residual' extraction method was used in combination with a 'oblimin' rotation

Factor Statistics

Summary

Factor	SS Loadings	% of Variance	Cumulative %
1	9.39	93.9	93.9

Assumption Checks

Bartlett's Test of Sphericity

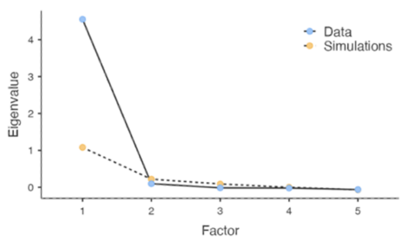
χ^2	df	p
3680	45	<.001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.878
INP1_BP	0.854
INP2_BU	0.870
INP3_DP	0.854
INP4_DU	0.866
INP5_MN	0.956
INP6_FN	0.840
INP7_MP	0.873
INP8_FP	0.946
INP9_HS	0.854
INP10_RESP	0.879

Eigenvalues

Scree Plot



Exploratory Factor Analysis - Cluster 2

Factor Loadings

	Factor	
	1	Uniqueness
OUT1_HOS	0.998	0.00387
OUT2_MOB	0.941	0.11399
OUT3_DEA	0.957	0.08347
OUT4_INT	0.945	0.10750
OUT5_REA	0.928	0.13915

Note. 'Minimum residual' extraction method was used in combination with a 'oblimin' rotation

Factor Statistics

Summary

Factor	SS Loadings	% of Variance	Cumulative %
1	4.55	91.0	91.0

Assumption Checks

Bartlett's Test of Sphericity

χ^2	df	p
1199	10	<.001

KMO Measure of Sampling Adequacy

	MSA
Overall	0.823
OUT1_HOS	0.740
OUT2_MOB	0.854
OUT3_DEA	0.822
OUT4_INT	0.899
OUT5_REA	0.825

Eigenvalues

Scree Plot

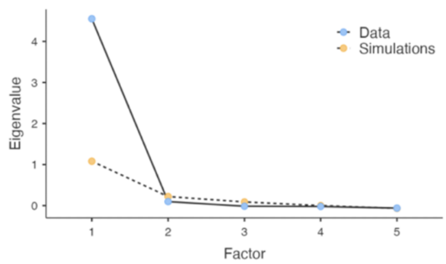


Figure 36 Factor Analysis results for the 2 clusters.

The exploratory factor analysis of the two clusters revealed significant findings. In Cluster 1, consisting of 10 variables, a single latent factor emerged that explains 93.9% of the total variance within the cluster. All variables within the cluster exhibit correlations above 0.96 with the factor, indicating a strong association among them. The Bartlett's test of sphericity confirmed the presence of a significant factor structure in the cluster, with a p-value below 0.001. Additionally, the KMO MSA indicates adequate data suitability for factor analysis in the cluster (0.878).

In Cluster 2, consisting of 5 variables, a single latent factor was identified that explains 91.0% of the total variance within the cluster. All variables within the cluster exhibit correlations above 0.95 with the factor, highlighting a strong association among them. The Bartlett's test of sphericity confirmed the presence of a significant factor structure in the cluster, with a p-value below 0.001. Additionally, the KMO MSA indicates adequate data suitability for factor analysis in the cluster (0.823).
The screening test based on the parallel analysis confirmed the importance of one factor in both clusters.
In conclusion, both clusters demonstrate significant factor structures and strong associations among variables. These findings indicate the presence of a latent factor in each cluster that can consistently explain the observed variations in their respective variables.

3.3 REASSESSMENT OF LIMITATIONS AND ADVANTAGES: RELIABILITY AND EXPLORATORY FACTOR ANALYSIS FOR APULIA AND EMILIA-ROMAGNA

The extension of the analysis from the sole context of Apulia to include Emilia-Romagna has introduced new dynamics, allowing for a more expansive and comprehensive view. This enlargement has also mitigated some of the initially observed limitations while further bolstering the advantages.
Updated Limitations:

- **Sample Size:** The inclusion of data from Emilia-Romagna has allowed for an expansion of the sample size, providing a more solid foundation for the Exploratory Factor Analysis and enhancing the capability to more precisely identify latent variables.
- **Sample Dependence:** The addition of another region has mitigated the dependence on the original sample, offering a more holistic view and reducing the risk of divergent results due to regional or temporal differences.
- **Subjective Interpretation:** The expansion of the analysis has allowed for cross-validation of results, reducing the risk of subjective interpretations and biases.

Strengthened Advantages:

- **Data Exploration:** Access to a broader data set has enriched exploration, allowing for greater depth and a wider understanding of latent structures in hospital data.
- **Robustness and Reliability:** Cross-validation across the two regions has strengthened the robustness and reliability of the analysis, confirming the consistency of the obtained results.
- **Foundation for Subsequent Analysis:** The extensive data base now available provides an even more solid foundation for further analysis, ensuring that future investigations are even more informed and reliable.

Implications:

- **Breadth of Analysis:** The geographical expansion of the analysis has captured a wider range of dynamics and factors, offering more generalizable and applicable insights to a broader context.

In summary, the update of the analysis, with the inclusion of data from Emilia-Romagna, has not only reinforced existing advantages but also significantly mitigated the initially perceived limitations. This extension has further solidified the analytical foundation, ensuring the results are not only robust but also representative of a broader context, thus providing a completer and more reliable analytical framework.

3.4 ADVANCING ANALYSIS WITH PCA AND POSITIVE TRANSFORMATION IN APULIA AND EMILIA-ROMAGNA

Aligning meticulously with the methodologies previously adopted for Apulia, we now expand our analytical horizon by applying Principal Component Analysis (PCA) to two distinct clusters, C1 and C2, within a macro-regional context that encompasses both Apulia and Emilia-Romagna. This extended perspective aims to overcome challenges related to discriminatory capacity in DEA studies, optimizing variable representation through PCA.
Detailed PCA Application:
Cluster 1 (Input Variables): The analysis unveils a dominant principal component, Input_PC1, capturing almost 95% of the total variance while retaining the informative richness of the original dataset.
Cluster 2 (Output Variables): Similarly, a principal component, Output_PC1, emerges, encapsulating nearly 93% of the total variance.
The clarity and acumen of the PCA analysis are reflected in the renaming of the principal components: Input_PC1 and Output_PC1, now known as Hospital Organization and Propension Hospitalization, respectively.
The fairly uniform distribution of incidence across all variables is manifestly evident, bolstering the robustness of the conducted analysis.
Graphical visualizations, presented in Figures 37 and 38, offer a detailed and intuitive overview of the PCA analysis results.

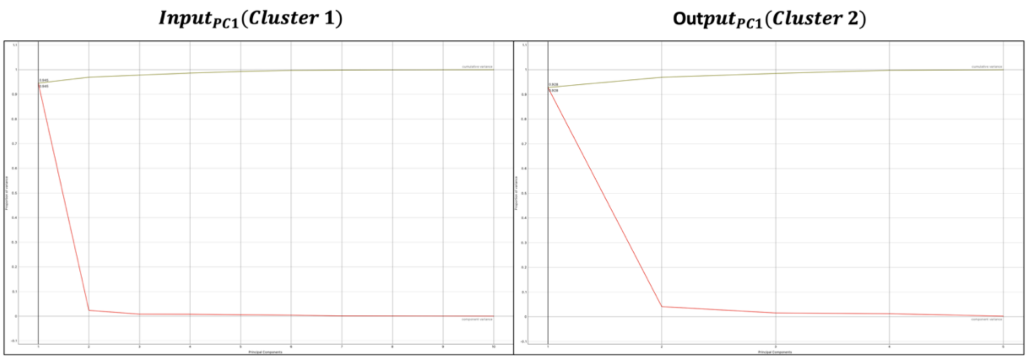


Figure 37 Principal Components variance representation.

Data Table

Data instances: 1
Features: 10
Meta attributes: 2

	components	variance	INP1_BP	INP2_BU	INP3_DP	INP4_DU	INP5_MN	INP6_FN	INP7_MP	INP8_FP	INP9_HS	INP10_RESP
1	PC1	0.945372	0.320254	0.318188	0.315744	0.31673	0.317466	0.317674	0.304771	0.320965	0.322324	0.307706

Data Table

Data instances: 1
Features: 5
Meta attributes: 2

	components	variance	OUT1_HOS	OUT2_MOB	OUT3_DEA	OUT4_INT	OUT5_REA
1	PC1	0.928004	0.45938	0.443622	0.448192	0.44479	0.439835

Figure 38 The impact of each individual original variable on the main components.

The transformation of principal components into exclusively positive values is a key step to neutralize the effect of variables with negative values in subsequent analyses, such as DEA analysis.
The addition of a specific positive constant to all principal components ensures the positivity of all values. The principal components are transformed into their positive counterparts using the formulas (4) and (5) of paragraph 1.1.9.
The judicious implementation of PCA, followed by the positive transformation of principal components, elevates the quality and depth of the analysis conducted on the Apulia and Emilia-Romagna regions. This sophisticated approach not only enhances the robustness of the analysis but also sheds new light on the dynamics of hospital efficiency in the two regions, laying a foundation for future insights and investigations in this critical sector.

3.5 DEA AS THIRD STEP: ENHANCING CPDA METHODOLOGY FOR A HOLISTIC ANALYSIS OF HOSPITAL EFFICIENCY IN APULIA AND EMILIA-ROMAGNA

In this section, we focus on the third step of our analytical methodology, emphasizing the application of Data Envelopment Analysis (DEA) within the context of hospital facilities in Apulia. The adoption of an output-oriented DEA model, integrated with variable returns to scale (VRS), allows for a comprehensive investigation into hospital efficiency, while keeping inputs constant. This model stands as an ideal tool for a detailed comparative analysis of hospital efficiency in the region, highlighting specific areas of strength and potential for improvement.
Following the implementation of the Principal Component Analysis (PCA), two new variables, Input_PC1 and Output_PC1, emerge as pillars for the subsequent DEA assessment. Table 22, post this operation, displays the hospital efficiency scores calculated both with the traditional DEA approach, listed in the PTE_DEA column, and with the CPDA methodology, outlined in the PTE_CPDA column.
This comparison, essential for a deep understanding of the Apulian hospital landscape, precedes a more extensive analytical discussion, scheduled in the subsequent sections of this work. The in-depth discriminatory analysis of efficiency scores will prove crucial in unveiling the impact of the different methodologies adopted on efficiency evaluations, providing a clear and detailed picture of the implications of the obtained results. This phase will be decisive in outlining concrete paths for enhancing hospital performance in Apulia, enriching the value and scope of the present study.

Table 22. Efficiency Scores Expressed for 59 Hospitals in Apulia Using DEA Analysis and Cluster-PCA-DEA Analysis.

REGION	ASL	NETWORK	LEVEL	HOSPITAL	PTE_DEA	PTE_CPDA
PUGLIA	ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura ' Monte Imperatore' - Noci (BA)	0,1693	0,574229
PUGLIA	ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura - Villa Lucia Hospital - Conversano (BA)	0,8717	0,662505
PUGLIA	ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Anthea - Bari (BA)	1	1
PUGLIA	ASL BA	PRIVATE	FIRST LEVEL	Casa Di Cura C.B.H. Mater Dei Hospital - Bari (BA)	1	0,785429
PUGLIA	ASL BA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Santa Maria - Bari (BA)	1	1
PUGLIA	ASL BA	PUBLIC	SECOND LEVEL	Consorziale Policlinico Bari - Bari (BA)	0,7822	0,685077
PUGLIA	ASL BA	PRIVATE	IRCCS	Ics Maugeri SPA Societa' Benefit - Bari (BA)	0,4799	0,504451
PUGLIA	ASL BA	PUBLIC	IRCCS	IRCCS 'Saverio De Bellis' - Castellana Grotte (BA)	0,8953	0,606006
PUGLIA	ASL BA	PUBLIC	IRCCS	Istituto Tumori Giovanni Paolo II - Bari (BA)	0,6819	0,498014
PUGLIA	ASL BA	PUBLIC	FIRST LEVEL	Ospedale Della Murgia - Perinei - Altamura (BA)	0,5612	0,532587
PUGLIA	ASL BA	PUBLIC	FIRST LEVEL	Ospedale Di Venere - Bari (BA)	0,8568	0,601823
PUGLIA	ASL BA	PUBLIC	BASE LEVEL	Ospedale Monopoli - Monopoli (BA)	0,8004	0,5953
PUGLIA	ASL BA	PUBLIC	BASE LEVEL	Ospedale Putignano - Putignano (BA)	0,2488	0,409595
PUGLIA	ASL BA	PRIVATE	FIRST LEVEL	Ospedale Regionale EE 'Miulli' - Acquaviva Delle Fonti (BA)	1	0,796713
PUGLIA	ASL BA	PUBLIC	FIRST LEVEL	Ospedale San Paolo - Bari (BA)	0,7054	0,447449
PUGLIA	ASL BR	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura 'Salus' - Brindisi (BR)	0,7137	0,621697
PUGLIA	ASL BR	PRIVATE	IRCCS	IRCCS E.Medea - Brindisi (BR)	1	0,779395
PUGLIA	ASL BR	PUBLIC	FIRST LEVEL	Ospedale Francavilla Fontana - Francavilla Fontana (BR)	1	0,600845
PUGLIA	ASL BR	PUBLIC	BASE LEVEL	Ospedale Ostuni - Ostuni (BR)	0,5684	0,551339
PUGLIA	ASL BR	PUBLIC	SECOND LEVEL	Ospedale Perrino - Brindisi (BR)	0,5651	0,470702
PUGLIA	ASL BT	PUBLIC	FIRST LEVEL	Ospedale Andria Andria (BT)	0,9588	0,592348
PUGLIA	ASL BT	PUBLIC	FIRST LEVEL	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	0,3841	0,377582
PUGLIA	ASL BT	PUBLIC	BASE LEVEL	Ospedale Bisceglie Bisceglie (BT)	0,0874	0,353443
PUGLIA	ASL BT	PRIVATE	PRIVATE NURSING HOMES	Ospedale Opera Don Uva Bisceglie (BT)	0,4495	0,500214
PUGLIA	ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura 'S.Michele' Gest. Brodetti Manfredonia (FG)	1	0,92914
PUGLIA	ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Universo Salute - Don Uva Foggia (FG)	0,7454	0,593456
PUGLIA	ASL FG	PRIVATE	PRIVATE NURSING HOMES	Case Cura Riunite Villa Serena-S. Francesco Foggia (FG)	1	0,71774
PUGLIA	ASL FG	PUBLIC	FIRST LEVEL	Ospedale Casa Sollievo Della Sofferenza San Giovanni Rotondo (FG)	0,6631	0,503695
PUGLIA	ASL FG	PUBLIC	BASE LEVEL	Ospedale Manfredonia Manfredonia (FG)	0,4332	0,527094
PUGLIA	ASL FG	PUBLIC	FIRST LEVEL	Ospedale San Severo - Teresa Masselli San Severo (FG)	0,7444	0,525847
PUGLIA	ASL FG	PUBLIC	SECOND LEVEL	Ospedali Riuniti Di Foggia Foggia (FG)	0,8331	0,607655
PUGLIA	ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Leonardo De Luca - Castelnuovo Della Daunia (FG)	0,3369	0,734874
PUGLIA	ASL FG	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Prof. Brodetti - Foggia (FG)	1	0,694102
PUGLIA	ASL FG	PUBLIC	FIRST LEVEL	Ospedale Cerignola 'S.Tatarella' - Cerignola (FG)	0,5521	0,451076
PUGLIA	ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Citta' Di Lecce Lecce (LE)	1	0,914946
PUGLIA	ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura 'Prof. Petrucciani' SRL Lecce (LE)	0,8998	0,57916
PUGLIA	ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Riabilitativa Euroitalia Casarano (LE)	0,0688	0,665311
PUGLIA	ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura San Francesco Galatina (LE)	1	0,523469
PUGLIA	ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Villa Bianca Lecce (LE)	1	0,746394
PUGLIA	ASL LE	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Villa Verde - Lecce (LE)	0	0,676674
PUGLIA	ASL LE	PUBLIC	BASE LEVEL	Ospedale Copertino Copertino (LE)	0,7125	0,594661
PUGLIA	ASL LE	PUBLIC	BASE LEVEL	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	0,3101	0,389855
PUGLIA	ASL LE	PUBLIC	FIRST LEVEL	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	0,5987	0,548963
PUGLIA	ASL LE	PUBLIC	SECOND LEVEL	Ospedale Lecce 'V. Fazzi' Lecce (LE)	0,7366	0,561039
PUGLIA	ASL LE	PRIVATE	FIRST LEVEL	Ospedale Regionale EE 'G. Panico' Tricase (LE)	1	0,892321
PUGLIA	ASL LE	PUBLIC	FIRST LEVEL	Ospedale Scorrano Scorrano (LE)	0,828	0,641578
PUGLIA	ASL LE	PUBLIC	BASE LEVEL	Ospedale Casarano - Casarano (LE)	0,5695	0,561852
PUGLIA	ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Bernardini Taranto (TA)	0,7875	0,663237
PUGLIA	ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura D'Amore SRL Taranto (TA)	0,8508	0,693802
PUGLIA	ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura San Camillo Taranto (TA)	0,5897	0,626125
PUGLIA	ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Santa Rita SRL Taranto (TA)	0,56	0,741463
PUGLIA	ASL TA	PRIVATE	FIRST LEVEL	Casa Di Cura Villa Bianca SRL - Martina Franca (TA)	0	0,732358
PUGLIA	ASL TA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Villa Verde SRL Taranto (TA)	1	0,738396
PUGLIA	ASL TA	PRIVATE	PRIVATE NURSING HOMES	Centro Medico Riabilitazione Ics Maugeri Ginosa (TA)	1	0,781836
PUGLIA	ASL TA	PRIVATE	PRIVATE NURSING HOMES	Fondazione Cittadella Della Carita` Taranto (TA)	0,536	0,715853
PUGLIA	ASL TA	PUBLIC	FIRST LEVEL	Ospedale Castellaneta Castellaneta (TA)	0,7923	0,530082
PUGLIA	ASL TA	PUBLIC	FIRST LEVEL	Ospedale Civile Martina Franca (TA)	0,8694	0,60505
PUGLIA	ASL TA	PUBLIC	BASE LEVEL	Ospedale Manduria 'Giannuzzi' Manduria (TA)	0,1974	0,428506
PUGLIA	ASL TA	PUBLIC	SECOND LEVEL	Presidio Ospedaliero centrale Taranto	0,5613	0,478863

EMILIA ROMAGNA	AUSL BO	PRIVATE	BASE LEVEL	Casa Di Cura Prof. Nobili SPA Castiglione Dei Pepoli (BO)	0,4331	0,570395
EMILIA ROMAGNA	AUSL BO	PUBLIC	IRCCS	Istituto Delle Scienze Neurologiche Bologna (BO)	1	0,733397
EMILIA ROMAGNA	AUSL BO	PUBLIC	IRCCS	Istituto Ortopedico Rizzoli Bologna (BO)	1	0,787316
EMILIA ROMAGNA	AUSL BO	PRIVATE	PRIVATE NURSING HOMES	Ospedale Privato Accreditato Casa Di Cura - Bologna (BO)	0	0,890071
EMILIA ROMAGNA	AUSL BO	PRIVATE	BASE LEVEL	Ospedale Privato Accreditato Nigrisoli Bologna (BO)	0,6194	0,559011
EMILIA ROMAGNA	AUSL BO	PRIVATE	BASE LEVEL	Ospedale Privato Accreditato Villa Chiara Casalecchio Di Reno (BO)	1	0,791015
EMILIA ROMAGNA	AUSL BO	PRIVATE	PRIVATE NURSING HOMES	Ospedale Privato Accreditato Villa Baruzziana - Bologna (BO)	0	0,668614
EMILIA ROMAGNA	AUSL BO	PRIVATE	BASE LEVEL	Ospedale Privato Accreditato Villa Laura Bologna (BO)	0,7304	0,533894
EMILIA ROMAGNA	AUSL BO	PRIVATE	BASE LEVEL	Ospedale Privato Accreditato Villa Regina Bologna (BO)	1	0,805941
EMILIA ROMAGNA	AUSL BO	PRIVATE	SECOND LEVEL	Ospedale Privato Accreditato Villa Torri Bologna (BO)	0,6246	0,618053
EMILIA ROMAGNA	AUSL BO	PRIVATE	PRIVATE NURSING HOMES	Ospedale Privato Santa Viola Bologna (BO)	0,1231	0,736787
EMILIA ROMAGNA	AUSL BO	PRIVATE	BASE LEVEL	Casa Di Cura Villa Erbosa Ospedale Privato - Bologna (BO)	0,6291	0,510149
EMILIA ROMAGNA	AUSL BO	PUBLIC	IRCCS	IRCCS Policlinico S. Orsola - Bologna (BO)	1	1
EMILIA ROMAGNA	AUSL BO	PRIVATE	PRIVATE NURSING HOMES	Ospedale Privato Accreditato Villa Bellombra - Bologna (BO)	0,0432	0,728297
EMILIA ROMAGNA	AUSL BO	PUBLIC	SECOND LEVEL	PRESIDIO OSPEDALIERO UNICO - AZIENDA DI BOLOGNA	1	0,980477
EMILIA ROMAGNA	AUSL BO	PRIVATE	PRIVATE NURSING HOMES	Villa Ranuzzi - Bologna (BO)	0	1
EMILIA ROMAGNA	AUSL FE	PUBLIC	SECOND LEVEL	Azienda Ospedaliero-Universitaria Ferrara (FE)	0,7536	0,676907
EMILIA ROMAGNA	AUSL FE	PRIVATE	BASE LEVEL	Casa Di Cura Quisisana SRL Ferrara (FE)	1	0,785199
EMILIA ROMAGNA	AUSL FE	PRIVATE	BASE LEVEL	Casa Di Cura Salus SRL Ferrara (FE)	1	0,778615
EMILIA ROMAGNA	AUSL FE	PUBLIC	FIRST LEVEL	PRESIDIO OSPEDALIERO UNICO - COMACCHIO - (FE)	0,4711	0,470732
EMILIA ROMAGNA	AUSL IML	PUBLIC	SECOND LEVEL	PRESIDIO OSPEDALIERO IMOLA	0,4831	0,463576
EMILIA ROMAGNA	AUSL IML	PUBLIC	BASE LEVEL	Ospedale Montecatone Rehabilitation Institute - Imola (BO)	0,0004	0,407358
EMILIA ROMAGNA	AUSL MO	PUBLIC	SECOND LEVEL	Azienda Ospedaliero-Universitaria Modena (MO)	0,6972	0,591601
EMILIA ROMAGNA	AUSL MO	PRIVATE	SECOND LEVEL	Hesperia Hospital Modena SRL Modena (MO)	1	0,703623
EMILIA ROMAGNA	AUSL MO	PUBLIC	FIRST LEVEL	Nuovo Ospedale Civile Di Sassuolo SPA Sassuolo (MO)	1	0,957779
EMILIA ROMAGNA	AUSL MO	PRIVATE	PRIVATE NURSING HOMES	Ospedale Privato Accreditato Villa Rosa - Modena (MO)	0	0,649048
EMILIA ROMAGNA	AUSL MO	PRIVATE	PRIVATE NURSING HOMES	Ospedale Privato "Villa Igea SPA" Modena (MO)	0,0046	0,419128
EMILIA ROMAGNA	AUSL MO	PRIVATE	BASE LEVEL	Prof. Fogliani Casa Di Cura SRL Modena (MO)	0,8135	0,649574
EMILIA ROMAGNA	AUSL MO	PRIVATE	BASE LEVEL	Villa Pineta SRL Pavullo Nel Frignano (MO)	0,4539	0,601571
EMILIA ROMAGNA	AUSL MO	PUBLIC	SECOND LEVEL	PRESIDIO OSPEDALIERO PROVINCIALE - CARPI - (MO)	1	1
EMILIA ROMAGNA	AUSL PC	PRIVATE	BASE LEVEL	Casa Di Cura Privata Piacenza SPA Piacenza (PC)	1	0,810436
EMILIA ROMAGNA	AUSL PC	PRIVATE	BASE LEVEL	Casa Di Cura Privata S. Antonino SRL Piacenza (PC)	0,1448	0,674232
EMILIA ROMAGNA	AUSL PC	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura S.Giacomo SRL - Ponte Dell'olio (PC)	0,1929	0,533984
EMILIA ROMAGNA	AUSL PC	PUBLIC	SECOND LEVEL	PRESIDIO UNICO PIACENZA - PIACENZA - (PC)	0,6691	0,670952
EMILIA ROMAGNA	AUSL PR	PUBLIC	SECOND LEVEL	Azienda Ospedaliero-Universitaria Parma (PR)	0,8466	0,798585
EMILIA ROMAGNA	AUSL PR	PRIVATE	BASE LEVEL	Casa Di Cura Citta' Di Parma Parma (PR)	1	0,988475
EMILIA ROMAGNA	AUSL PR	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Villa Igea - Salsomaggiore Terme (PR)	0	0,771325
EMILIA ROMAGNA	AUSL PR	PRIVATE	PRIVATE NURSING HOMES	Centro Cardinal Ferrari SRL Fontanellato (PR)	0,0036	0,556136
EMILIA ROMAGNA	AUSL PR	PRIVATE	BASE LEVEL	Hospital Piccole Figlie Parma (PR)	0,9525	0,612529
EMILIA ROMAGNA	AUSL PR	PRIVATE	BASE LEVEL	Hospital Val Parma Langhirano (PR)	1	0,534342
EMILIA ROMAGNA	AUSL PR	PRIVATE	PRIVATE NURSING HOMES	Fondazione Don Carlo Gnocchi Onlus - Parma (PR)	0,5756	0,600578
EMILIA ROMAGNA	AUSL PR	PUBLIC	FIRST LEVEL	PRESIDIO OSPEDALIERO AZIENDALE - AUSL PAR - FIDENZA - (PR)	0,7001	0,647642
EMILIA ROMAGNA	AUSL PR	PRIVATE	PRIVATE NURSING HOMES	Villa Maria Luigia - Montechiarugolo (PR)	0	0,539678
EMILIA ROMAGNA	AUSL RE	PRIVATE	BASE LEVEL	Casa Di Cura Privata Polispecialistica Reggio Nell'emilia (RE)	0,6372	0,578566
EMILIA ROMAGNA	AUSL RE	PRIVATE	BASE LEVEL	Salus Hospital (Casa Di Cura Privata) Reggio Nell'emilia (RE)	0,8206	0,634182
EMILIA ROMAGNA	AUSL RE	PUBLIC	SECOND LEVEL	PRESIDIO OSPEDALIERO PROVINCIALE - REGGIO NELL'EMILIA - (RE)	1	0,992597
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Casa Di Cura Privata San Lorenzino SPA Cesena (FC)	0,7751	0,602686
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Casa Di Cura Prof. E. Montanari Morciano Di Romagna (RN)	0,6182	0,532602
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	PRIVATE NURSING HOMES	Casa Di Cura Privata Villa Azzurra - Riolo Terme (RA)	0	0,741289
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Casa Di Cura San Francesco Ravenna (RA)	0,0867	0,602251
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Casa Di Cura Villa Maria Rimini (RN)	0,7565	0,618792
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	IRCCS	I.R.S.T. SRL IRCCS Meldola (FC)	0,0131	0,406521
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	PRIVATE NURSING HOMES	Luce Sul Mare - Bellaria-igea Marina (RN).	0	0,696636
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Malatesta Novello Cesena (FC)	1	0,67692
EMILIA ROMAGNA	AUSL ROMAGNA	PUBLIC	FIRST LEVEL	Ospedale "Degli Infermi " Faenza (RA)	0,5939	0,579597
EMILIA ROMAGNA	AUSL ROMAGNA	PUBLIC	SECOND LEVEL	Ospedale "Santa Maria Delle Croci" Ravenna (RA)	0,7704	0,629709
EMILIA ROMAGNA	AUSL ROMAGNA	PUBLIC	FIRST LEVEL	Ospedale "Umberto i" Lugo (RA)	0,5193	0,527429
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Ospedale Privato Accreditato Villa Igea Forli' (FC)	0,6279	0,586772
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Ospedale Privato Domus Nova SPA Ravenna (RA)	0,9971	0,62774
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Ospedale Privato San Pier Damiano Faenza (RA)	0,5967	0,541527
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Sol Et Salus Rimini (RN)	0,7245	0,59967
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	SECOND LEVEL	Villa Maria Cecilia Hospital Cotignola (RA)	1	0,786186
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	PRIVATE NURSING HOMES	Villa Salus SRL - Rimini (RN)	0	0,859121
EMILIA ROMAGNA	AUSL ROMAGNA	PRIVATE	BASE LEVEL	Villa Serena Forli' (FC)	0,7736	0,631872
EMILIA ROMAGNA	AUSL ROMAGNA	PUBLIC	SECOND LEVEL	PRESIDIO OSPEDALIERO CESENA - CESENA - (FC)	0,6165	0,533267
EMILIA ROMAGNA	AUSL ROMAGNA	PUBLIC	SECOND LEVEL	PRESIDIO OSPEDALIERO FORLI' - FORLI' - (FC)	0,7797	0,633316
EMILIA ROMAGNA	AUSL ROMAGNA	PUBLIC	FIRST LEVEL	PRESIDIO OSPEDALIERO RICCIONE-CATTOLICA - RICCIONE - (RN)	0,8277	0,690173
EMILIA ROMAGNA	AUSL ROMAGNA	PUBLIC	SECOND LEVEL	PRESIDIO OSPEDALIERO RIMINI (RN)	0,7445	0,619921

The DEA methodology has been applied in line with the analysis previously conducted for the Apulia region alone, as detailed in paragraph 1.20. This approach allows for a consistent and comparable evaluation of hospital efficiency across the two regions, ensuring a homogeneous understanding of the results obtained.

3.6 COMPARATIVE ANALYSIS OF HOSPITAL NETWORK EFFICIENCY BETWEEN APULIA AND EMILIA ROMAGNA: AN ANOVA PERSPECTIVE

Efficiency scores, such as Pure Technical Efficiency (PTE), can vary in their distribution. To ensure the application of standard statistical techniques, such as ANOVA, the scores should ideally follow a normal distribution. To achieve this, the PTE scores were normalized using a logarithmic transformation. This normalization step not only allows for the use of ANOVA but also ensures that the data is well-suited for such an analysis.
Before applying ANOVA, certain assumptions need to be met. These include the homogeneity of variances and the normality of the distribution.
Homogeneity of Variances: As presented in Figure 39, both Levene's and Bartlett's tests were employed. The p-values obtained (Levene's test p=0.136 and Bartlett's test p=0.113) indicate that the variances are assumed to be equal across the groups.
Normality of the Distribution: The normality of the distribution was also checked using multiple tests. The results, illustrated in Figure 39, showed that the distribution is assumed to be normal based on the Shapiro-Wilk (p=0.699), Kolmogorov-Smirnov (p=0.979), and Anderson-Darling (p=0.887) tests.
A QQ-Plot was used to provide a visual representation of the distribution of efficiency scores, which further confirmed the normality of the data.
ANOVA – The Fourth Step of the CPDA Methodology: After the necessary assumptions were verified, the ANOVA was applied to evaluate efficiency differences among different hospital networks in Apulia and Emilia Romagna, as shown in Figure 39. The analysis revealed significant differences based on both the region and the type of network (private or public). The interaction term (REGION ✻ NETWORK) also showed significant differences, indicating that the efficiency of hospital networks varies not only by region but also differently within each region based on whether the network is public or private.

ANOVA

ANOVA - LOG_PTE_CPDA

	Sum of Squares	df	Mean Square	F	p
REGION	0.0425	1	0.04246	5.60	0.019
NETWORK	0.0980	1	0.09804	12.94	<.001
REGION ✻ NETWORK	0.1736	1	0.17359	22.91	<.001
Residuals	0.9319	123	0.00758		

Assumption Checks

Homogeneity of Variances Tests

	Statistic	df	df2	p
Levene's	1.88	3	123	0.136
Bartlett's	5.97	3		0.113

Note. Additional results provided by *moretests*

Normality tests

	statistic	p
Shapiro-Wilk	0.992	0.699
Kolmogorov-Smirnov	0.0418	0.979
Anderson-Darling	0.197	0.887

Note. Additional results provided by *moretests*

Q-Q Plot

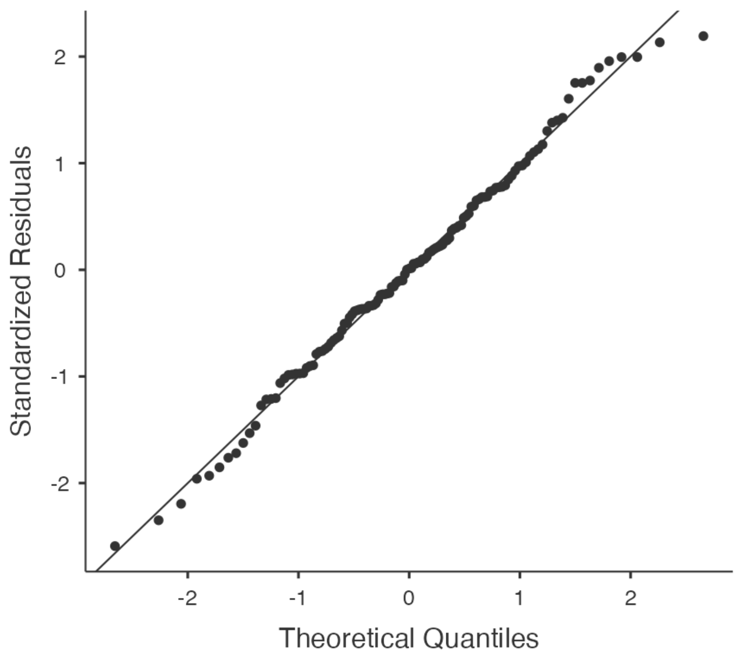


Figure 39 ANOVA Analysis of Hospital Network Efficiency in Apulia and Emilia Romagna.

4. DISCUSSION AND RESULTS

In this section, we will present a detailed analysis of hospital efficiency within the Apulia-Emilia Romagna macroregion. We utilized the CPDA methodology to derive our efficiency scores, offering a robust and comprehensive assessment of hospital performance. This analysis delves into the differences between the two regions in terms of pure technical efficiency and scale efficiency. Additionally, we will examine how these efficiency measures impact the perceived quality of healthcare, as denoted by the hospitalization propensity of resident patients.

4.1 RESEARCH QUESTION FOUR

Within the scope of our analysis based on the CPDA methodology, the propensity for hospitalization was considered as the primary indicator of the quality of healthcare perceived by patients. Five facilities achieved a pure technical efficiency (PTE) score of 1:

- Casa Di Cura Anthea – Bari;
- Casa Di Cura Santa Maria – Bari;
- IRCCS Policlinico S. Orsola – Bologna;
- Villa Ranuzzi – Bologna;
- Presidio Ospedaliero Provinciale - Carpi.

Among these, the Casa Di Cura Anthea - Bari both achieved a pure technical efficiency score and a scale efficiency score of 1, positioning itself as the only hospital in the macroregion to achieve technical efficiency (TE). These findings indicate that, while several facilities maximized their efficiency in terms of resource transformation into outputs (PTE), only one achieved both pure technical efficiency and optimal operational size (SE). The trend graph and the related comparison of efficiency components are shown in Figure 40.

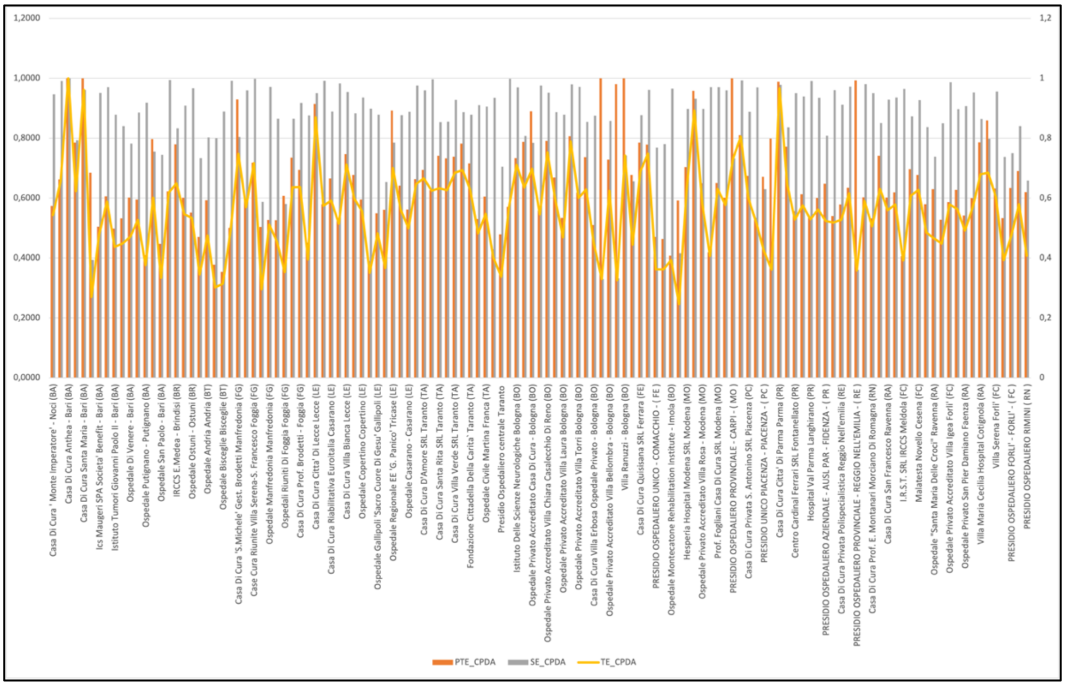


Figure 40 Technical Efficiency decomposition of Apulia – Emilia Romagna hospitals.

To further delve into these differences, we conducted a post-hoc test, the specifics of which are detailed in Table 23. This test allowed us to pinpoint the exact locations of differences between the groups following an ANOVA analysis. The detailed results provide a clearer picture of the differences between hospital facilities in the two regions in terms of pure technical efficiency.

Table 23. Post-hoc comparisons of hospital facilities in the Apulia and Emilia Romagna regions concerning Pure Technical Efficiency (PTE).

Comparison									
REGION	NETWORK		REGION	NETWORK	Mean Difference	SE	df	t	p_{tukey}
EMILIA ROMAGNA	PRIVATE	-	EMILIA ROMAGNA	PUBLIC	-0.0190	0.0226	123	-0.843	0.834
		-	PUGLIA	PRIVATE	-0.0387	0.0204	123	-1.894	0.236
		-	PUGLIA	PUBLIC	0.0954	0.0206	123	4.621	< .001
	PUBLIC	-	PUGLIA	PRIVATE	-0.0197	0.0244	123	-0.805	0.852
		-	PUGLIA	PUBLIC	0.1144	0.0246	123	4.649	< .001
PUGLIA	PRIVATE	-	PUGLIA	PUBLIC	0.1341	0.0227	123	5.915	< .001

Note. Comparisons are based on estimated marginal means

The results provided come from a post-hoc comparison, often conducted after an ANOVA when the latter indicates significant differences but doesn't specify where these differences lie. The post-hoc test provides pairwise comparisons to find out which groups differ from each other. It's important to note that the post-hoc test and the ANOVA were conducted on the logarithmically normalized value of PTE.

EMILIA ROMAGNA (PRIVATE) vs. EMILIA ROMAGNA (PUBLIC): The average difference in Pure Technical Efficiency (PTE) between private and public hospitals in Emilia Romagna is -0.0190. The standard error (SE) of this difference is 0.0226. The t-value is -0.843, and the p-value is 0.834, which is not significant (as it is above the common alpha level of 0.05). This means there isn't a significant difference in PTE between private and public hospitals within Emilia Romagna.

EMILIA ROMAGNA (PRIVATE) vs. APULIA (PRIVATE): The average difference is -0.0387, with an SE of 0.0204. The t-value is -1.894 and the p-value is 0.236, which again is not significant. This suggests that private hospitals in Emilia Romagna and Apulia don't significantly differ in terms of PTE.

EMILIA ROMAGNA (PRIVATE) vs. APULIA (PUBLIC): The average difference is 0.0954, with an SE of 0.0206. The t-value is 4.621 and the p-value is less than 0.001, highly significant. This means there's a significant difference in PTE between private hospitals in Emilia Romagna and public hospitals in Apulia, with the latter showing higher PTE.

EMILIA ROMAGNA (PUBLIC) vs. APULIA (PRIVATE): The average difference is -0.0197, with an SE of 0.0244. The t-value is -0.805 and the p-value is 0.852, indicating no significant difference.

EMILIA ROMAGNA (PUBLIC) vs. APULIA (PUBLIC): The average difference is 0.1144, with an SE of 0.0246. The t-value is 4.649 and the p-value is less than 0.001. This is again highly significant, suggesting that public hospitals in Apulia have significantly higher PTE compared to those in Emilia Romagna.

APULIA (PRIVATE) vs. APULIA (PUBLIC): The average difference is 0.1341, with an SE of 0.0227. The t-value is 5.915 and the p-value is less than 0.001. This result indicates a significant difference between private and public hospitals within Apulia, with public hospitals showing higher PTE.

In summary, the most pronounced differences in PTE are found between private hospitals in Emilia Romagna and public hospitals in Apulia, and between public

hospitals of the two regions. Public hospitals in Apulia consistently show higher PTE scores in these comparisons.
The analysis of scale efficiency revealed intriguing differences across regions, different networks, and various hospital levels. Let's begin by examining the scale efficiency differences among healthcare facilities in the Apulia and Emilia-Romagna regions (Table 24). The non-parametric Kruskal-Wallis test, applied to the SE_CPDA data, produced a $\chi 2$ value of 0.142 with one degree of freedom. The associated p-value of 0.706 suggests that there are no statistically significant differences in scale efficiency across hospitals in these two regions.

Table 24. Kruskal-Wallis Test Results for Scale Efficiency across the Apulia and Emilia Romagna Regions.

	χ^2	df	p
SE_CPDA	0.142	1	0.706

However, the picture changes when scale efficiency is analyzed distinguishing between the private and public hospital networks of the macro-region (Table 25). The Kruskal-Wallis test indicated a $\chi 2$ value of 33.4 with one degree of freedom, and a p-value less than 0.001, signaling statistically significant differences. Pairwise comparisons underscored a pronounced difference between the private and public networks: a W value of -8.17 with a p-value less than 0.001 denotes that the public network exhibits greater scale efficiency compared to its private counterpart.

Table 25. Kruskal-Wallis Test Results for Scale Efficiency across the private and public hospital networks of the macro-region Apulia and Emilia Romagna Regions.

	χ^2	df	p
SE_CPDA	33.4	1	< .001

Further analyses were conducted to explore scale efficiency differences across various hospital levels (Table 26). The Kruskal-Wallis test returned a $\chi 2$ value of 58.7 with 4 degrees of freedom and a p-value less than 0.001, pointing to the presence of significant differences across levels.

Table 26. Kruskal-Wallis Test Results for Scale Efficiency across various hospital levels of the macro-region Apulia and Emilia Romagna Regions.

	χ^2	df	p
SE_CPDA	58.7	4	< .001

Table 27. Dwass – Steel – Critchlow – Fligner pairwise comparisons for Scale Efficiency across various hospital levels of the macro-region Apulia and Emilia Romagna Regions.

		W	p
BASE LEVEL	FIRST LEVEL	-8.14	< .001
BASE LEVEL	IRCCS	-2.10	0.573
BASE LEVEL	PRIVATE NURSING HOMES	-1.80	0.708
BASE LEVEL	SECOND LEVEL	-7.90	< .001
FIRST LEVEL	IRCCS	2.43	0.425
FIRST LEVEL	PRIVATE NURSING HOMES	5.99	< .001
FIRST LEVEL	SECOND LEVEL	-5.44	0.001
IRCCS	PRIVATE NURSING HOMES	1.12	0.933
IRCCS	SECOND LEVEL	-3.45	0.105
PRIVATE NURSING HOMES	SECOND LEVEL	-7.54	< .001

Pairwise comparisons (Table 27) provided additional insights:

- The comparison between base-level hospitals and first-level hospitals yielded a W value of -8.14 and a p-value less than 0.001, indicating statistically significant differences.
- The difference between base-level hospitals and IRCCS isn't statistically significant, with a W value of -2.10 and a p-value of 0.573.
- There are no significant differences between base-level hospitals and private nursing homes, with a W value of -1.80 and a p-value of 0.708.
- Comparing base-level and second-level hospitals, significant differences arise with a W value of -7.90 and a p-value less than 0.001.
- There are no statistically significant differences between first-level hospitals and IRCCS, and between IRCCS and private nursing homes.
- However, significant differences exist between first level and second level hospitals ($W = -5.44$, $p = 0.001$), and between private nursing homes and second level hospitals ($W = -7.54$, $p < 0.001$).

These findings highlight the varied performance in terms of scale efficiency across healthcare facilities in the Apulia-Emilia Romagna macro-region. While healthcare facilities across the two regions don't markedly differ in terms of scale efficiency, clear differences arise when considering the private or public nature of the facilities and their service level.

4.2 PERCEIVED QUALITY INFLUENCED BY HOSPITAL EFFICIENCY IN APULIA AND EMILIA ROMAGNA

In line with the analysis previously conducted solely for the Apulia region, as described in Section 1.22.1, we expanded our investigation to also include the Emilia Romagna region. The relationship between the quality perceived by patients, measured through their propensity for hospitalization, and the efficiency of a hospital is a crucial aspect to consider when assessing the effectiveness of a hospital's organizational structure and inpatient practices in the Apulia and Emilia Romagna regions.
This correlation can offer insights into how efficiently a hospital can meet patients' needs and expectations, as well as potential discrepancies between the perceived quality of care and the actual efficiency in providing it. Understanding this relationship can help in pinpointing potential areas for improvement in hospital management and resource allocation, ultimately leading to better outcomes and increased patient satisfaction. Moreover, grasping the connection between hospital efficiency and perceived quality of care can also have significant policy implications.
This study aimed to investigate the influence of the identified hospital efficiency in both Apulia and Emilia Romagna on the hospitalization propensity of resident patients. Before applying linear regression, outliers were removed. We assessed the Spearman correlation coefficient between the target variable and the SE feature, which are not normally distributed, both for the entire hospital network in Apulia and Emilia Romagna and for the private and public hospital networks, using the "correlation" widget of the Orange software. The correlation analysis, shown in Table 28, pertains to the entire model of the Apulia and Emilia-Romagna macro-region.

Table 28. Spearman correlation coefficient between the target variable and the two features for Apulia and Emilia – Romagna regions.

NETWORK	TARGET VARIABLE	SE
PUBLIC HOSPITAL NETWORK	PROPENSION HOSPITALIZATION	- 0.863

The results of the linear regression model, illustrated in Figure 41, highlight a significant relationship between hospital scale efficiency, represented by the "SE_CPDA" variable, and the patients' "Propension Hospitalization". Examining the model's fit measures, we observe a correlation coefficient R of 0.894, indicating a strong linear relationship between the variables. The high percentage of the determination coefficient R^2 (79.9%) suggests that "SE_CPDA" explains a significant portion of the variance in "Propension Hospitalization".
The Omnibus ANOVA test revealed an F statistic of 167 with a p-value less than 0.001. This confirms the statistical significance of the model and indicates "SE_CPDA" as a relevant predictor of "Propension Hospitalization". Analyzing the model's coefficients, the intercept is 7.56, representing the predicted "Propension Hospitalization" when "SE_CPDA" is zero. The coefficient for "SE_CPDA" is -9.51, revealing a decrease of 9.51 units in "Propension Hospitalization" for every unitary increase in "SE_CPDA". This negative relationship is further supported by the standardized estimate of -0.894.
Regarding the linear regression assumptions, normality tests, such as Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling, indicate a normal distribution of the residuals. Moreover, heteroskedasticity tests, like Breusch-Pagan, Goldfeld-Quandt, and Harrison-McCabe, show no evidence of heteroskedasticity, confirming the constancy of the residuals' variance across the levels of the independent variable.
In conclusion, the results suggest a marked negative linear relationship between "SE_CPDA" and "Propension Hospitalization" in the public hospital network of the Apulia-Emilia Romagna macroregion. The fundamental assumptions of linear regression, such as the normality of residuals and homoscedasticity, are met, making the model suitable for the analysis of the provided data.

Linear Regression

Model Fit Measures

Model	R	R^2
1	0.894	0.799

Omnibus ANOVA Test

	Sum of Squares	df	Mean Square	F	p
SE_CPDA	100.1	1	100.103	167	<.001
Residuals	25.2	42	0.601		

Note. Type 3 sum of squares

Model Coefficients - PROPENSION HOSPITALIZATION

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	7.56	0.557	13.6	<.001	
SE_CPDA	-9.51	0.737	-12.9	<.001	-0.894

Assumption Checks

Normality Tests

	Statistic	p
Shapiro-Wilk	0.980	0.637
Kolmogorov-Smirnov	0.0837	0.892
Anderson-Darling	0.329	0.507

Note. Additional results provided by *moretests*

Heteroskedasticity Tests

	Statistic	p
Breusch-Pagan	2.14	0.144
Goldfeld-Quandt	1.82	0.094
Harrison-McCabe	0.402	0.190

Note. Additional results provided by *moretests*

Q-Q Plot

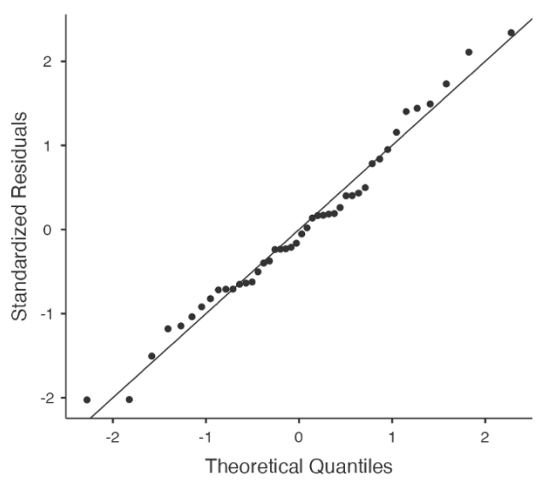


Figure 41 Linear Regression between "SE_{CPDA}" and "Propension Hospitalization" in the Apulia and Emilia Romagna Regions.

5. RESPONSES TO II SESSION RESEARCH QUESTIONS

The answers to the research question proposed in II Session is as follows:
A_4: In the Apulia-Emilia Romagna macroregion, hospital efficiency displays variability between public and private sectors and across regions. Using the CPDA methodology, notable facilities achieved optimal technical efficiency, with a standout in Apulia also reaching optimal scale efficiency. Public hospitals consistently showcased superior scale efficiency compared to private ones. Crucially, a strong negative linear relationship was identified between scale efficiency and patient's propensity for hospitalization, indicating that hospital efficiency directly influences the perceived quality of healthcare by resident patients.

6. CONCLUSIONS OF II SESSION

In the context of the detailed analysis of hospital efficiency within the Apulia-Emilia Romagna macroregion, a complex picture of performance and differences between the two regions emerged. Using the CPDA methodology, we were able to derive efficiency scores that offer a robust and comprehensive evaluation of hospital performance. While five facilities, spread between Apulia and Emilia Romagna, achieved a pure technical efficiency (PTE) score of 1, only one of these, located in Apulia, also achieved a scale efficiency score of 1, highlighting its excellence both in terms of resource transformation into outputs and optimal operational size. Delving further into regional differences, it emerged that, although there are facilities in both regions operating with optimal technical efficiency, there is a slight predominance of high-efficiency facilities in Emilia Romagna. However, when considering the private or public nature of the facilities, the differences become more pronounced, with the public hospital network showing greater scale efficiency compared to private facilities, regardless of the region.

These findings, combined with the analysis of hospitalization propensity as an indicator of the quality of healthcare perceived by patients, suggest a direct correlation between hospital efficiency and the quality of care. In particular, the negative relationship between the scale efficiency "SE_{CPDA}" and "Propension Hospitalization" underscores the importance of efficient resource management and operational size in influencing patient perception.

In summary, while both Apulia and Emilia Romagna feature hospitals with high performance, there are significant differences in terms of technical and scale efficiency, which directly influence the quality of healthcare perceived by patients. These discoveries provide valuable insights for further reflections on hospital management practices and regional health policies.

Limitations and Future Perspectives: Despite the comprehensive analysis provided, this study has inherent limitations. We relied on data that might not capture recent developments in hospital structures or healthcare within the examined regions. While the CPDA methodology is robust, it does not negate the validity or utility of other analytical approaches. Moreover, hospitalization propensity, though significant, captures only one facet of care quality. Future analyses might consider factors like hospital funding or staff expertise for a more holistic understanding. Moving forward, monitoring how efficiency and perceived quality evolve with technological innovations, regulatory shifts, and new healthcare policies is essential. Extending research to other regions or international contexts could also yield further insights and identify global best practices in hospital efficiency.

III SESSION

ASSESSMENT OF PUBLIC HEALTH PERFORMANCE IN RELATION TO HOSPITAL ENERGY DEMAND, SOCIO-ECONOMIC EFFICIENCY AND QUALITY OF SERVICES

1. INTRODUCTION

Good health is essential to sustainable development, as established by the 2030 Agenda. Indeed, health has a central position in Sustainable Development Goal 3, which concerns "Good Health and Well-being." The goal also aims to achieve universal health coverage and provide access to safe and effective medicines and vaccines for all. It is closely linked to over a dozen targets in other goals related to urban health, equal access to treatments, and non-communicable diseases, among others. In fact, it represents a unique opportunity to promote public health through an integrated approach to public policies across different sectors. Specifically, Target 3.8 "Achieve Universal Health Coverage" aims to provide universal health coverage, including financial risk protection, access to quality essential health care services, and access to safe, effective, quality, and affordable essential medicines and vaccines for all.

Subsequent paragraphs, however, are indented. Despite this awareness underpinned by ambitious goals in recent decades, there has been a steady increase in the number of disasters, including pandemics, which have had significant impacts on societies and economies. The COVID-19 pandemic has shown that many countries around the world have been caught unprepared to deal with such a threat, and the safety net provided by the health infrastructure has failed even in the most developed countries, with considerable fallout and repercussions on the health sector in general. This has led to an increase in healthcare costs, with forecasts of a further upward trend, without an integrated system for monitoring the efficiency of the healthcare system in general and the quality of services in particular.

Therefore, the interest of academics, healthcare managers, and policymakers has increased in identifying measures to contain healthcare spending while ensuring service quality. In this context, various attempts have been made to improve provider efficiency, including activating competition between hospitals and the implementation of incentive-based payment systems. Proposed models aimed at maximizing public administration and social housing have demonstrated effectiveness in meeting public needs while ensuring fair compensation for private entities (Morano et al., 2021). Numerous studies have been conducted to assess hospital efficiency and its variation over time, in order to provide an accurate estimate of hospital productivity and costs, which can be used as a criterion for payment for hospital services and to improve national health provision. As a matter of fact in many countries, public policies concerning the reduction of beds and medical and nursing staff, hospital mergers and acquisitions, and lower investments in health infrastructure are now being reassessed. In addition, the decentralised organisation of the healthcare system is also being questioned, and in many cases, the re-centralisation of the system is being considered. Italy is one of the Western countries that has significantly reduced healthcare spending and decentralised management to the regional level. However, the effects of these policies are still being debated. Moreover, despite the changes in healthcare organisation, inequalities between regional systems have not decreased over the last 20 years. In

the literature, there is still a broad consensus that these decentralisation plans, and the decentralisation process in general, have had an impact on increasing health inequalities not only between but also within regions.
Moreover nowadays, healthcare managers must consider the impact of exogenous economic factors, such as the progressive increase in the cost of energy resources throughout Europe, an increase that is even more relevant following the outbreak of the conflict in Ukraine and the inflationary spiral that is still in progress. According to the National Agency for Regional Health Services (AGENAS), funding of EUR 1.6 billion has been provided for the National Health Service entities for the year 2022 to counter the effects of the increase in the prices of energy sources.
Given this evidence, the assessment of the territorial health services usually crosses several dimensions to obtain a comprehensive composite indicator useful for classification and comparisons. Therefore the first contribution of the present work is to identify and measure the main latent variables that summarize the organizational components of public nosocomial facilities, as well as variables that can express patients' preferences for hospital choice based on best activity or outcome criteria. The second contribution of this paper presents a machine learning methodology using machine learning algorithms that can assist decision-makers in their choices. The third contribution explains the interaction between identified hospital components and per capita health expenditure on electricity through a linear regression model.
The third section is organized as follows: after introducing the problem and the objectives of the work, the following paragraph presents the methodological background and the case study to which it was applied. This is followed by a paragraph detailing the conducted study, where the most interesting results are presented and subsequently critically discussed, complemented by an evaluation of the proposed approach's innovativeness, potential, and limitations. The conclusions offer insights into the implications of the methodology for decision support at various scales and suggest possible future developments.

2. BACKGROUND

In the literature, it has been observed that various demographic and economic factors play a role in the choice of healthcare facilities, such as income, propensity to travel, level of education, age, type of illness, need for frequent treatment, and trust and reputation of the facility and its operators.
The patient's decision-making process when considering the quality of services (real or perceived) can be divided into several stages, including information gathering, risk/benefit assessment, consultation with the physician, and choice of treatment. The patient experience is complex and depends on several factors, including satisfaction, quality of care, and effectiveness of the healthcare system (Wolf et al., 2021). The perceived quality of healthcare is related to the actual quality of care provided (Doyle et al., 2013). Patients' satisfaction with the healthcare system is influenced by their experience of care and their perception of adequate attention to their care needs (Bleich et al., 2009). Active patient involvement in treatment choice improves patient satisfaction (Shay & Lafata, 2015). From the perspective of hospital facilities, a significant factor for perceived quality is the facility's ability to effectively treat complex and specialized illnesses, i.e., the facility's specialization and the complexity of the clinical cases treated. Another essential aspect is the reputation of the doctors working there. These

factors can be emphasized by marketing policies pursued (Falavigna & Ippoliti, 2013).
According to the literature, two main points of view need consideration when evaluating possible measures: social welfare and stakeholder theory. Hospitals should not only ensure high-quality medical services at reasonable costs to improve health in society, but also be concerned about the well-being of their customers and all other stakeholders involved in the process (Hajiagha et al., 2022).
Based on these assumptions, hospital management should be aware of all the variables that express the structure's human, financial, and technological resources, as well as the outcomes produced and the quality of life of patients. To obtain timely decisions by healthcare management with the aid of streamlined procedures and tools, methodologies based on Multi-Agent Simulation to support Decision-Making in healthcare infrastructures for the organizational management and actionable choices in health risk (Esposito et al., 2020), as well as methods based on Dynamic Network Visualization of space use to support spatial redesign related decisions to improve workflow effectiveness and patient well-being (Esposito & Abbattista, 2020), have been proposed.
Artificial intelligence-based approaches composed of optimization and machine learning (Mirmozaffari et al., 2022) have been conducted and applied in different fields and organizations to calculate, for example, public hospital efficiency (Hajiagha et al., 2023), rather than in the industrial and service sectors (Guede-Cid et al., 2021).
As such methodologies involve large volumes of data, they require data mining techniques such as feature extraction, selection, and classification to derive meaningful information from the data. Feature selection is a technique used to reduce dimensionality to prune the feature space and, consequently, reduce computational cost and improve classification accuracy by means of Principal Component Analysis (Alomari et al., 2022).
Section Three will present an extension of the work by Santamato et al., 2023, applying the 2020 data from the Apulia Region described in Section One and the 2021 data from the Apulia Region in Section Two. A linear regression analysis will be conducted between the two main components identified in the previous two sections (Hospital Organization and Propensity for Hospitalization) and the hospital energy cost. Subsequently, an ANOVA analysis of the predictive results will be carried out. In a secondary analysis, a linear regression will be applied between the efficiency scores identified in the previous sections, the number of hospital devices, and the cost of hospital energy.
The analyses carried out in this study will be conducted in a machine learning environment.
This session presents a case study of the Apulia region (Italy). The regional health system encompasses both public and private accredited facilities within a given region, forming an organized complex known as the regional health industry (Falavigna & Ippoliti, 2013).
The Regional Health Service in Apulia is represented by six Aziende Sanitarie Locali (ASLs), as shown in Table 1a. ASLs are public legal entities with autonomy in organizational, managerial, technical, administrative, patrimonial, and accounting matters, as well as entrepreneurial autonomy (in accordance with Article 3 of Legislative Decree No. 502 of 30 December 1992). ASLs are part of the National Health Service.
This study focuses on the regional public hospital network in Apulia, specifically analyzing 28 facilities as indicated in the National Health Service Data Bank of the

Ministry of Health. The public network consists of 24 ASL Direct Hospitals, one Hospital Authority integrated with the National Health System (NHS), one Hospital Authority integrated with the University, and one Public Institute for Hospitalization and Scientific Care. Some facilities, as specified in the single hospital reorganization document approved by the Apulian Regional Council on 03/07/2019, have connected plexuses and/or hospitals located in different places from the main structure. Thus, for greater accuracy in measuring the distance traveled by patients, all the physical plexuses indicated in the National Outcomes Plan have been considered. The total number of facilities considered for measuring the active mobility of Apulian patients is 37 (see Figure 42), including admissions made by ASL and by territorial ambit, as indicated in the National Outcomes Plan.

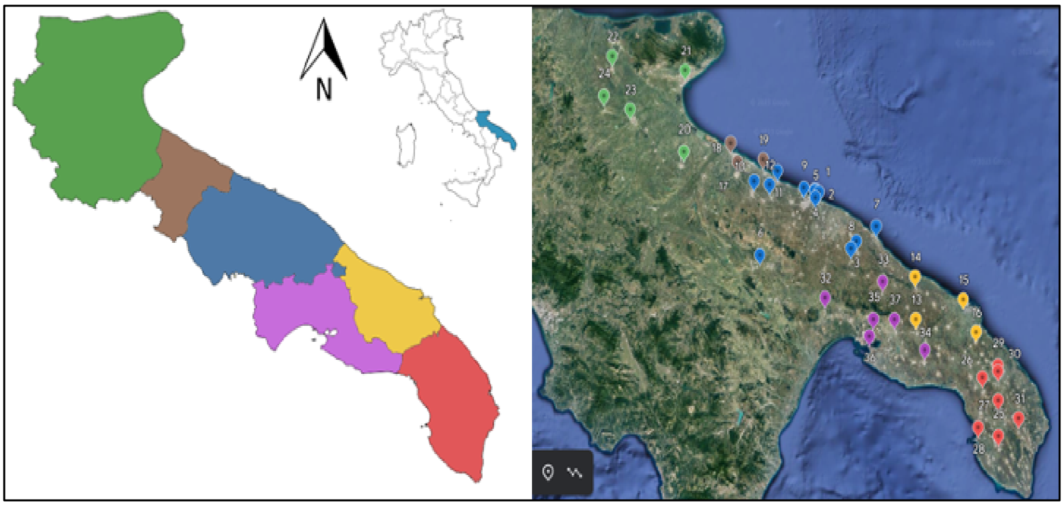


Figure 42 Division of regional territory into provinces and public hospitals in Apulia.

In terms of hospital classification in Apulia, there are 5 second-level hospitals, 12 first-level hospitals, 2 Scientific Hospitalization and Care Institutes (IRCCS), and 9 basic hospitals. The differences between these types of hospitals are described in the Ministry of Health's regulations on the definition of qualitative, structural, technological, and quantitative standards for hospital care, which are implemented under Article 1, paragraph 169, of Law no. 311 of 30 December 2004 and Article 15, paragraph 13, letter c) of Decree-Law no. 95 of 6 July 2012, converted with amendments by Law no. 135 of 7 August 2012. The hospital levels differ primarily based on catchment area, number of wards, and complexity.

2.1 KEY RESEARCH QUESTIONS ADDRESSING ENERGY COSTS AND HOSPITAL EFFICIENCY IN APULIA

In the context of our investigation into the energy and management efficiency of hospitals in Apulia, we identified several fundamental questions that required further exploration. These questions became the backbone of our research and guided our subsequent analyses:

Q_5: How do the variables of hospital organization and patients' propensity for hospitalization correlate with the per capita energy cost in public health facilities in Apulia?

This question aims to explore the dynamics between managerial and organizational decisions within hospital facilities and their impact on energy costs. Understanding

this relationship is crucial for formulating recommendations on how to optimize resources without compromising the quality of care.
Q_6: What is the relationship between the pure technical efficiency (PTE) of public hospitals in Apulia and the number of medical devices, in relation to the per capita energy cost?
With the rise of medical technology and the importance of medical devices in care provision, it is essential to understand how these factors influence energy costs. This understanding can offer insights into how to balance the adoption of new technologies with energy sustainability.
These research questions were posed in light of the preliminary findings and trends observed in the initial stages of our analysis. The answers to these questions are vital to provide concrete recommendations to policymakers and hospital managers on how to address the challenges of energy management in the healthcare sector.

3. METHODOLOGY

3.1 METHODOLOGICAL APPROACH TO HOSPITAL AND ENERGY EFFICIENCY

In an era marked by the urgent need to address climate change and pursue energy efficiency, the strategy with which hospital resources are managed becomes of paramount importance. Our study, drawing on a systematic review of the literature, identified a range of crucial variables for assessing hospital efficiency based on data sourced from the National Health Service and the National Outcomes Plan. These variables span a range of dimensions, from human resources, capacity, and productivity to quality of care, length of stay, and patient satisfaction. Our analysis methodology, detailed in paragraph 1.17 of SESSION I, is based on Principal Component Analysis (PCA). Through this, variables were segmented into two main categories: hospital organization and propensity for patient admission. The analysis was conducted using the Orange software, known for its advanced data mining capabilities as outlined in paragraph 1.9 of SESSION I. The depth and breadth of this analysis, exploring the intersection between energy efficiency, human resource management, and hospital performance, are extensively discussed in the work by Santamato et al., 2023.
Table 29 presents the variables in each group.

Table 29. Organization variables of Apulia Public Hospital.

Organization variables	Definition	Reference No.	Data sources
Var_1	No. of day hospital beds	(Santamato et al., 2023)	NHS Database
Var_2	No. of day surgery beds		
Var_3	No. of beds in ordinary hospitalization		
Var_4	No. of beds used		
Var_5	No. of departments used		
Var_6	Total no. of physicians		
Var_7	Total no. of nurses		
Var_8	Total no. of hospital staff		

In the second group, we have included outcome variables that express the hospital's performance in terms of services provided and outcomes produced, rather than active mobility and thus the attractiveness of the facility (Table 30). The combination of these factors will contribute to the overall perceived quality of care by patients.

Table 30. Outcome variables of Apulia Public Hospital.

Outcome variables	Definition	Reference No.	Data sources
Var_9	No. of hospitalizations	(Santamato et al., 2023)	NHS Database
Var_10	Intra-regional mobility active by territorial scope		National Outcomes Plan
Var_11	No. of deaths at 30 days after hospitalization		
Var_12	No. of interventions according		
Var_13	No. of hospital readmissions at 30 days after hospital discharge		
Var_14	No. of inpatient days		NHS Database
Var_15	No. of available days indicated		
Var_16	No. of surgical discharges		

The methodology proposed in this paper involves a first phase of applying Principal Component Analysis (PCA) to the initial dataset. The dataset will be reduced in size by applying two distinct PCAs to the two groups of variables identified. The Principal Component for the first group will express Hospital Organisation, while the Principal Component for the second group will express Patient Admission Propensity.

The effects of climate change, along with the recent energy crisis, have brought energy efficiency issues in hospitals and the increased demand for more research on energy efficiency in buildings into the spotlight (Psillaki et al., 2023). Therefore, we included an additional study variable: the per capita cost of medical electricity per Apulian resident in relation to the number of physicians in the year 2020. This variable was calculated as the ratio of the resident population of the Apulia Region in the year 2020 divided by the number of physicians in each hospital (Gutierrez-Romero et al., 2021), multiplied by the value of the per capita health expenditure related to the energy costs of Apulia in the same year (AGENAS).

The methodological workflow, shown in Figure 43, is a graphical representation of the complex analyses applied in this study using data analysis and machine learning tools, via the widgets available in the Orange software.

We chose to use the Orange software for our analyses because it represents a robust data mining tool (Mirmozaffari et al., 2022).

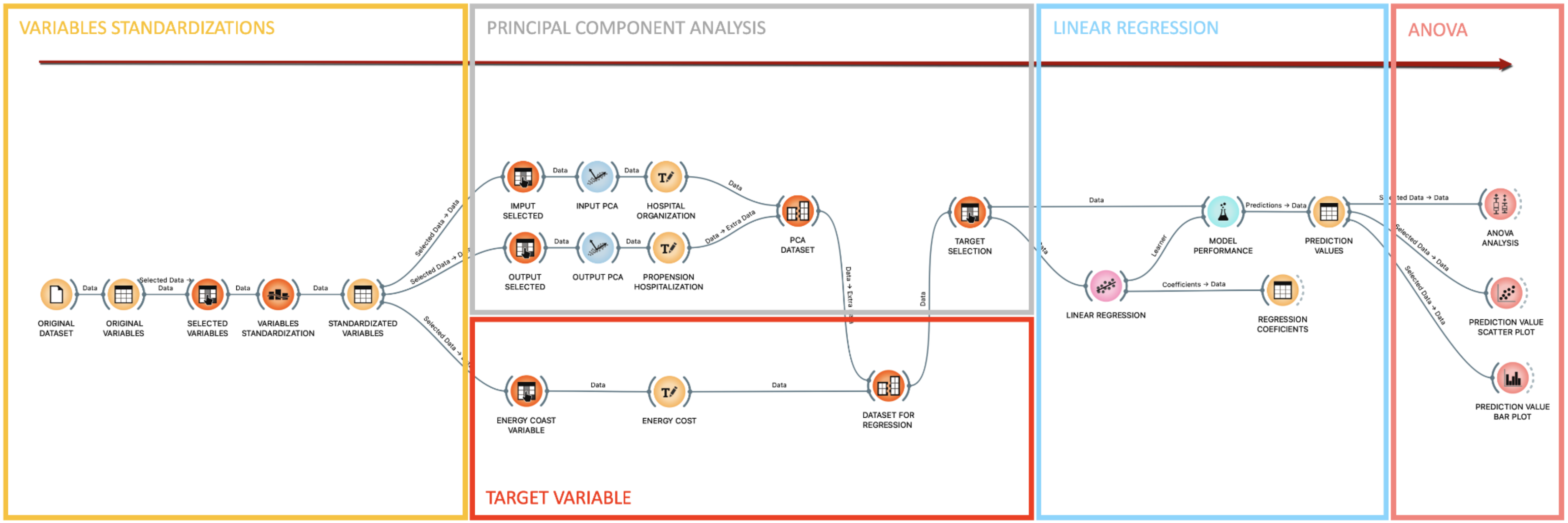


Figure 43. Methodological workflow implemented in Orange.

The new dataset, which includes the two identified principal components and the value of energy expenditure per capita of the population distributed by doctor, will be subjected to training and testing by means of 10-fold cross validation. Linear regression will be used for this process. The learning algorithm (regressor) will process the input dataset and produce as output a prediction model capable of identifying the value of healthcare energy expenditure for given values of hospital organization and propensity to hospitalization of patients.

3.2 PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is a machine learning technique that reduces the complexity of a dataset by transforming a set of original variables into a new set of linearly independent variables, called principal components. This method simplifies the representation of the data while retaining the most important information.

By reducing the size of the original variables, it is optimized while preserving as much information as possible. The transformation of the data takes place in a new coordinate system, where the new variables are orthogonal and arranged in order of importance (Hastie et al., 2009). PCA was used as a pre-processing phase of the data in a machine learning environment. We used Orange's select column widgets to split the initial dataset into the two groups of Input Variables (Table 24) and Output Variables (Table 25). We then linked the two groups to the PCA widgets, as illustrated in Figure 44.

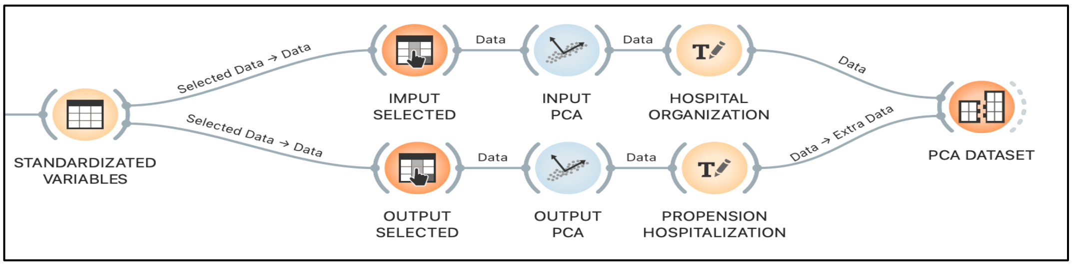


Figure 44. PCA workflow.

Applying PCA to the first group of identified variables (Table 24), we obtained a principal component (Input_PC1) preserving almost 90% of the total variance with minimal loss of information. The graphical representation is shown in Figure 45.

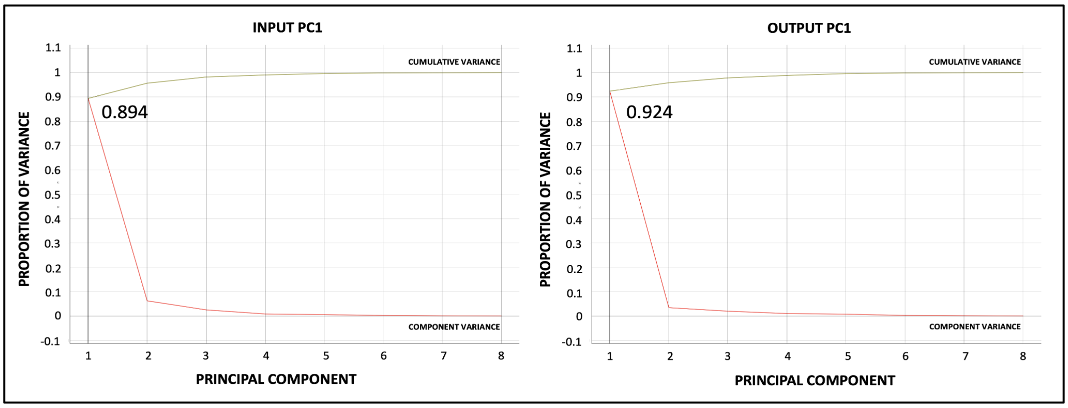


Figure 45. Input PC1 and Output PC1 variance representation.

Applying PCA to the second group of identified variables (Table 25), one main components (Output_PC1) was identified, preserving almost 93% of the total variance. The graphical representation is illustrated in Figure 46.
The incidence of the individual original variables on the main components can be visualised by means of the data table widget. The results produced and represented in figure 6 for the first principal component show a distribution in terms of incidence, which is fairly homogeneous across all variables and therefore PC_1 was renamed Hospital Organisation.

Data instances: 1
Features: 8
Meta attributes: 2

	components	variance	VAR_1	VAR_2	VAR_3	VAR_4	VAR_5	VAR_6	VAR_7	VAR_8
1	PC1	0.894017	0.331244	0.290424	0.368028	0.361608	0.367425	0.369782	0.369161	0.362916

Figure 46. Incidence of Input variables on Input PC1.

The main output component was renamed Hospitalisation Propensity. The results are illustrated in figure 47.

Data instances: 1
Features: 8
Meta attributes: 2

	components	variance	VAR_9	VAR_10	VAR_11	VAR_12	VAR_13	VAR_14	VAR_15	VAR_16
1	PC1	0.923857	0.365744	0.338618	0.33546	0.3532	0.349674	0.364054	0.359203	0.36118

Figure 47. Incidence of Output variables on Output PC1.

3.3 MACHINE LEARNING ALGORITHM

Linear regression is a statistical model that attempts to establish a linear relationship between a dependent variable (target) and one or more independent variables (features) (Fang & Lahdelma, 2016).
The linear regression model produces a linear function that attempts to predict the value of the dependent variable based on the values of the independent variables.
The model in multiple linear regression consists of more than one predictor variable:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P + \varepsilon \qquad (22)$$

where Y is the response variable, $X_1; X_2; \dots X_P$ is the predictor variables with p as the number of variables, $\beta_0;\ \beta_1;\ \beta_2 \cdots \beta_P$ are the regression coefficients, and ε is an error to account for the discrepancy between predicted data and the observed data (Fumo & Rafe Biswas, 2015).

3.4 TARGET VARIABLE

Sustainability issues have become fundamental in all their various environmental, social and economic facets. One of the main challenges to be overcome by hospitals in this regard is energy management, based on environmental sustainability, which is used as a strategic means to achieve competitiveness and focuses on energy efficiency that includes policies, strategies and technologies designed to reduce energy consumption, pollutant gas emissions and costs (Borges de Oliveira et al., 2021).

Therefore, efficient energy management in hospitals has the potential to improve energy efficiency on the one hand and better management of public expenditure on health energy on the other.

Based on this assumption, in this study the per capita expenditure of the Apulian population in the calendar year 2020, distributed by hospital doctors, was identified as the target variable.

To calculate the cost of energy, we first calculated the ratio between the resident population in Apulia in 2020 ($Tot_{Res_{ASL}}$) and the total number of doctors in both public and private accredited hospitals, for each ASL ($Tot_{Doctors_{ASL}}$).

We then multiplied this ratio by the number of doctors in each public hospital, weighted by ASL ($Doctors_{Hospi}$) The resulting value for each facility, which is an expression of the population catchment area, was then multiplied by the per capita health energy cost of € 21.45 for the Apulian population in the year 2020, as indicated by the National Agency for Regional Health Services (AGENAS).

$$Energy\ cost\ = \left(\frac{Tot_{Res_{ASL}}}{Tot_{Doctors_{ASL}}} \right) * Doctors_{Hospi} * 21.45 \qquad (23)$$

4. DISCUSSION AND RESULTS

It describes the second phase of the study, in which the target variable is "Per-capita energy cost" and the features are "Hospital Organisation" and "Hospital Propensity".

The Linear regression widget is used to provide the prediction algorithm with the dataset containing the variables to be analysed, and the performance of the model is evaluated using the Test and scores widget with a cross-validation of 10 folds. The results of the evaluation are described in Figure 48.

Sampling type: 10-fold Cross validation

Scores

Model	MSE	RMSE	MAE	R2
LINEAR REGRESSION	0.06876586014125485	0.26223245440115694	0.19946152373923498	0.9312341398587451

Figure 48. Prediction models performances.

The linear regression model has an MSE of 0.06, an RMSE of 0.2 and an MAE of 0.19, which indicates that the mean prediction error is relatively low. Furthermore, the R2 of 0.93 suggests that the model explains about 93% of the variance in the data, indicating a good level of fit.

The coefficients of the regression model for the 28 public hospitals in Apulia are described in Figure 49.

Data instances: 3
Features: 1
Meta attributes: 1

	name	coef
1	intercept	2.41838e-16
2	HOSPITAL ORGANIZATION	0.170574
3	PROPENSION HOSPITALIZATION	0.19381

Figure 49. Regression model coefficients.

The dataset resulting from the analysis model proposed in the study, with the relevant predictive values generated by the linear regression, for the 28 public hospital facilities in the Apulia Region, is shown in figure 50.

Data instances: 28
Features: 2
Meta attributes: 3
Target: Numeric variable 'ENERGY COST'

	ENERGY COST	STRUCTURE NAME	LINEAR REGRESSION	Fold	HOSPITAL ORGANIZATION 0.894017	PROPENSION HOSPITALIZATION 0.923857
1	-0.745769	OSPEDALE OSTUNI	-0.665551	1	-1.86611	-1.78255
2	-0.460426	OSPEDALE FRANCAVILLA FONTANA	-0.583272	1	-1.6726	-1.52833
3	-0.378578	OSPEDALE CIVILE MARTINA FRANCA	-0.371675	1	-1.25215	-0.805925
4	-0.469693	OSPEDALE DELLA MURGIA - PERINEI	-0.365568	2	-0.851991	-1.1361
5	0.100588	OSPEDALE BARLETTA - 'MONS. R. DIMICCOLI'	-0.103025	2	-0.304612	-0.268841
6	-0.424145	OSPEDALE SCORRANO	-0.30743	2	-1.25587	-0.508038
7	0.29843	OSPEDALE ANDRIA - 'L. BONOMO'	-0.116555	3	-0.483352	-0.117045
8	1.6364	AO UNIV. 'OO RR FOGGIA'	1.91123	3	5.22458	5.05528
9	-0.502842	OSPEDALE CASARANO	-0.408186	3	-1.16271	-1.01687
10	0.761344	OSPEDALE BARI 'SAN PAOLO'	0.626752	4	1.83688	1.67788
11	-0.417241	OSPEDALE BISCEGLIE	-0.6698	4	-1.46354	-2.06907
12	-0.600637	OSPEDALE SAN SEVERO - TERESA MASSELLI	-0.45893	4	-1.24921	-1.22612
13	-0.687382	OSPEDALE MONOPOLI	-0.484848	5	-1.23629	-1.48059
14	0.151812	OSPEDALE BARI "DI VENERE" - TRIGGIANO	0.23771	5	0.213875	1.13649
15	3.22886	AO UNIV. CONS. POLICLINICO BARI	3.73946	5	10.0078	8.73097
16	-0.813643	OSPEDALE PUTIGNANO	-0.66868	6	-1.52997	-1.94286
17	-0.843701	OSPEDALE MANFREDONIA	-0.811186	6	-1.84558	-2.34117
18	1.35276	OSPEDALE LECCE - 'V FAZZI' (SAN CESARIO)	1.89555	6	3.32307	5.42516
19	-0.523552	OSPEDALE GALATINA 'S. CATERINA NOVELLA'	-0.667666	7	-1.35007	-2.06443
20	-0.522171	OSPEDALE MANDURIA 'GIANNUZZI'	-0.633389	7	-1.5107	-1.79218
21	-0.322481	OSPEDALE CASTELLANETA	-0.546519	7	-1.22505	-1.59083
22	-0.832146	IRCCS 'SAVERIO DE BELLIS'	-0.702665	8	-1.9766	-1.90177
23	-0.638146	OSPEDALE COPERTINO	-0.60114	8	-1.70254	-1.5926
24	1.92981	PRESIDIO OSPEDALERO CENTRALE TARANTO	1.13679	8	3.11079	3.55219
25	1.40147	OSPEDALE BRINDISI 'PERRINO'	1.18844	9	3.20641	3.40553
26	-0.70165	OSPEDALE CERIGNOLA 'G.TATARELLA'	-0.493111	9	-1.08998	-1.65949
27	-0.560034	ISTITUTO TUMORI GIOVANNI PAOLO II	-0.454713	10	-1.25174	-1.30056
28	-0.417241	OSPEDALE GALLIPOLI 'SACRO CUORE DI GESU'	-0.265152	10	-0.642778	-0.858102

Figure 50. Hospital component scores.

In particular, the coefficient of the variable 'Company organisation' is 0.17. This means that an increase of one unit in the variable 'Company organisation' is associated with an increase of approximately 0.17 units in the target variable 'Per capita expenditure on health energy'. Thus, an increase in business organisation (e.g. greater efficiency or better resource management) is associated with an increase in per capita expenditure on health energy.
Similarly, the coefficient of the variable 'propensity to hospitalise' is 0.19. This means that an increase of one unit in the variable "Propensity to hospitalise" is associated with an increase of about 0.19 units in the target variable "Per capita expenditure on health care energy". Thus, an increase in the propensity to hospitalise (e.g. an increased need for hospital care) is associated with an increase in per capita expenditure on health care energy.

The model intercept represents the value of the target variable 'Per capita expenditure on health energy' when all other variables in the model are zero. In practice, the intercept represents a kind of "expenditure floor" that the model predicts even in the absence of changes in the other independent variables.
Hospitals in their business complexity and their economic, social and environmental importance are large consumers of energy due to the continuous operation involving the use of complex electronic equipment to support clinical procedures.
A better management of healthcare resources in terms of a greater workforce, i.e. more hospital staff, rather than an increase in instrumental resources, beds and/or expansion of wards, which would contribute to an increase of an overall unit of the hospital corporate organisation, implies an increase in per capita healthcare expenditure on energy of 17%. From a managerial point of view, an assessment for a potential investment in corporate organisation would imply the same percentage increase for the energy cost item.
Analysing, at the same time, the hospital outcome factor counterbalancing the economic one, it can be seen that the relationship between the propensity to hospitalise patients and the relative increase in energy costs is quite robust. The final health outcomes understood as the reduction of discomfort, the prolongation of life, the decrease in the incidence of disease, rather than the satisfaction of users, family members, and the general population with the perceived overall quality and various aspects of care, are translated into the propensity to hospitalise patients. A propensity that is an expression of a need for care that sees its own unit increase associated with a 19% increase in the cost of energy.
A choice of resource allocation in the short/medium term can be optimised with the methodological approach proposed in this study, considering, on the one hand, the relationship between the hospital components and the cost of health care energy, but on the other taking note of the annual increase in the cost of health care energy per capita, which has increased in the three-year period 2020 - 2023, by about 142.65%, going to affect the regional health budget (AGENAS).
Observing the results of our linear regression analysis, represented graphically by means of the bar plot widget, by ASL and by hospital facility denomination (Fig. 51), it can be seen that the five regional public 2nd level facilities and two 1st level hospitals of ASL BA, have outliers with respect to the overall distribution of the regression values. These results are confirmed by the scatter plot (figure 52) distributing the predictive values by different type of level.

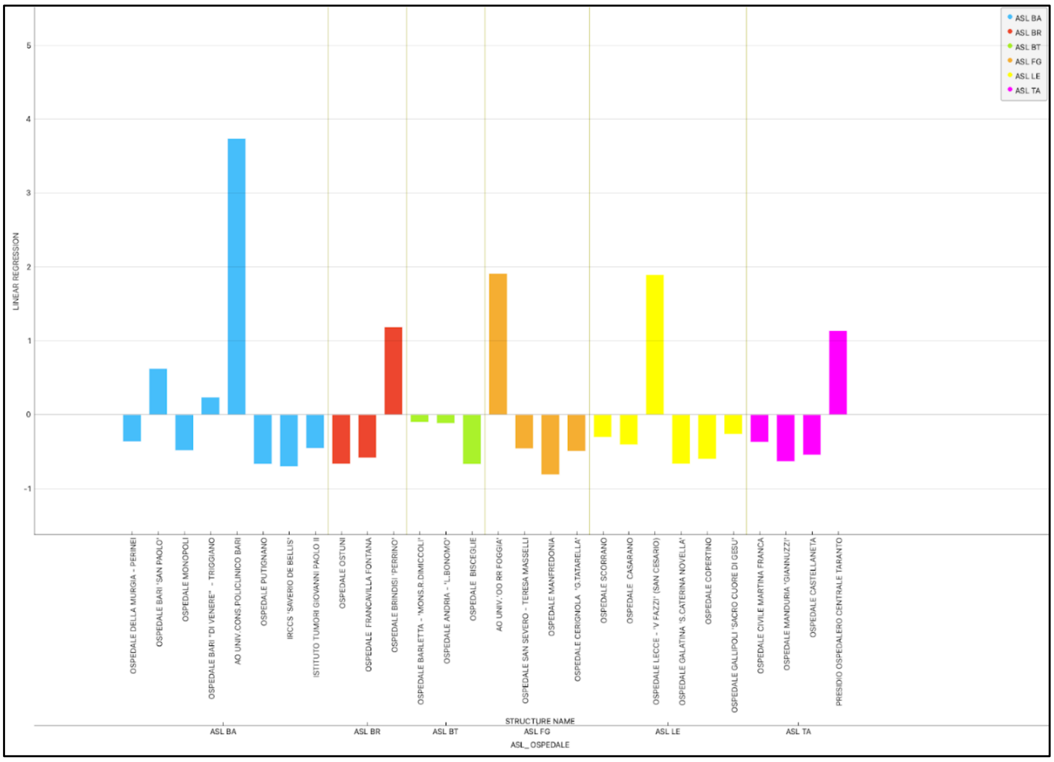


Figure 51 Bar plot of linear regression predictive values.

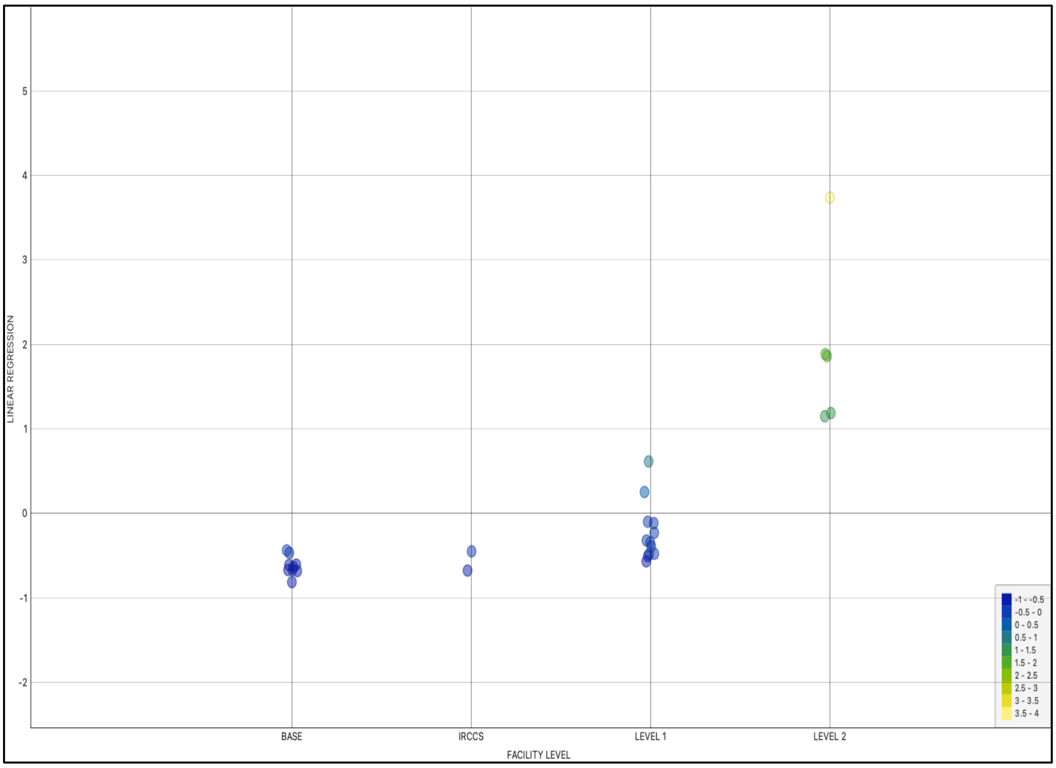


Figure 52 Scatter plot of linear regression predictive values.

This assumption is confirmed by the ANOVA analysis applied to the same predictive regression values. With a p-value of <0.005, the analysis confirms a significant difference between the different levels of the 28 public hospitals. As

depicted in Figure 53, there is a difference between the macro group consisting of basic, first level, and IRCCS hospitals compared to the second level hospitals.
The "LINEAR REGRESSION" column in Figure 10 provides the estimated regression coefficients for each facility. For basic hospitals, IRCCSs, and level 1 hospitals, the estimated regression coefficients indicate that for a unit increase in hospital organization and patient admission propensity, there is an associated decrease in energy cost.
The opposite situation is verified for second level hospitals, in which an increase in hospital organization and an increase in the propensity to hospitalize will produce a higher energy cost.

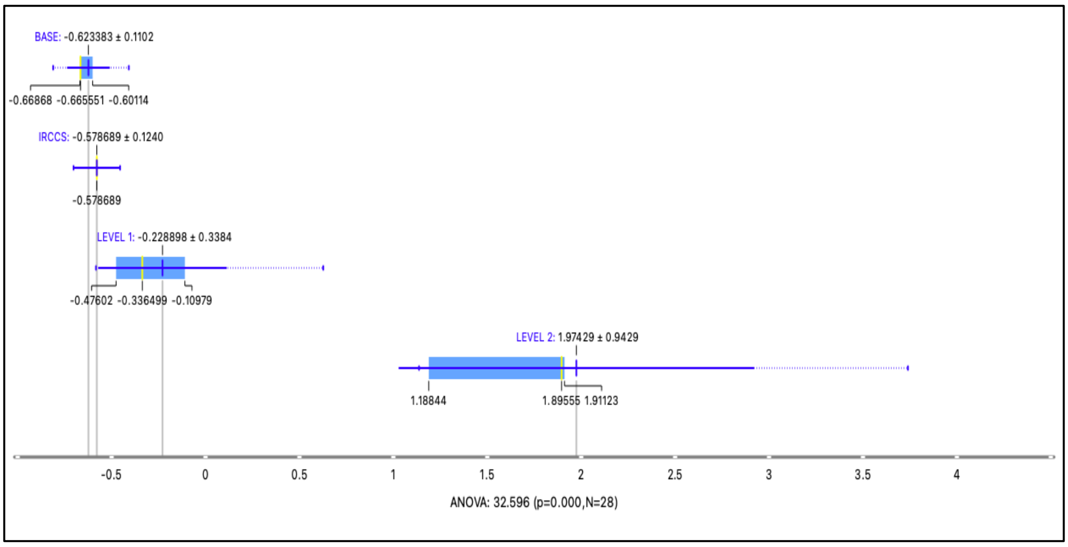


Figure 53 Anova analysis applied to predictive regression values.

It is clear that there is no homogeneity in energy management practices among all hospitals. The management of healthcare companies should develop guidelines aimed at promoting a change in the organizational culture, by creating an energy consumption management plan following the ISO 50001 guidelines and prioritizing the acquisition of alternative or renewable energy. Additionally, they should focus on designing, constructing, and managing hospital buildings with a focus on energy efficiency and developing energy-related social responsibility programs (Borges de Oliveira et al., 2021).
The study found that an increase in organizational efficiency is associated with an increase in energy costs, while an increase in patient hospitalization rates is associated with an increase in energy costs. The analysis also highlighted some exceptions among hospital structures, with some showing a higher energy cost per capita than the average. The research suggests that better management of human resources could be more effective in reducing energy costs than purchasing new equipment or expanding structures. Furthermore, the study emphasized that energy costs have been increasing in recent years and that resource allocation choices must consider these rising costs.
Policies and decisions made by policymakers should aim to incentivize the public hospital network on the quality of services offered and not solely on economic productivity derived from DRGs, achieving a dual optimal allocation. An efficient allocation of economic resources that at the same time promotes an efficient redistribution of regional admissions offers optimal outcomes in terms of perceived quality.

4.1 DEEP DIVE: RELATIONSHIP BETWEEN PURE TECHNICAL EFFICIENCY, MEDICAL DEVICES, AND PER CAPITA ENERGY COST

A key component of our study revolves around the exploration of the relationship between hospitals' Pure Technical Efficiency (PTE), the number of medical devices, and the per capita energy cost. Using a regression model, we examined how these factors impact the per capita energy cost.

From the regression model, we derive an R^2 coefficient of 0.797, indicating that the model explains approximately 79.7% of the variance in the per capita energy cost. The correlation coefficient R of 0.893 suggests a strong positive relationship between the model's variables and the per capita energy cost.

Examining the model coefficients, we note that the PTE score has a positive effect on energy cost with an estimated coefficient of 0.270 ($p = 0.037$). This indicates that a unitary increase in the PTE score is associated with a 0.270 unit increase in the per capita energy cost. Similarly, the number of medical devices has an even more pronounced effect, with an estimated coefficient of 0.688 ($p < 0.001$), suggesting that an increase in the number of medical devices leads to a significant rise in the per capita energy cost.

Normality tests, including Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling, suggest that the model residuals follow a normal distribution. Additionally, heteroskedasticity tests, such as Breusch-Pagan, Goldfeld-Quandt, and Harrison-McCabe, suggest the residuals' variance is constant across the levels of the independent variables.

In summary, the findings suggest that while adopting advanced medical devices may enhance care quality, it can also lead to increased energy consumption. Thus, hospitals need to strike a balance between adopting advanced technologies and managing energy resources efficiently, also taking technical efficiency into account.

Refer to Figure 54 for a comprehensive graphical representation of the regression model outcomes.

Linear Regression

Model Fit Measures

Model	R	R²
1	0.893	0.797

Model Coefficients - COSTO_ENERGIA_PRO_CAPITE_2020

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	-3.72e–17	0.0902	-4.13e–16	1.000	
Score_PTE	0.270	0.1225	2.20	0.037	0.270
MEDICAL_DEVICE	0.688	0.1225	5.61	<.001	0.688

Assumption Checks

Normality Tests

	Statistic	p
Shapiro-Wilk	0.959	0.337
Kolmogorov-Smirnov	0.186	0.254
Anderson-Darling	0.581	0.118

Note. Additional results provided by *moretests*

Heteroskedasticity Tests

	Statistic	p
Breusch-Pagan	2.81	0.245
Goldfeld-Quandt	0.817	0.628
Harrison-McCabe	0.549	0.624

Note. Additional results provided by *moretests*

Q-Q Plot

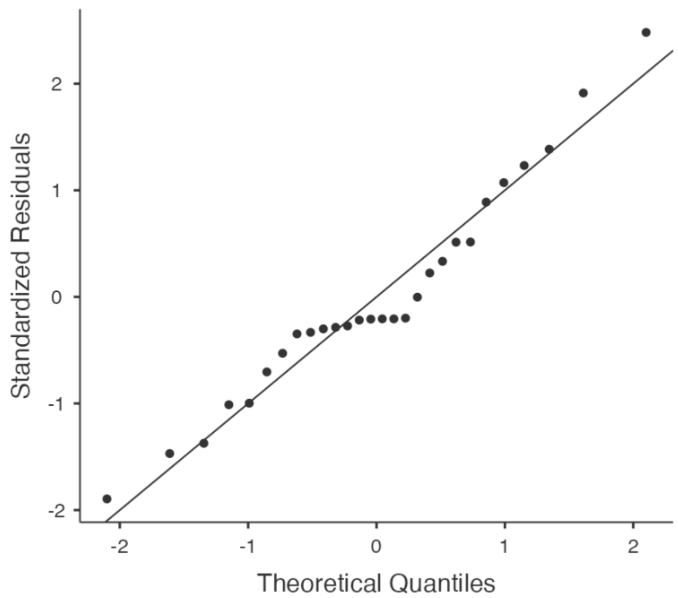


Figure 54 Regression results linking Technical Efficiency, medical devices, and per capita energy cost.

4.2 IMPACTS OF TECHNICAL EFFICIENCY AND MEDICAL DEVICES ON ENERGY COSTS: AN ANOVA ANALYSIS

In advancing our investigation, we adopted an additional methodological step by normalizing the predictive values through the NORM function. This procedure was adopted to ensure the data were on a common scale, facilitating further statistical analyses.

Normality tests, like Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling, suggest the normalized predictive values follow a normal distribution, making ANOVA analysis suitable. The ANOVA analysis revealed significant differences between hospital structure levels concerning the normalized predictive values, with an F-value of 5.48 and a p-value of 0.005.

Further tests on variance homogeneity, including Levene's and Bartlett's tests, indicate homogeneity in the variances across the groups. This is a key assumption for ANOVA analysis and indicates the analysis yields valid results.

Post-hoc comparisons based on the estimated marginal means reveal some key differences between hospital structure levels. Specifically, there's a significant difference between base hospitals and second-level hospitals, with a mean difference of -2.557 and a p-value of 0.004. Furthermore, a significant difference was detected between first and second-level hospitals, with a mean difference of -2.184 and a p-value of 0.011. This suggests that, even when accounting for technical efficiency and the number of medical devices, there's a significant variation in the per capita energy costs across these different hospital structure levels.

In conclusion, the ANOVA analysis provides further evidence of the complexities in the relationships between hospital efficiency, medical device availability, and energy costs. While efficiency and technology play roles, the hospital structure level also can have a significant impact on energy costs.

For a detailed visualization of the ANOVA results and post-hoc comparisons, please refer to Figure 55.

ANOVA

ANOVA - NORM_PREDICTIONS

	Sum of Squares	df	Mean Square	F	p
FACILITY LEVEL	23.5	3	7.84	5.48	0.005
Residuals	34.3	24	1.43		

[3]

Assumption Checks

Homogeneity of Variances Tests

	Statistic	df	df2	p
Levene's	0.0485	3	24	0.985
Bartlett's	0.127	3		0.988

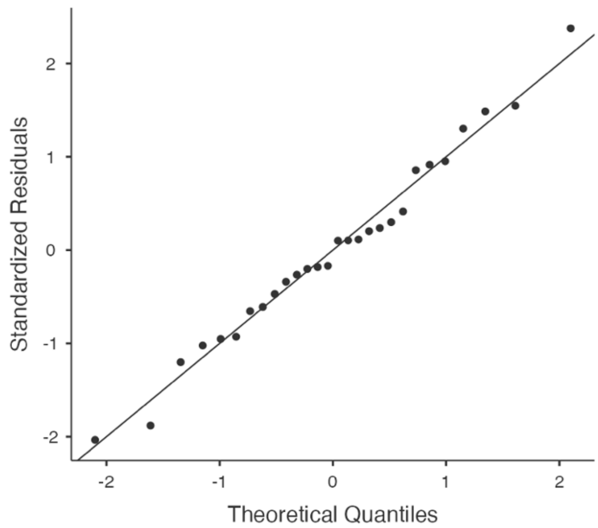
Note. Additional results provided by *moretests*

Normality tests

	statistic	p
Shapiro-Wilk	0.985	0.953
Kolmogorov-Smirnov	0.103	0.901
Anderson-Darling	0.209	0.847

Note. Additional results provided by *moretests*

Q-Q Plot



Post Hoc Tests

Post Hoc Comparisons - FACILITY LEVEL

Comparison							
FACILITY LEVEL		FACILITY LEVEL	Mean Difference	SE	df	t	p_{tukey}
BASE	-	IRCCS	-1.237	0.935	24.0	-1.322	0.558
	-	LEVEL 1	-0.373	0.527	24.0	-0.707	0.893
	-	LEVEL 2	-2.557	0.667	24.0	-3.833	0.004
IRCCS	-	LEVEL 1	0.864	0.914	24.0	0.945	0.781
	-	LEVEL 2	-1.320	1.001	24.0	-1.319	0.560
LEVEL 1	-	LEVEL 2	-2.184	0.637	24.0	-3.430	0.011

Note. Comparisons are based on estimated marginal means

Figure 55 ANOVA analysis of normalized regression predictions across different hospital facility levels.

4.3 INTEGRATION OF HOSPITAL CARE AND ENERGY SUSTAINABILITY: ADDRESSING RESEARCH QUESTIONS

In our extensive exploration of the energy and management efficiency of hospitals in Apulia, the two pivotal research questions have significantly shaped our analytical approach. Here's how our research addressed these questions:

A_5: Correlation between Hospital Organization, Patients' Propensity for Hospitalization, and Per Capita Energy Cost:

Our detailed analysis revealed a nuanced interplay between hospital organizational structures, the inclination of patients towards hospitalization, and the resultant energy costs. Specifically, we observed that efficient managerial practices, while enhancing the quality of care, could inadvertently lead to an increase in energy expenditure. This increase is especially pronounced in facilities with a higher number of patient admissions, underscoring the challenge of balancing patient care with energy conservation. Hospitals, therefore, need to adopt a multifaceted approach, optimizing managerial decisions without compromising the quality and extent of care, all while being mindful of energy costs.

A_6: Relationship between PTE, Medical Devices, and Per Capita Energy Cost:

The integration of modern medical technology in hospitals has undeniably improved patient care. However, our research indicated a direct correlation between the proliferation of medical devices in Apulian public hospitals and the rise in energy costs. Hospitals with a higher PTE score, indicating greater technical efficiency, also showed a pronounced increase in energy expenditure with the addition of more medical devices. This finding emphasizes the importance of a balanced approach where the adoption of new medical technologies, essential for improved patient outcomes, should be complemented with strategies for energy sustainability.

In conclusion, while Apulian hospitals are at the forefront of providing quality healthcare, there is an underlying challenge of managing energy costs. Addressing this requires a judicious blend of efficient managerial practices, sensible adoption of medical technologies, and a continual focus on energy conservation strategies. Decision-makers and hospital managers must be equipped with this knowledge to make informed choices, ensuring a sustainable, efficient, and patient-centric healthcare environment.

5. CONCLUSION OF III SESSION

The lack of homogeneity in energy management among public hospitals is evident from our studies. Better management of human resources could prove more effective in reducing energy costs than purchasing new equipment and expanding facilities.

Our in-depth analyses highlighted the importance of Pure Technical Efficiency (PTE) scores and the number of medical devices in determining the per capita energy cost. These variables, together with hospital organization and propensity for admission, play a crucial role in shaping energy costs.

The link between organizational efficiency and energy costs is particularly significant. For instance, an increase in organizational efficiency is associated with an increase in energy costs, while an increase in patient hospitalization rates is correlated with a rise in energy costs. These relationships underscore the need to carefully balance managerial decisions between resource optimization and energy cost management.

It's clear that there is no homogeneity in energy management practices among all hospitals. Healthcare company management should develop guidelines aimed at promoting a change in organizational culture, by creating an energy consumption management plan following the ISO 50001 guidelines and prioritizing the acquisition of alternative or renewable energy.
Policy decisions should aim to incentivize the public hospital network on the quality of services offered and not solely on economic productivity derived from DRGs, achieving a dual optimal allocation. An efficient allocation of economic resources that simultaneously promotes an efficient redistribution of regional admissions offers optimal outcomes in terms of perceived quality.

SESSION IV

ANALYSIS OF OPERATIONAL EFFICIENCY IN PUBLIC HOSPITALS: AN INNOVATIVE MACHINE LEARNING APPROACH

1. INTRODUCTION

In an era marked by unprecedented global health challenges, the efficiency of healthcare systems has become a focal point of scientific inquiry. The COVID-19 pandemic, which emerged in 2020, served as a severe test for healthcare infrastructures worldwide, exposing both their resilience and vulnerabilities. For instance, a comparative study on the efficiency of hospital bed management in four European countries before the pandemic outbreak highlighted how France and Germany were better prepared compared to Italy and Spain, underscoring the importance of a robust hospital structure (Pecoraro et al., 2020). This evolving landscape necessitates a rigorous examination and enhancement of hospital operational efficiency, not only as a response to the immediate crisis but as a fundamental element for the future trajectory of public health. In this context, transformative initiatives in hospital management emerge, aimed at improving the energy efficiency and environmental sustainability of healthcare facilities. A recent study highlighted the importance of integrating proposed guidelines with the adoption of ISO 50001 energy management systems to achieve the United Nations Sustainable Development Goal – SDG 7 clean and affordable energy" (Dion et al., 2023).
An analysis of the efficiency of emergency departments during the pandemic, using the Data Envelopment Analysis (DEA) model, revealed areas of potential improvement (Taghipour et al., 2023). Furthermore, the research developed three integrated conceptual strategic frameworks towards energy efficiency, green hospital initiatives, and corporate governance, providing recommendations for hospital managers and policymakers on how to effectively implement and manage energy efficiency initiatives in healthcare facilities (Dion & Evans, 2023). An additional study explored the impact of adopting sustainable management practices in a hospital setting, highlighting how the integration of energy efficiency strategies can not only reduce operational costs but also improve the quality of healthcare, offering a replicable model for other healthcare facilities (Dulce-Chamorro & Martinez-de-Pison, 2021).
The present study fits into this crucial context, using data related to 2020, a year that marked a significant turning point in the history of public health. Our analysis delves into the specific challenges faced by hospitals in the Apulia region, with the intent to extrapolate broader trends and dynamics of hospital efficiency within Italy and the wider European context. A recent investigation revealed how energy management in hospitals in Apulia is closely linked to healthcare performance, highlighting that an increase in organizational efficiency can lead to higher energy costs (Santamato et al., 2023). A study on the hospital on the island of Rhodes showed how adapting to organizational changes can increase efficiency and productivity in response to pandemic pressure (Androutsou et al., 2022).
The research is driven by a dual objective: to provide a comprehensive and data-based analysis of hospital efficiency during one of the most critical periods for the global healthcare system and to outline strategies for long-term improvement and optimization.

The pressing need for an efficient and responsive healthcare system was underscored by the rapid spread of the pandemic. A hospital's ability to effectively manage a sudden increase in care demands while maintaining high standards of quality emerged as a key indicator of preparedness and resilience. The Apulia region, characterized by unique demographic, economic, and health challenges, represents a particularly relevant case study in this context. A critical challenge identified in this scenario is the escalation of energy costs, intensified by global circumstances and a growing emphasis on sustainability. This aspect imposes the need for hospitals to devise new efficient resource management strategies. The necessity to balance energy consumption with high standards of healthcare has become a crucial factor in evaluating hospital efficiency. Our study explores how hospitals in the Apulia region are addressing this challenge, providing insights that could have broader applicability at an international level, as demonstrated by a study on hospital resource management against COVID-19 in Peru (Ninamango Origuela & Sovero Rivera, 2022).
We aim to achieve an optimal balance between operational costs, quality of care, and environmental sustainability. The analysis of 2020 data seeks to unveil how hospitals have adapted to these pressing issues, identifying areas of success and those requiring further improvements. This in-depth understanding is crucial for guiding future decisions, encompassing public health policy, energy management, and sustainability considerations. The paper is organized as follows: after introducing the problem and aims of the work, the next paragraph presents the methodological background and the case study to which it was applied. This is followed by a chapter with details of the study conducted, where most interesting results are presented and are then critically discussed in the following chapter along with an assessment of the proposed approach's innovativeness, potential, and limitations. The conclusions offer considerations of the methodology's implications for decision support at different scales and outlines possible follow-ups.

2. BACKGROUND

The context of our study is complex and multifaceted, characterized by a series of health, economic, and social challenges that have profoundly influenced the public health sector in Italy and around the world. The COVID-19 pandemic has underscored the importance of resilient health systems capable of rapidly adapting to crisis situations. In this context, the need for a more sustainable and circular approach to healthcare resource management clearly emerges. Within the framework of the National Recovery and Resilience Plan (NRRP) and national and international energy policies, the Apulia region faces the challenge of integrating principles of circular economy and green practices into hospital management. In response to these challenges, numerous countries, including Italy, have implemented reform and investment plans in the healthcare sector, such as the National Recovery and Resilience Plan (NRRP), aimed at strengthening the response capacity of the health system and promoting innovation.
The Apulia region, with its specific demographic configuration and peculiarities in the Italian healthcare landscape, offers a unique study context. The region faced specific challenges during the pandemic, including an immediate response in terms of hospital capacity, logistics, and resource management, highlighting the crucial importance of optimal operational efficiency.
Healthcare facilities, to effectively meet emerging needs and ensure long-term sustainability, must adopt strategies that go beyond the traditional linear model of

resource consumption. This includes the implementation of recycling practices, waste reduction, efficient energy use, and the adoption of renewable sources. Furthermore, the transition to green healthcare implies a change in procurement models, the use of technologies, and the management of hospital waste, laying the groundwork for a more contained environmental impact and greater social responsibility.

Recent research highlights the significant impact of integrating green practices and circular economy principles in enhancing sustainable performance in the service sector (Obeidat et al., 2023). This strategic intent towards green practices is crucial for healthcare systems, particularly in regions like Apulia, where such integration can lead to more efficient and sustainable operations.

The pressure on energy costs is a global phenomenon that has had significant repercussions for hospitals. The rise in energy prices and the need to adopt more sustainable practices have led to a reconsideration of how hospital resources are managed. In this scenario, the Apulia region presents itself as an exemplary case study, as it faces challenges common to many health systems but also local specificities that influence its response.

Examining how Apulian hospitals manage energy costs in the context of a global pandemic offers a unique learning opportunity. This study aims to analyze how resource optimization can be achieved without compromising the quality of care, exploring innovative and sustainable solutions. Moreover, the analysis aims to provide a broader perspective on the impact of resource management decisions at the regional and national level, fitting into a context of energy and environmental policies increasingly central in the public agenda.

Additionally, the pandemic has brought to light the critical issue of healthcare waste management. Insights from a study on the impact of COVID-19 on healthcare waste emphasize the importance of sustainable management within a life cycle and circular economy framework (Dihan et al., 2023). This perspective is crucial for the Apulia region as it navigates the complexities of healthcare waste management in a post-pandemic era, emphasizing the need for sustainable and circular approaches to ensure environmental and public health safety.

2.1 RELATED WORKS

In the field of research on healthcare system efficiency, the use of machine learning to improve resource management and energy efficiency in public hospitals emerges as a topic of significant scientific interest. Previous studies have highlighted the potential of these models in predicting and optimizing the use of hospital resources, emphasizing the importance of organizational efficiency and energy consumption management. In particular, research has shown that improving the organizational efficiency of hospitals can lead to increased energy costs, underscoring the need for a balanced approach that considers both operational efficiency and environmental sustainability (Santamato et al., 2023).

The use of machine learning systems in real-world contexts highlights ethical issues and the complexity of decision-making contexts (Kent & Ménager, 2023). Another significant research has underscored the critical importance of data in machine learning-based models, emphasizing the need for high-quality data to improve the effectiveness of healthcare services (Torra, 2023). Methodological approaches focused on the application of artificial intelligence can increase the efficiency of healthcare processes (Dubey et al., 2023).

In our study, we developed a neural network model that integrates concepts from both intelligent auditing and efficiency evaluation in the healthcare sector. Drawing inspiration from the use of an enhanced Self-Organizing Map (SOM) neural network for intelligent auditing of hospital financial vouchers (Wang, 2022), our model focuses on a categorical variable derived from cluster analysis. This approach is particularly effective in handling key features such as hospital energy costs and non-medical staff costs, crucial for predicting the scale efficiency category of a hospital.
Our methodology also aligns with the combined use of Data Envelopment Analysis (DEA) and Artificial Neural Networks (ANN) for evaluating hospital efficiency (Tosun, 2012). By categorizing hospitals based on efficiency through cluster analysis and then applying neural network processing, we offer an advanced method for managing complex variables and large datasets. This integration enhances predictive accuracy and operational efficiency analysis in hospital environments, providing a comprehensive tool for assessing and improving hospital efficiency, especially in the context of scale efficiency and resource management.
The use of machine learning in solving recurrence relations represents a promising area, combining the analysis of complex patterns with the processing of large amounts of data to optimize resources and improve decision-making processes in healthcare (Klemen et al., 2023).
Concurrently, the adoption of machine learning models in electronic structure calculations and multimodal learning is becoming a key factor in the analysis of complex healthcare data. These models offer the ability to manage and interpret large data sets, providing greater accuracy in identifying trends and patterns relevant to public health (Fiedler et al., 2023) (Xu et al., 2023).
Furthermore, the integration of machine learning with data assimilation techniques is transforming the way health information is processed and used, allowing for more effective and informed management (Cheng et al., 2023). Quantifying uncertainty in machine learning models, especially in biomedical applications, is crucial to ensure the safety and reliability of decisions based on these technologies (Nemani et al., 2023).
Finally, the application of machine learning to renewable energy systems offers new perspectives for energy management in hospitals, contributing to the reduction of environmental impact and improving the operational efficiency of healthcare facilities (Zaparoli Cunha et al., 2023).
These works collectively indicate the growing importance of machine learning in the healthcare sector, demonstrating its applicability in a variety of contexts, from resource management to operational efficiency, and highlighting its potential in addressing contemporary challenges in the field of public health.

2.2 MACHINE LEARNING ALGORITHMS APPLIED

In our study on hospital efficiency, we adopted a meticulous approach in applying machine learning algorithms to ensure precise and informative analyses.
STANDARDIZATION OF VARIABLES: The standardization algorithm for numerical health variables is a process that transforms numerical variables so that they have a mean of 0 and standard deviation of 1. This is important because when analyzing data from different sources, the units of measurement may be different, and therefore variables may have different scales. Standardization allows all variables to be put on the same scale, so that they can be compared fairly and accurately. Furthermore, standardization is often used as a first step before applying

multivariate analysis techniques such as PCA, in order to have a common starting point for all numerical variables. Standardizing data is an important step in data analysis, including the use of the data mining algorithm for PCA. Standardization is necessary to ensure that different variables within the dataset have the same scale, to avoid variables with higher values dominating those with lower values. This can affect PCA and produce inconsistent or misleading results. By standardizing the dataset, a more accurate analysis and better understanding of the data can be obtained (Mohammed et al., 2023).

PRINCIPAL COMPONENT ANALYSIS (PCA): The PCA (Principal Component Analysis) algorithm is a multivariate analysis technique used to reduce the dimensionality of a dataset by identifying linear combinations of input variables that capture most of the variance in the data. This process not only simplifies data understanding but also improves the visualization of relationships between variables. In the hospital context, applying PCA to input variables helped identify key factors influencing hospital performance. Similarly, applying PCA to output variables identified key factors affecting the quality of hospital care. Using PCA to reduce the number of original variables to two principal components simplifies the understanding of health data and provides useful information on the effectiveness of hospital organization and the quality of healthcare provided by the hospital. This approach highlights the importance of PCA in analyzing complex data, allowing for a deeper understanding and more accurate interpretation of underlying dynamics in various contexts (Jollife & Cadima, 2016), including healthcare.

DATA ENVELOPMENT ANALYSIS (DEA): Applying the DEA output-oriented algorithm to evaluate hospital efficiency (Ferreira et al., 2023) in producing Propensity for hospitalization, considering hospital organization as input and Propensity for hospitalization as output. We have focused our attention on Scale Efficiency (SE) to assess the operational efficiency of hospitals in relation to their size and resource management capacity. This approach is in line with the findings of several studies, conducted in different countries such as Tanzania, southeastern Nigeria, and Malaysia, which have emphasized the importance of assessing efficiency of scale in public hospitals. These studies highlight the need for resource reallocation, the use of health ratio indicators, and the implementation of technical and scale efficiency measures to optimize hospital performance and resource utilization (A Rahim et al., 2021; Aloh et al., 2020; Fumbwe et al., 2021).

CLUSTER ANALYSIS (LOUVAIN ALGORITHM): We employed the Louvain algorithm to categorize the numerical values (Xu et al., 2023) of hospitals based on their scale efficiency scores. This step helped us to meaningfully group hospitals based on their scale efficiency. Louvain’s algorithm, known for its efficiency and effectiveness in various contexts, has been shown to outperform other clustering methods in terms of accuracy and execution performance. it has demonstrated significant improvements in execution time and communication efficiency in distributed memory implementations (Ghosh et al., 2019). The robustness and versatility of Louvain's algorithm in clustering complex datasets make it an ideal choice for our analysis of hospital efficiency scores.

NEURAL NETWORK: In our study, we implemented a neural network model to analyze operational efficiency in hospitals, focusing on energy costs and non-medical staff costs. This approach draws inspiration from advanced methods of predicting energy consumption in hospital settings, as explored in recent research in the field of artificial intelligence applied to energy management (Panagiotou & Dounis, 2022). Integrating these methods into our model allowed for a more accurate assessment of the impact of energy costs on the scale efficiency of hospitals. Using a categorical variable derived from cluster analysis as the target, our model groups hospitals based on their operational efficiency. The use of artificial neural networks optimizes the prediction of scale efficiency, specifically considering energy consumption and non-medical staff costs. This method not only enhances the accuracy of the analysis but also provides strategic insights for more efficient and sustainable hospital management.

SHAP (SHAPLEY ADDITIVE EXPLANATIONS): A key element of our methodology was the use of the SHAP (SHapley Additive exPlanations) algorithm to interpret the results of the neural network. The effectiveness of SHAP in interpreting complex models has been demonstrated (Lundberg & Lee, 2017). This tool allowed us to gain a detailed understanding of the impact of each feature in the predictive model, particularly regarding hospital energy costs and non-medical staff costs. The application of SHAP enabled a more precise and in-depth evaluation of hospital efficiency.

ANOVA (ANALYSIS OF VARIANCE): We utilized ANOVA in hospital settings (Noudeh et al., 2022) to analyze the impact of cost features on scale efficiency scores and to examine significant differences among various hospital levels in terms of efficiency. This allowed us to assess how specific cost variables influence hospital efficiency. The methodological approach adopted in our study, which utilizes ANOVA to analyze the impact of cost features on hospital efficiency, finds a similarity in the use of ANOVA in the pedestrian safety study (Chang et al., 2022). In both contexts, ANOVA is used to examine the effect of multiple independent variables (cost factors in the hospital setting, environmental factors for pedestrian safety) on a specific dependent variable (hospital efficiency, pedestrian fatality incidents). This allows for a more detailed isolation and understanding of the impact of each variable, also analyzed through the SHAP approach, providing a deeper analysis of the complex dynamics in question.

This integrated approach provides us with a detailed framework of hospital efficiency, highlighting key areas for improvement and innovation in the healthcare sector, and offering valuable insights for resource management.

2.3 APPLICATION CONTEXT

The Apulia region faced specific challenges during the pandemic, including acute pressure on healthcare services due to a rapid increase in COVID-19 cases. This situation necessitated an immediate response in terms of hospital capacity, logistics, and resource management, highlighting the crucial importance of optimal operational efficiency. Additionally, the region had to balance the demands of the health emergency with the need to maintain high standards of care for all patients, not just those affected by COVID-19.

Another significant aspect was adapting to social distancing measures and restrictions implemented to contain the virus spread, requiring significant modifications in the organization of hospital spaces and services. These measures directly impacted the daily operations of hospitals, influencing staff management, internal logistics, and patient care and reception procedures.

Parallel to the immediate challenges posed by the pandemic, the Apulia region also had to address the rising energy costs, a concern made even more pressing by the need to operate in health emergency conditions. Efficient energy management became a critical factor, not only for reducing operational costs but also for contributing to environmental sustainability goals, in line with the directives of the

National Recovery and Resilience Plan (PNRR) and national and international energy policies.

The regional health system includes both public and private accredited facilities within the region, forming an organized complex known as the regional health industry (Falavigna & Ippoliti, 2013). The Regional Health Service in Apulia is represented by six Local Health Authorities (ASLs). ASLs are public legal entities with autonomy in organizational, managerial, technical, administrative, asset, and accounting matters, as well as entrepreneurial autonomy (according to Article 3 of Legislative Decree No. 502 of December 30, 1992). ASLs are part of the National Health Service.

This study focuses on the regional public hospital network in Apulia, specifically analyzing 28 facilities as indicated in the National Health Service Data Bank of the Ministry of Health. The public network consists of 24 ASL Direct Hospitals, one Hospital Authority integrated with the National Health System (NHS), one Hospital Authority integrated with the University, and one Public Institute for Hospitalization and Scientific Care. Some facilities, as specified in the single hospital reorganization document approved by the Apulian Regional Council on 03/07/2019, have connected plexuses and/or hospitals located in different places from the main structure. Thus, for greater accuracy in measuring the distance traveled by patients, all the physical plexuses indicated in the National Outcomes Plan have been considered. The total number of facilities considered for measuring the active mobility of Apulian patients is 37, including admissions made by ASL and by territorial ambit, as indicated in the National Outcomes Plan.

Regarding the classification of hospitals in Apulia, there are 5 second-level hospitals, 12 first-level hospitals, 2 Scientific Hospitalization and Care Institutes (IRCCS), and 9 basic hospitals. The differences between these types of hospitals are described in the Ministry of Health's regulations on the definition of qualitative, structural, technological, and quantitative standards for hospital care, which are implemented under Article 1, paragraph 169, of Law no. 311 of December 30, 2004, and Article 15, paragraph 13, letter c) of Decree-Law no. 95 of July 6, 2012, converted with amendments by Law no. 135 of August 7, 2012. The hospital levels differ primarily based on catchment area, number of wards, and complexity. According to the Ministry of Health:

Hospitals of First Level: They provide basic services such as emergency care, diagnostics, regular hospitalization, and outpatient services. They are usually present in various regions of Italy and provide primary level care to the local community.

Hospitals of Second Level: They are more specialized than first-level hospitals. They offer more complex services such as specialized surgery, intensive care, and hemodynamics services. They are present in numerous regions of Italy and serve as reference points for the provision of advanced care.

Basic Hospitals: They primarily perform primary care functions. They provide basic care, outpatient services, and primary level diagnostics. They are present in various regions of Italy and serve as a link between primary care and more specialized hospital facilities.

Institutes of Scientific Research and Care (IRCCS): They are specialized hospital facilities dedicated to scientific research and highly specialized healthcare. They are present in various regions of Italy and offer highly specialized care, playing an important role in medical research and the development of new therapies.

This context provided a unique basis for our study, allowing us to examine how hospital management and optimization strategies can be implemented in extreme crisis situations and regulatory change. Our analysis focuses on identifying effective and sustainable practices, which can serve as a model for other regions and health systems facing similar challenges. Ultimately, the study aims to provide valuable insights that can positively influence health policies and resource management at the regional.

2.4 STUDY OBJECTIVES AND RESEARCH QUESTION

The primary objective of this study is to utilize advanced machine learning models to analyze operational efficiency in public hospitals, with a specific focus on integrating considerations of environmental sustainability and effective resource management. The research question guiding our work is:
Q_7: "How can machine learning models contribute to the analysis and improvement of operational efficiency in public hospitals, considering the needs for sustainability and efficient resource management?".
By using the Apulia region as a sample area, we aim to create a detailed and applicable analytical framework that can provide relevant insights for decision-makers in the healthcare sector. The intention is to explore how machine learning practices can be transferred and used in different healthcare contexts, thereby offering guidance for optimizing hospital operations on a broader scale.
Specific objectives of the study include identifying the main factors that influence efficiency in public hospitals and assessing the integration of sustainability strategies to address current and future challenges. The purpose is to provide decision-makers in the healthcare sector with a data-based knowledge base and strategies that can be employed to guide informed decisions and improvements in healthcare services.
Therefore, the ambition of this study is to make a significant contribution to the scientific literature on machine learning applied to healthcare management and, at the same time, to offer a practical and replicable model that can guide decision-makers in transforming towards a more efficient, effective, and sustainable public hospital management.

3. METHODOLOGY

In the analysis of hospital efficiency, the use of advanced methodologies and machine learning tools is essential for an accurate and detailed evaluation. This study adopts an innovative approach, utilizing Orange software, a machine learning environment (Mirmozaffari et al., 2022), to explore and analyze the operational and allocative efficiency of hospitals in the Apulia region.
The methodology is illustrated in Figure 56, a mind map describing the methodological workflow. This conceptual map emphasizes the use of neural networks to analyze hospital cost variables and hospital allocative efficiency (SE) as the target variable.

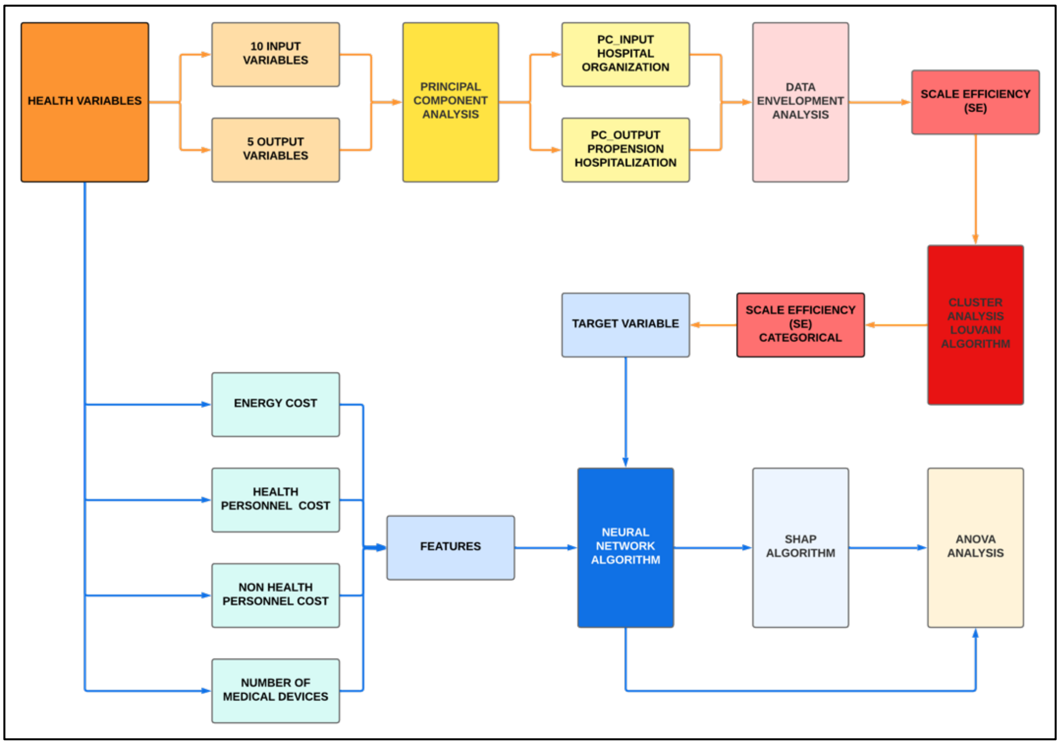


Figure 56 Methodological Workflow in Hospital Efficiency Analysis

At the core of our analysis is a neural network, designed to examine three types of hospital costs (energy cost, health and non-health staff cost) with the number of medical devices as features, and hospital allocative efficiency (SE) as the target variable. The process begins with the application of Principal Component Analysis (PCA) on the initial input and output variables, creating the "Hospital Organization" and "Hospitalization Propensity" variables, respectively. These variables are then employed in the Data Envelopment Analysis (DEA) output-oriented with variable returns to scale, applied to the 59 hospitals in the Apulia hospital network. In this context, the use of a three-staged DEA approach, as demonstrated in studies on public hospitals in emerging economies, can provide a more in-depth and nuanced assessment of operational efficiency (Hajiagha et al., 2023). This method allows for a comprehensive evaluation of performance metrics, which is crucial in settings with diverse challenges and resource constraints.
We decomposed Technical Efficiency (TE) into Pure Technical Efficiency (PTE) and Scale Efficiency (SE), obtaining SE values ranging from 0 to 1. Cluster analysis, conducted with the Louvain algorithm on SE, identified two distinct

clusters: high efficiency and low efficiency, transforming SE into a dichotomous target variable for hospital allocative efficiency.
In our methodological workflow, we integrated the use of the SHAP algorithm to analyze the influence and impact of the input features on the output and the model. This approach allowed us to quantify the importance of each feature in predicting hospital efficiency, providing a deeper understanding of how different variables influence the outcome.
The predictive results obtained from the neural network analysis, performed in the Orange machine learning environment, were further examined through ANOVA analysis and graphically represented. This methodological approach, supported by Orange software, provided a comprehensive and detailed view of hospital efficiency, integrating both operational and scale variables.
The analysis highlighted how designed features influence the allocative efficiency of hospitals, offering significant insights for management and strategic planning in the healthcare sector. Furthermore, the categorization of scale efficiency through cluster analysis allowed for a clearer understanding of efficiency dynamics within the hospital system.
The use of machine learning in the analysis of hospital efficiency has proven to be an effective approach, allowing for an in-depth and multidimensional analysis of hospital dynamics. This study not only contributes to the literature on hospital efficiency but also provides practical tools for decision-makers in the healthcare sector to optimize resources and improve the quality of services offered.

3.1 CRITERIA FOR VARIABLE SELECTION IN HOSPITAL ANALYSIS

In our study on hospital efficiency, we adopted the variable structure used in the study conducted by Santamato et al. (2023), introducing specific modifications to refine the analysis. Among these, we distinguished human components by gender, recognizing the importance of gender diversity in healthcare personnel, and included the catchment area for each hospital. These modifications were made to achieve a holistic and detailed view, which will be crucial in the subsequent step of studying hospital allocative efficiency. We focused on how allocative efficiency correlates with energy costs, non-medical staff costs, health personnel costs, and the number of hospital medical devices, critical factors in the management and sustainability of modern healthcare facilities. Figure 57 illustrates the variables selected for this study. The data were collected from various sources, including the National Health Service Database (NHS Database) of the Ministry of Health, the National Outcomes Program of the National Agency for Regional Health Services, and the new National Statistical Institute Database (ISTAT Database), for the year 2020. The data on medical device, analyzed as one of the features in the neural network of our study, come from the comprehensive list of large medical equipment cataloged in the national inventory by the Ministry of Health, as per the decree of April 22, 2014.

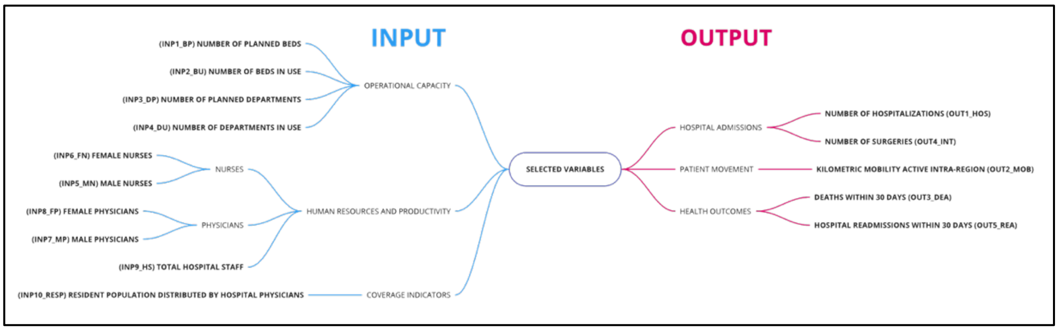


Figure 57 Overview of Selected Variables for Hospital Study.

The input section of the chart examines variables through a multidimensional perspective, assessing operational capacity in terms of available and utilized infrastructure. This segment outlines a quantitative analysis of the use of physical resources, highlighting the correspondence between the number of beds and departments planned and those in use, reflecting a hospital's ability to adapt to fluctuations in the demand for healthcare services.
The human component, crucial in the healthcare context, is analyzed through the distribution of staff, including doctors and nurses, with a gender distinction that illuminates workplace diversity and inclusion dynamics. The total hospital staff emerges as a key indicator of a hospital's ability to manage workload and sustain high standards of care.
Coverage indicators, such as the catchment area, are used to assess the reach and accessibility of the services offered.
The output section is assessed through variables that reflect the hospital's effectiveness. The number of hospitalizations and interventions serves as a direct measure of clinical activity. Intra-regional active kilometric mobility provides insight into logistics and accessibility to care for patients. Health outcomes, such as the number of deaths and hospital readmissions within 30 days, are considered essential parameters for evaluating the quality of care and the effectiveness of clinical interventions.

3.2 DERIVEDED VARIABLES

Apulian resident population distributed by hospital physicians:
Regarding the input variables, we used the resident population of Apulia as of December 31, 2020, distributed based on the number of hospital physicians, to estimate the size of the facility in terms of user and service basin.
To calculate the Apulian resident population distributed by hospital physicians, we first identified the number of residents in each municipality in Apulia for the year 2020. Next, we identified the municipalities that make up each ASL (Local Health Authority) and summed up the number of residents to obtain the total population for each ASL ($Tot_{res_{Asl}}$). Then, we determined the total number of physicians for each ASL by summing up the physicians working in hospitals within each ASL ($Tot\ Physicians_{ASL}$). Finally, we calculated the Apulian resident population distributed based on the number of hospital physicians for 2020 using the following formula:

$$Population_{res} = \frac{Tot_{res_{Asl}}}{Tot\ Physicians_{ASL}} x\ Physicians_{Hospitals} \tag{24}$$

Intra-regional active mobility of patients residing in Apulia:
In the present study, a patient's decision to seek treatment where they perceive better quality, subject to their economic availability and the medical offer proposed, is considered. 'Positive mobility' is defined as the flow of 'immigrants,' residents in the Apulia Region in 2020, who reach a hospital located in a different ASL from the one where the patient is a resident. Only intra-regional movements, i.e., within the region, made by patients' resident in the region, have been evaluated. Therefore, admissions of non-resident patients are not considered.
To calculate intra-regional active mobility in kilometers, we first calculated the interpolated distance between the patient's ASL of residence and the city where the hospital providing the service is located ($Dist\_km_{Hospital}$). Next, we summed the total number of active hospitalizations for each ASL (($Hospi_{ASL}$) and for each territorial area ($Hospi_{Area}$) within the region. Finally, we calculated intra-regional active mobility in kilometers using the following formula:

$$Active\ mobility_{infra-regional} = (Hospi_{ASL} + Hospi_{Area})x\ \ Dist_{km_{Hospital}} \tag{25}$$

This variable of intra-regional active mobility in kilometers represents the distance traveled by patients within the same region to access hospital services provided by different ASLs. It can be used to assess patient preference in choosing a hospital and may be correlated with the perceived quality of hospital services.

Energy Cost:
To calculate the cost of energy, we first calculated the ratio between the resident population in Apulia in 2020 ($Tot_{Res_{ASL}}$) and the total number of doctors in both public and private accredited hospitals, for each ASL ($Tot\ Physicians_{ASL}$).
We then multiplied this ratio by the number of doctors in each public hospital, weighted by ASL ($Physicians_{Hospitals}$) The resulting value for each facility, which is an expression of the population catchment area, was then multiplied by the per capita health energy cost of € 21.45 for the Apulian population in the year 2020, as indicated by the National Agency for Regional Health Services (AGE.NA.S).

$$Energy\ cost\ = \frac{Tot_{res_{Asl}}}{Tot\ Physicians_{ASL}} x\ \ Physicians_{Hospitals}\ x\ 21.45 \tag{26}$$

Cost of non-health personnel:
To calculate the cost of non-health staff, we initially identified the number of non-health personnel by subtracting the total number of doctors and nurses from the total hospital staff ($Tot\ non-health\ personnel_{ASL}$). Subsequently, we identified the cost items related to non-health staff for the year 2020, aggregated by ASL company ($Cost_{non-healt\ pers_{ASL}}$) (Data source: NSIS – SP, CE consolidated regional models. Extraction as of April 29, 2022. Economic data of the Regional Health Services. Economic-financial trend for the years 2019 - 2020, Apulia Region, AGE.NA.S.).
The cost of non-health hospital staff was calculated as the ratio of the aggregated cost per ASL to the number of non-health staff per ASL, then multiplied by the corresponding number of non-health hospital staff ($Non-health\ Personnel_{Hospitals}$), using the following formula:

$$Cost\ of\ non-health\ personnel = \frac{Cost_{non-healt\ pers_{ASL}}}{Tot\ Non-health\ personnel_{ASL}}\ x\ Non-health\ Personnel_{Hospitals} \quad (27)$$

Cost of health personnel:
To calculate the cost of health staff, we initially identified the number of health personnel by subtracting the total number of doctors and nurses from the total hospital staff ($Tot\ health\ personnel_{ASL}$). Subsequently, we identified the cost items related to health staff for the year 2020, aggregated by ASL company ($Cost_{healt\ pers_{ASL}}$) (Data source: NSIS – SP, CE consolidated regional models. Extraction as of April 29, 2022. Economic data of the Regional Health Services. Economic-financial trend for the years 2019 - 2020, Apulia Region, AGE.NA.S.).
The cost of health hospital staff was calculated as the ratio of the aggregated cost per ASL to the number of health staff per ASL, then multiplied by the corresponding number of health hospital staff ($Health\ Personnel_{Hospitals}$), using the following formula:

$$Cost\ of\ health\ personnel = \frac{Cost_{healt\ pers_{ASL}}}{Tot\ Health\ personnel_{ASL}}\ x\ Health\ Personnel_{Hospitals} \quad (28)$$

Hospital Scale Efficiency (SE):
Hospital Scale Efficiency is derived from decomposing the overall hospital efficiency, which is obtained through the application of Data Envelopment Analysis (DEA) with an output-oriented approach and variable returns to scale. This analysis considers "Hospital Organization" as the input and "Propension Hospitalization" as the output. The formula for calculating Technical Efficiency (TE) is given by:

$$TE\ (Technical\ Efficiency) = PTE\ (Pure\ Technical\ Efficiency)\ x\ SE\ (Scale\ Efficiency) \quad (29)$$

In this context, Pure Technical Efficiency (PTE) reflects the ability of a hospital to maximize outputs with given inputs, independent of its size and scale of operations. Scale Efficiency (SE), on the other hand, measures the efficiency of the hospital's size or scale of operations. By multiplying PTE with SE, we obtain the overall Technical Efficiency (TE), which provides a comprehensive measure of a hospital's operational effectiveness, considering both its operational processes and scale of operations. Subsequently, cluster analysis is applied to categorize the Scale Efficiency into two groups: high efficiency and low efficiency. This categorization process helps to simplify the analysis and understanding of the hospital's performance in terms of scale efficiency.

3.3 SELECTION OF VARIABLES AND ANALYSIS OF IMPACT, ADVANTAGES, AND LIMITATIONS IN THE INTERPRETATION OF HOSPITAL EFFICIENCY THROUGH NEURAL NETWORKS: A FOCUS ON COSTS AND SUSTAINABILITY

In the context of neural network analysis, we have chosen to use energy costs, non-medical staff costs, healthcare personnel costs, and the number of hospital medical devices as the main features. These variables are essential not only for understanding the dynamics of expenditure and resource management within hospitals but also crucial for assessing operational efficiency from a sustainability perspective.

Scale Efficiency (SE) has been selected as the target variable for analysis. SE represents a key indicator of a hospital's operational efficiency, reflecting an institution's ability to maximize outputs in relation to its operational scale. The goal of the neural network analysis is to predict hospital operational efficiency based on various factors, including management costs and the environmental and social impact of hospital operations.

Including healthcare personnel costs and the number of medical devices allows us to examine how investment in qualified human resources and advanced technologies influences efficiency and sustainability. These variables can significantly impact not only operational efficiency but also the quality of care, reduction of environmental impact, and promotion of sustainable healthcare practices.

However, the analysis must consider limitations related to data quality and the risk of model overfitting. Moreover, the "black box" nature of neural networks can make interpreting results difficult, especially in terms of impacts on sustainability.

Despite these challenges, integrating these variables into the neural network model offers the opportunity to explore more deeply the interaction between operational efficiency and sustainability. The analysis can reveal how resource optimization and the adoption of advanced technologies can contribute to more sustainable hospital operations, improving care quality and reducing environmental impact.

In conclusion, integrating these specific variables into a neural network model, with a focus on sustainability, provides a powerful tool for analyzing and improving hospital operational efficiency. This approach not only aims to optimize resources and improve service quality but also promotes sustainable healthcare practices, essential for the future of the healthcare sector.

3.4 IDENTIFICATION OF HOSPITAL ALLOCATIVE EFFICIENCY AS A TARGET VARIABLE IN NEURAL NETWORK ANALYSIS

In the realm of healthcare efficiency analysis, the identification of key performance indicators is crucial for enhancing operational effectiveness and resource allocation. This study introduces an innovative methodological approach in the utilization of neural network analysis, focusing on the identification of Hospital Allocative Efficiency (SE) as the target variable. This approach represents a significant advancement in the field, offering a more nuanced and comprehensive understanding of hospital efficiency dynamics.

Standardization and Principal Component Analysis:

Initially, we standardized the groups of 10 input variables and 5 output variables (Figure 2) for all 59 hospitals in the Apulia hospital network, normalizing them with a mean of 0 and a variance of 1.

The input variables, denoted as $X_1, X_2, \dots, X_{10}$ and the output variables, denoted $Y_1, Y_2, \dots, Y_5$, were standardized to ensure uniformity and comparability. This standardization process was achieved by subtracting the mean (μ) and dividing by the standard deviation (σ) for each variable, as indicated in the formulas $Z_{X_i} = \frac{(X_i - \mu_{X_i})}{\sigma_{X_i}}$ for input variables and $Z_{Y_j} = \frac{(Y_j - \mu_{Y_j})}{\sigma_{Y_j}}$ for output variables.

Subsequently, we formed the matrices of standardized variables, M_{input} and M_{output}, representing the sets of standardized input and output variables, respectively. These matrices were used to calculate the covariance matrices, C_{input} and C_{output},
using the formula $C_{input} = \frac{1}{n-1}\, x\, (M_{input})^T\, x\, M_{input}$ and $C_{output} = \frac{1}{n-1}\, x\, (M_{output})^T\, x\, M_{output}$. This step is crucial for understanding the internal relationships among the variables and for setting the stage for Principal Component Analysis (PCA).
PCA was then applied to extract the principal components from the covariance matrices. We calculated the eigenvectors and eigenvalues for each covariance matrix and selected the eigenvectors corresponding to the largest eigenvalues. This led to the identification of the principal components for the input and output groups, named “Hospital Organization” and “Propension Hospitalization”, respectively, calculated as $PC_{input} = M_{input}\, x\, Eigenvector_{\max\_input}$ and $PC_{output} = M_{output}\, x\, Eigenvector_{\max\_output}$.
This process confirms and aligns with the techniques employed in the study by Santamato et al., 2023, demonstrating the consistency and effectiveness of such methodologies in the field of hospital efficiency.

Positive Shift e Data Envelopment Analysis:
We implemented a crucial step known as "Positive Shift" on the principal components, PC_{input} and PC_{output}. This step is essential to ensure that all values are positive, a necessary condition for applying Data Envelopment Analysis (DEA). The Positive Shift was realized using the following formulas:

For PC_{input}: $PC_{input}^{shifted} = PC_{input} + \min\left(min\left(PC_{input}\right), min\left(PC_{output}\right)\right) + 1$ (30)
For PC_{output}: $PC_{output}^{shifted} = PC_{output} + \min\left(min\left(PC_{input}\right), min\left(PC_{output}\right)\right) + 1$ (31)

This method ensures that both principal components are transformed in such a way that all values are positive, making them suitable for DEA analysis. By using the minimum value between PC_{input} and PC_{output} for the shift, we ensure that the analysis is based on consistent data, correctly prepared for efficiency analysis.
we proceeded with the Output-Oriented DEA with variable returns to scale (VRS). This model was chosen for its ability to maximize outputs considering the operational scale of hospitals. The key formulas for the VRS DEA include:

1.Output – Oriented VRS DEA Model:
$max\ \theta$
Subject to:

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{i0}, i = 1,2,\dots,m. \tag{32}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq \theta y_{r0}, r = 1,2,\dots,s. \tag{33}$$

$$\sum_{j=1}^{n} \lambda_j = 1, \lambda \geq 0 \tag{34}$$

These formulas represent the core of the DEA analysis, where x_{ij} and y_{rj} are the inputs and outputs of hospitals ($PC_{input}^{shifted}$ and $PC_{output}^{shifted}$), λ_j are the weighting variables, θ is the efficiency parameter to be maximized and n (=59) is the total number of hospitals in the network..

2.Decomposition of Total Technical Efficiency (TE):

$$TE = PTE\ x\ SE \quad (35)$$

Here, TE represents the total technical efficiency, PTE is the pure technical efficiency, and SE is the scale efficiency. This decomposition allows for a more detailed analysis of different aspects of hospital efficiency.

Cluster Analysis:

The use of cluster analysis to categorize hospitals based on their scale efficiency (SE) was a critical step in transforming SE from a continuous variable into a categorical variable, suitable for use as a target variable in predictive analysis using neural networks. This transformation enabled the identification of significant patterns and trends in hospital operational efficiency, facilitating the interpretation and practical application of the findings.

The Louvain algorithm was chosen for its effectiveness in identifying communities or clusters within large networks. This algorithm optimizes modularity Q, defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (36)$$

where A_{ij} represents the weight of the edge between nodes i and j, k_i and k_j are the degrees of nodes i and j, m is the sum of the weights of all edges in the network, and $\delta(c_i, c_j)$ is 1 if i and j are in the same cluster and 0 otherwise.

Applying the Louvain algorithm to the SE values for the 59 hospitals resulted in the identification of two distinct clusters: one with 14 hospitals (C1) and one with 45 hospitals (C2) (Fig. 58).

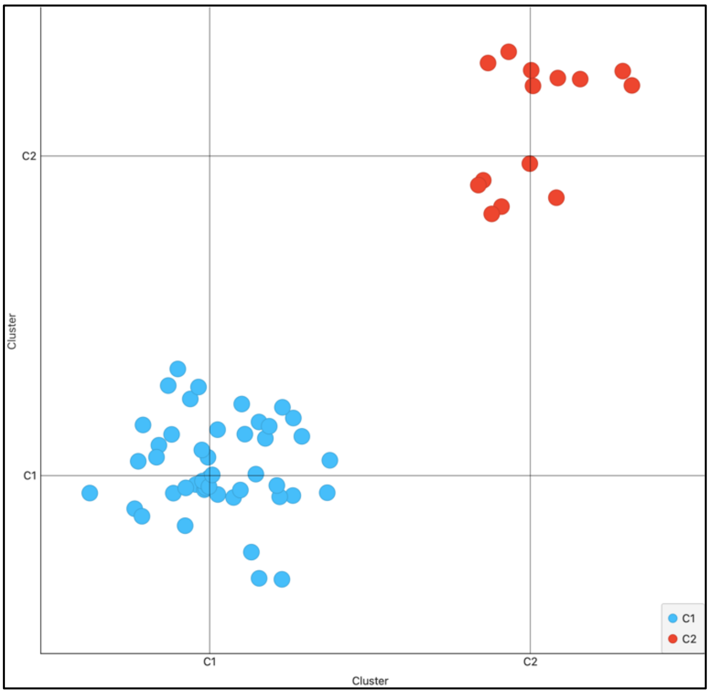


Figure 58 Scatter plot of 2 clusters identified by Louvain algorithm.

The average scale efficiency for C1 was 0.951996, while for C2 it was 0.637046. We labeled C1 as the “high efficiency” cluster and C2 as the “low efficiency” cluster. The average silhouette index, an indicator of the internal consistency of the clusters, was 0.714279, suggesting a good separation between the two groups. This categorization allowed for the transformation of the numerical SE variable into a categorical variable, which was then used as the target variable for subsequent analysis with neural networks.

3.5 PARAMETERS USED FOR DEFINING THE TARGET VARIABLE: HOSPITAL ALLOCATIVE EFFICIENCY (SE)

Standardizzazione delle Variabili di Input e Output:
Objective: Normalize the data for comparability and reduce the influence of outliers.
Method: Z-Score Standardization.
Application: Applied to the 10 input variables and 5 output variables for the 59 hospitals in the Apulian Region.
Parameters: $\mu = 0, \sigma = 1$.

Principal Component Analysis (PCA):
Objective: To reduce dimensionality and identify principal components.
Application: Applied separately to the standardized input and output variable groups.
Parameters:
For 10 Input variables:

- 1 Principal Component
- Explained variance 95%
- The variance of the components is homogeneous (Fig. 59).
- Component Loading: varimax rotation was used.
- Bartlett’s Test of Sphericity: $p<0.001$
- KMO Measure of Sampling Adequacy: MSA 0.887

For 5 Output variables:

- 1 Principal Component
- Explained variance 88%
- The variance of the components is homogeneous (Fig. 59).
- Component Loading: varimax rotation was used.
- Bartlett’s Test of Sphericity: $p<0.001$
- KMO Measure of Sampling Adequacy: MSA 0.917

Data Table

Data instances: 1
Features: 10
Meta attributes: 2

	components	variance	INP7_MP	INP10_RESP	INP8_FP	INP5_MN	INP6_FN	INP9_HS	INP2_BU	INP3_DP	INP4_DU	INP1_BP
1	PC1	0.954227	0.318983	0.314576	0.313861	0.316871	0.315813	0.319617	0.308111	0.319368	0.317284	0.317624

Data Table

Data instances: 1
Features: 5
Meta attributes: 2

	components	variance	OUT4_INT	OUT5_REA	OUT3_DEA	OUT1_HOS	OUT2_MOB
1	PC1	0.88791	0.447566	0.453067	0.44015	0.454725	0.44035

Figure 59 The impact of each individual original variable on the main components.

Data Envelopment Analysis:
Objective: Evaluate the efficiency of hospitals in transforming inputs into outputs.
Application: Applied to the positively shifted principal components derived from PCA.
Parameters:

- Variable Scale Return (VRS)
- Output Oriented
- No weight for variables
- 1PCA_Input and 1PCA_Output
- Scale Efficiency (SE) = $\frac{\text{Output−Oriented CRS (Constant Returns to Scale) Efficiency}}{\text{Output−Oriented VRS (Variable Returns to Scale) Efficiency}}$
- Output: A score of 1 indicates efficiency, while scores below 1 indicate relative inefficiency.

Cluster Analysis:
Objective: Categorize hospitals into distinct groups based on their allocative efficiency.
Method: Louvain Clustering.
Application: Applied to the SE scores derived from DEA.
Parameters:

- Two clusters identified: High Efficiency (C1) and Low Efficiency (C2).
- Average Scale Efficiency: C1=0.951996, C2=0.637046
- Silhouette Index: 0.714279
- Normalize data: yes.
- PCA preprocessing: yes, 1 component.
- Metric: Cosine
- K neighbors: 9
- Resolution: 1.0
- Output: Transformation of SE into a categorical variable for neural network analysis.

Although the analysis was conducted on the entire network of 59 hospitals in the Apulia region, our focus is on the 28 public hospitals. This selection was driven by the importance of providing a more detailed and specific analysis, one that considers the unique peculiarities and challenges of the public hospital sector. The resulting dataset (Fig. 60), which includes the target variable for neural network analysis, is thus an accurate and targeted reflection of the efficiency dynamics specific to the public hospitals in Apulia.

Data Table

Data instances: 28
Features: 4
Meta attributes: 4
Target: Class 'SCALE EFFICIENCY (Categorical)'

	SCALE EFFICIENCY (Categorical)	HOSPITAL	NETWORK	ASL	HOSPITAL LEVEL	ENERGY COST	NON-HEALTH PERSONNEL COST	HEALTH PERSONNEL COST	MEDICAL DEVICES
1	LOW EFFICIENCY	Ospedale San Paolo Bari (BA)	PUBLIC	ASL BA	FIRST LEVEL	0.681493	2.0976	1.77669	1.2874
2	HIGH EFFICIENCY	Ospedale Monopoli Monopoli (BA)	PUBLIC	ASL BA	BASE LEVEL	-0.699023	-0.907956	-0.405389	-0.655263
3	HIGH EFFICIENCY	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	PUBLIC	ASL FG	FIRST LEVEL	-0.717041	-0.47306	-0.156241	-0.439412
4	LOW EFFICIENCY	Consorziale Policlinico Bari Bari (BA)	PUBLIC	ASL BA	SECOND LEVEL	2.9271	2.56925	2.5865	2.58251
5	HIGH EFFICIENCY	Ospedale Casarano Casarano (LE)	PUBLIC	ASL LE	BASE LEVEL	-0.5004	-0.433544	-0.553225	-0.439412
6	HIGH EFFICIENCY	Ospedale Scorrano Scorrano (LE)	PUBLIC	ASL LE	FIRST LEVEL	-0.463917	-0.356179	-0.44288	-0.439412
7	LOW EFFICIENCY	Ospedale Andria Andria (BT)	PUBLIC	ASL BT	FIRST LEVEL	0.105352	0.264786	0.0731797	0.208142
8	HIGH EFFICIENCY	Ospedale San Severo - Teresa Masselli San Severo (FG)	PUBLIC	ASL FG	FIRST LEVEL	-0.560857	0.108296	-0.062447	-0.871114
9	HIGH EFFICIENCY	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	PUBLIC	ASL LE	BASE LEVEL	-0.342284	-0.439991	-0.587555	-0.871114
10	HIGH EFFICIENCY	Ospedale Putignano Putignano (BA)	PUBLIC	ASL BA	BASE LEVEL	-0.454877	-0.971291	-0.6886	-0.871114
11	LOW EFFICIENCY	Ospedale Lecce 'V. Fazzi' Lecce (LE)	PUBLIC	ASL LE	SECOND LEVEL	1.87134	1.88741	1.85966	2.1508
12	HIGH EFFICIENCY	Ospedale Castellaneta Castellaneta (TA)	PUBLIC	ASL TA	FIRST LEVEL	-0.684583	-0.56071	-0.646426	-0.439412
13	HIGH EFFICIENCY	Istituto Tumori Giovanni Paolo II Bari (BA)	PUBLIC	ASL BA	IRCCS	-0.801511	-0.946635	-0.72402	0.639845
14	HIGH EFFICIENCY	Ospedale Ostuni Ostuni (BR)	PUBLIC	ASL BR	BASE LEVEL	-0.49739	-0.38766	-1.2457	-0.871114
15	LOW EFFICIENCY	Ospedali Riuniti Di Foggia Foggia (FG)	PUBLIC	ASL FG	SECOND LEVEL	1.2736	1.59692	1.3619	0.208142
16	LOW EFFICIENCY	Ospedale Di Venere Bari (BA)	PUBLIC	ASL BA	FIRST LEVEL	-0.277034	0.974835	0.792191	0.855696
17	HIGH EFFICIENCY	Ospedale Francavilla Fontana Francavilla Fontana (BR)	PUBLIC	ASL BR	FIRST LEVEL	-0.654094	-0.798362	-0.6455	-0.439412
18	HIGH EFFICIENCY	Ospedale Civile Martina Franca (TA)	PUBLIC	ASL TA	FIRST LEVEL	-0.649141	-0.78657	-0.630893	-0.655263
19	HIGH EFFICIENCY	Ospedale Bisceglie Bisceglie (BT)	PUBLIC	ASL BT	BASE LEVEL	0.132361	-0.656388	-0.599182	-0.871114
20	HIGH EFFICIENCY	Ospedale Copertino Copertino (LE)	PUBLIC	ASL LE	BASE LEVEL	-0.597699	-0.607616	-0.727325	-0.655263
21	LOW EFFICIENCY	Presidio Ospedaliero centrale Taranto (TA)	PUBLIC	ASL TA	SECOND LEVEL	1.64301	1.19085	1.46959	1.93495
22	LOW EFFICIENCY	Ospedale Perrino Brindisi (BR)	PUBLIC	ASL BR	SECOND LEVEL	2.0387	0.492415	1.41753	1.50325
23	HIGH EFFICIENCY	Ospedale Manfredonia Manfredonia (FG)	PUBLIC	ASL FG	BASE LEVEL	-0.807913	-0.398314	-0.675712	-0.655263
24	HIGH EFFICIENCY	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	PUBLIC	ASL LE	FIRST LEVEL	-0.419311	-0.278814	-0.460045	-0.439412
25	LOW EFFICIENCY	Ospedale Della Murgia - Perinei Altamura (BA)	PUBLIC	ASL BA	FIRST LEVEL	-0.439806	-0.182477	-0.205792	-0.00770898
26	HIGH EFFICIENCY	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	PUBLIC	ASL BA	IRCCS	-0.858785	-1.085	-1.02226	-0.439412
27	LOW EFFICIENCY	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	PUBLIC	ASL BT	FIRST LEVEL	0.334897	-0.000777403	-0.109791	-0.871114
28	HIGH EFFICIENCY	Ospedale Manduria 'Giannuzzi' Manduria (TA)	PUBLIC	ASL TA	BASE LEVEL	-0.582182	-0.911023	-0.748257	-0.439412

Figure 60 Dataset Report for Neural Network Analysis on the 28 Public Hospitals in the Apulia Region.

Figure 5 illustrates the dataset report that will be processed for neural network analysis. The dataset includes 28 data instances, corresponding to the public hospitals in the Apulia Region. The four main features are standardized variables, with a mean of zero and a variance of one, related to hospital energy costs, number of medical devices, health and non-health staff costs. Additionally, the dataset includes four meta-attributes: the name of the hospital, its affiliation with the public hospital network, the local health authority (ASL) to which the hospital belongs, and the hospital level. The selected target variable for the analysis is Scale Efficiency, categorized to reflect the allocative efficiency of the hospitals.

3.6 APPLICATION OF NEURAL NETWORKS AND MACHINE LEARNING IN HOSPITAL EFFICIENCY ANALYSIS

In our study on hospital efficiency in the Apulia region, we employed advanced machine learning techniques, with a particular focus on the use of a neural network. The key variables analyzed include number of medical devices, energy costs, health and non-health staff costs, treated as the main inputs in the neural network model, aiming to classify hospital allocative efficiency into two categories: high efficiency and low efficiency.

The neural network model was configured with the following parameters: 100 hidden layers, ReLU activation function, Adam solver, alpha 0.0001, and a maximum of 200 iterations. We opted for an integrated approach in training and validating our neural network model, using the entire dataset without dividing it into separate training and validation sets. This decision was made considering the limited size of the dataset, where division could potentially compromise representativeness and reliability of the results. To mitigate the risk of overfitting and ensure the generalizability of the model, we implemented cross-validation techniques, thus allowing for a complete and iterative use of the dataset. Moreover, we placed particular emphasis on data quality, ensuring accurate preparation and

cleaning to provide precise and informative inputs to the model. This approach allowed us to maximize the use of available information, while ensuring the robustness and validity of the model in a context with a limited number of observations. The training was replicable, with a stratified 10-fold cross-validation sampling type. The results obtained showed high model accuracy, with an AUC of 0.994, an accuracy (CA) of 0.929, an F1 score of 0.926, precision (PREC) of 0.936, recall of 0.929, and an MCC of 0.849.
For a more detailed evaluation of the model's performance, we included the analysis of the confusion matrix and ROC curves. The confusion matrix revealed a high match between the model's predictions and the actual classifications: 90.0% of high-efficiency cases were correctly identified (High-High), while 100% of low-efficiency cases were accurately classified (Low-Low). Only 10.0% of high-efficiency cases were mistakenly classified as low efficiency (Low-High).
The ROC curves for both classes (high and low efficiency) were included to provide a visual representation of the model's ability to distinguish between these two categories. These curves demonstrate excellent class separation, with high AUC values indicating a strong capability of the model to correctly classify hospitals based on their allocative efficiency (Fig. 61). The inclusion of these visual analyses not only enriches our understanding of the model's performance but also provides a solid foundation for further investigations and practical applications in the healthcare sector.

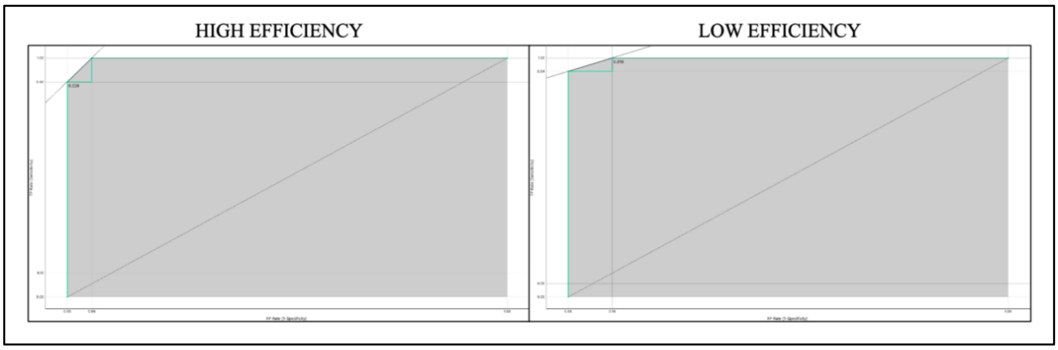


Figure 61 ROC Curves for High and Low Efficiency Classes in Hospital Allocative Efficiency Analysis.

Figure 62 provides a clear and detailed graphical representation of the analysis results, effectively illustrating the dynamics and correlations emerged from the study. This figure synthesizes the 28 instances of the hospitals studied, highlighting the four main variables (energy costs, number of medical devices, medical and non-medical staff costs) and the four meta-attributes (hospital name, hospital level, ASL, neural network analysis) along with the target variable. This layout offers a comprehensive overview of the adopted approach and the importance of each attribute in the analysis of hospital efficiency.

Data Table

Data instances: 28
Features: 4
Meta attributes: 4
Target: Class 'SCALE EFFICIENCY (Categorical)'

	SCALE EFFICIENCY (Categorical)	ASL	HOSPITAL LEVEL	HOSPITAL	Neural Network	ENERGY COST	NON-HEALTH PERSONNEL COST	HEALTH PERSONNEL COST	MEDICAL DEVICES
1	LOW EFFICIENCY	ASL BA	FIRST LEVEL	Ospedale San Paolo Bari (BA)	LOW EFFICIENCY	0.681493	2.0976	1.77669	1.2874
2	HIGH EFFICIENCY	ASL BA	BASE LEVEL	Ospedale Monopoli Monopoli (BA)	HIGH EFFICIENCY	-0.699023	-0.907956	-0.405389	-0.655263
3	HIGH EFFICIENCY	ASL FG	FIRST LEVEL	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	HIGH EFFICIENCY	-0.717041	-0.47306	-0.156241	-0.439412
4	LOW EFFICIENCY	ASL BA	SECOND LEVEL	Consorziale Policlinico Bari Bari (BA)	LOW EFFICIENCY	2.9271	2.56925	2.5865	2.58251
5	HIGH EFFICIENCY	ASL LE	BASE LEVEL	Ospedale Casarano Casarano (LE)	HIGH EFFICIENCY	-0.5004	-0.433544	-0.553225	-0.439412
6	HIGH EFFICIENCY	ASL LE	FIRST LEVEL	Ospedale Scorrano Scorrano (LE)	HIGH EFFICIENCY	-0.463917	-0.356179	-0.44288	-0.439412
7	LOW EFFICIENCY	ASL BT	FIRST LEVEL	Ospedale Andria Andria (BT)	LOW EFFICIENCY	0.105352	0.264786	0.0731797	0.208142
8	HIGH EFFICIENCY	ASL FG	FIRST LEVEL	Ospedale San Severo - Teresa Masselli San Severo (FG)	HIGH EFFICIENCY	-0.560857	0.108296	-0.062447	-0.871114
9	HIGH EFFICIENCY	ASL LE	BASE LEVEL	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	HIGH EFFICIENCY	-0.342284	-0.439991	-0.587555	-0.871114
10	HIGH EFFICIENCY	ASL BA	BASE LEVEL	Ospedale Putignano Putignano (BA)	HIGH EFFICIENCY	-0.454877	-0.971291	-0.6886	-0.871114
11	LOW EFFICIENCY	ASL LE	SECOND LEVEL	Ospedale Lecce 'V. Fazzi' Lecce (LE)	LOW EFFICIENCY	1.87134	1.88741	1.85966	2.1508
12	HIGH EFFICIENCY	ASL TA	FIRST LEVEL	Ospedale Castellaneta Castellaneta (TA)	HIGH EFFICIENCY	-0.684583	-0.56071	-0.646426	-0.439412
13	HIGH EFFICIENCY	ASL BA	IRCCS	Istituto Tumori Giovanni Paolo II Bari (BA)	HIGH EFFICIENCY	-0.801511	-0.946635	-0.72402	0.639845
14	HIGH EFFICIENCY	ASL BR	BASE LEVEL	Ospedale Ostuni Ostuni (BR)	HIGH EFFICIENCY	-0.49739	-0.38766	-1.2457	-0.871114
15	LOW EFFICIENCY	ASL FG	SECOND LEVEL	Ospedali Riuniti Di Foggia Foggia (FG)	LOW EFFICIENCY	1.2736	1.59692	1.3619	0.208142
16	LOW EFFICIENCY	ASL BA	FIRST LEVEL	Ospedale Di Venere Bari (BA)	LOW EFFICIENCY	-0.277034	0.974835	0.792191	0.855696
17	HIGH EFFICIENCY	ASL BR	FIRST LEVEL	Ospedale Francavilla Fontana Francavilla Fontana (BR)	HIGH EFFICIENCY	-0.654094	-0.798362	-0.6455	-0.439412
18	HIGH EFFICIENCY	ASL TA	FIRST LEVEL	Ospedale Civile Martina Franca (TA)	HIGH EFFICIENCY	-0.649141	-0.78657	-0.630893	-0.655263
19	HIGH EFFICIENCY	ASL BT	BASE LEVEL	Ospedale Bisceglie Bisceglie (BT)	HIGH EFFICIENCY	0.132361	-0.656388	-0.599182	-0.871114
20	HIGH EFFICIENCY	ASL LE	BASE LEVEL	Ospedale Copertino Copertino (LE)	HIGH EFFICIENCY	-0.597699	-0.607616	-0.727325	-0.655263
21	LOW EFFICIENCY	ASL TA	SECOND LEVEL	Presidio Ospedaliero centrale Taranto (TA)	LOW EFFICIENCY	1.64301	1.19085	1.46959	1.93495
22	LOW EFFICIENCY	ASL BR	SECOND LEVEL	Ospedale Perrino Brindisi (BR)	LOW EFFICIENCY	2.0387	0.492415	1.41753	1.50325
23	HIGH EFFICIENCY	ASL FG	BASE LEVEL	Ospedale Manfredonia Manfredonia (FG)	HIGH EFFICIENCY	-0.807913	-0.398314	-0.675712	-0.655263
24	HIGH EFFICIENCY	ASL LE	FIRST LEVEL	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	HIGH EFFICIENCY	-0.419311	-0.278814	-0.460045	-0.439412
25	LOW EFFICIENCY	ASL BA	FIRST LEVEL	Ospedale Della Murgia - Perinei Altamura (BA)	HIGH EFFICIENCY	-0.439806	-0.182477	-0.205792	-0.00770898
26	HIGH EFFICIENCY	ASL BA	IRCCS	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	HIGH EFFICIENCY	-0.858785	-1.085	-1.02226	-0.439412
27	LOW EFFICIENCY	ASL BT	FIRST LEVEL	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	HIGH EFFICIENCY	0.334897	-0.000777403	-0.109791	-0.871114
28	HIGH EFFICIENCY	ASL TA	BASE LEVEL	Ospedale Manduria 'Giannuzzi' Manduria (TA)	HIGH EFFICIENCY	-0.582182	-0.911023	-0.748257	-0.439412

Figure 62 Detailed Analysis of the 28 Hospital Instances with Neural Network.

The details of this analysis and the discussion of the results will be further explored in the Results and Discussion section, where the implications of the collected data and their relevance in the context of management and strategic planning in the healthcare sector will be examined.

3.7 SELECTION OF KEY FEATURES FOR NEURAL NETWORK IN HOSPITAL OPERATIONS

In the realm of hospital operations, the integration of a neural network necessitates the careful selection of features that significantly impact operational efficiency. Among these, four critical features stand out: energy cost, healthcare staff cost (both medical and non-medical), and the number of medical devices (such as CT scanners and MRI machines).

Energy Cost: The cost of energy in hospitals is a substantial component of operational expenses. Efficient energy management not only reduces costs but also aligns with sustainable healthcare practices. A study on hospital management practices revealed that better-managed cardiac units paid lower prices for cardiac devices, indicating a correlation between efficient management and cost savings, which can extend to energy usage as well (Grennan et al., 2022).

Healthcare Staff Cost: Staffing costs, encompassing both medical and non-medical personnel, represent a significant portion of hospital budgets. Efficient management of these costs is crucial for financial sustainability. Research indicates that hospitals can achieve cost reductions through improved management practices, which include efficient staffing strategies (Abe et al., 2016).

Number of Medical Devices: The number and type of medical devices, such as CT scanners and MRI machines, are indicative of a hospital's capacity to provide advanced medical care. However, these devices also contribute to operational costs. Studies have shown that hospitals can control costs by managing their medical device inventory effectively, balancing the need for advanced technology with financial constraints (Robinson & Brown, 2014).

Medical Devices: The cost associated with acquiring and maintaining medical devices is a critical factor. Hospitals need to navigate the balance between having state-of-the-art medical equipment and the associated costs. Research has

highlighted the importance of strategic purchasing and maintenance of medical devices to minimize costs (Çakmak & Yol, 2019).
Incorporating these features into a neural network model allows for a comprehensive analysis of hospital operations, enabling the identification of areas for cost optimization and efficiency improvements. This approach not only aids in financial management but also ensures the delivery of quality healthcare services.

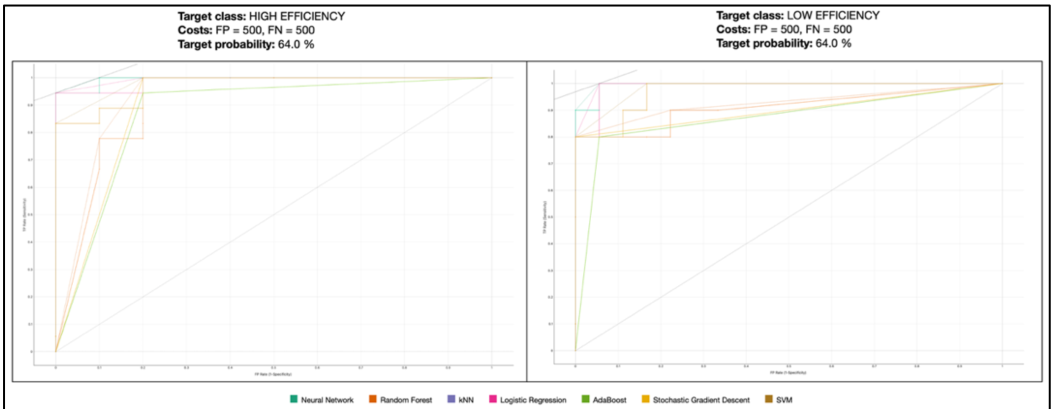
4. DISCUSSION AND RESULTS

In this section of our study, we delve into a thorough discussion of the results obtained, analyzing how various machine learning models compare in the context of hospital efficiency. We begin by examining the effectiveness of different analytical approaches, including neural networks, logistic regression, random forests, KNN, AdaBoost, Stochastic Gradient Descent, and Support Vector Machines (SVM), assessing their performance through key metrics. This comparative analysis allows us to identify the most effective model for interpreting and predicting the factors influencing hospital efficiency.
We then explore the impact of specific operational variables, such as energy costs, health and non-health personnel costs, and the number of medical devices, on the efficiency of healthcare facilities. Advanced techniques like SHAP analysis are used to decompose the relative importance of these contributing factors, providing a detailed view of which elements contribute most significantly to the efficiency or inefficiency of healthcare facilities.
Finally, we focus on analyzing the impact of these operational costs and the number of medical devices on hospital allocative efficiency, using statistical analysis methodologies to discern how different categories of expenditure and resources influence hospital performance. Using techniques like Kruskal-Wallis ANOVA and scatterplots, we offer an in-depth assessment of how such costs and resources influence healthcare service delivery.
Overall, this section aims to provide an insightful and comprehensive discussion of our research findings, with a particular focus on how data analysis and machine learning methodologies can be used to inform and improve management practices in the healthcare sector.

4.1 MACHINE LEARNING MODEL DEVELOPMENT AND COMPARISONS

We conducted a comparative analysis of various machine learning models, including Neural Network, Logistic Regression, SVM, Random Forest, Stochastic Gradient Descent (SGD), KNN, and AdaBoost, using metrics such as AUC, Accuracy, F1 Score, Precision, Recall, and MCC. The final choice of the Neural Network was guided by a combination of high performance and robustness in predictions.
In our rigorous evaluation methodology for the machine learning models, we employed the 10-fold Cross Validation technique for each of the tested models. This approach ensured a robust and reliable evaluation of each model's performance, minimizing the risk of overfitting and providing a more accurate estimate of their generalization ability.
For instance, in a study on the diagnosis of Alzheimer's Disease, various machine learning algorithms, including SGD, k-Nearest Neighbors, Logistic Regression, Random Forest, AdaBoost, Neural Network, and SVM, were successfully applied,

demonstrating the effectiveness of these models in a complex clinical context (Arjaria et al., 2022).
Additionally, another study compared seven algorithms in predicting one-year mortality and clinical progression to AIDS in a small cohort of children living with HIV, showing that machine learning models outperform logistic regression even with limited sample sizes (Rodriguez et al., 2022).
We chose to train the models on the entire dataset, consisting of 28 instances, due to the relatively limited number of data available. This decision was made to maximize the use of available information, ensuring that each model had access to the widest possible variety of data during the training phase.
The Neural Network model stood out for its exceptional performance, with an AUC of 0.994 and an Accuracy of 0.929. Its F1 Score of 0.926 highlighted an excellent balance between precision and recall. Other models, such as Logistic Regression, SVM, Random Forest, SGD, KNN, and AdaBoost, also showed good performance, but with some limitations in terms of AUC and ability to handle data complexity.
A further example of practical application of these models was provided by a study that used a machine learning model based on ultrasound image features to assess the risk of sentinel lymph node metastasis in breast cancer patients, demonstrating the high diagnostic performance of the XGBoost model (Zhang et al., 2022).
Moreover, a comparison between machine learning and logistic regression for prognostic modeling in individuals with non-specific neck pain showed that machine learning could offer an improvement in prediction performance, highlighting the greater non-linearity between baseline predictors and clinical outcome (Liew et al., 2022).
The choice of the Neural Network as the primary model was supported by this analysis, highlighting its superiority in balancing sensitivity and specificity, as well as its overall robustness in predictions.
Figure 63 illustrates the parameters and ROC curves for each model, visually reflecting their discriminative capacity.



Model	AUC	CA	F1	Prec	Recall	MCC
Neural Network	0.994	0.929	0.926	0.936	0.929	0.849
Random Forest	0.911	0.929	0.926	0.936	0.929	0.849
kNN	0.894	0.929	0.926	0.936	0.929	0.849
Logistic Regression	0.989	0.929	0.926	0.936	0.929	0.849
AdaBoost	0.872	0.893	0.891	0.893	0.893	0.764
Stochastic Gradient Descent	0.900	0.929	0.926	0.936	0.929	0.849
SVM	0.972	0.929	0.926	0.936	0.929	0.849

Figure 63 Comparative ROC Curves and Performance Parameters of Machine Learning Models.

Figure 64 graphically illustrates the Cumulative Gain Curves and associated parameters for each model, providing a visual representation of their discriminative ability and overall performance. These curves were crucial in measuring the effectiveness of the models in distinguishing between different classes, especially in scenarios with imbalanced classes.
For the high efficiency target, we observed the following results: the probability thresholds were 0.898 for the Neural Network, 0.95 for the Random Forest, 1.0 for the KNN, 0.844 for Logistic Regression, 1.0 for AdaBoost, 1.0 for Stochastic Gradient Descent, and 0.858 for SVM. The area under the curve (AUC) for the Neural Network was 0.659, while for Logistic Regression it was 0.657, both outperforming Random Forest (0.518), KNN (0.319), AdaBoost (0.304), Stochastic Gradient Descent (0.286), and SVM (0.651). For the low efficiency target, the probability thresholds were 0.108 for the Neural Network, 0.1 for the Random Forest, 0.0 for the KNN, 0.165 for Logistic Regression, 0.0 for AdaBoost, 0.0 for Stochastic Gradient Descent, and 0.15 for SVM. The AUC for the Neural Network was 0.8, while for Logistic Regression it was 0.796, both outperforming Random Forest (0.761), KNN (0.618), AdaBoost (0.543), Stochastic Gradient Descent (0.571), and SVM (0.786).
These results demonstrate the superiority of the Neural Network and Logistic Regression in recognizing both high and low efficiency instances, with a particular emphasis on the precision and accuracy of the Neural Network in both scenarios.

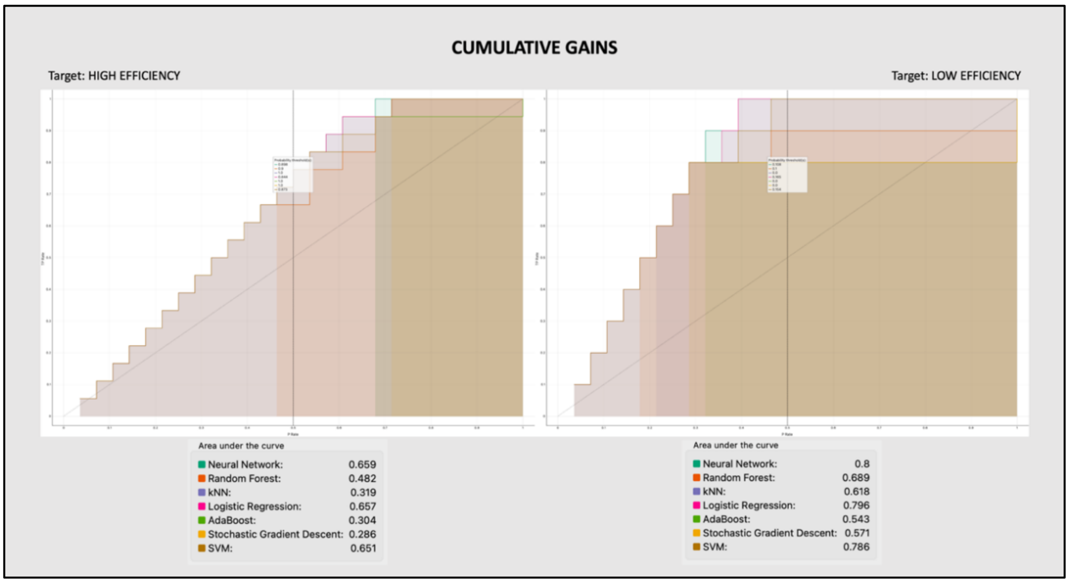


Figure 64 Cumulative Gain Curves for the Analyzed Machine Learning Models.

Figure 65 describes the Precision-Recall Curves for each model. These curves provide detailed insights into the models' performance in terms of precision and recall at various probability threshold levels. For the high efficiency target, the Neural Network shows a probability threshold of 0.009, while the Random Forest has a threshold of 0.025, KNN at 0, Logistic Regression at 0.018, AdaBoost at 0, Stochastic Gradient Descent at 0, and SVM at 0.069. The area under the curve (AUC) for the Neural Network is 0.187, comparable to that of Logistic Regression and SVM (both at 0.187), but superior to Random Forest (0.173), KNN (0.134), AdaBoost (0.284), and Stochastic Gradient Descent (0.08). This indicates the effectiveness of the Neural Network in maintaining a high level of precision while capturing a significant number of positive instances.

For the low efficiency target, the probability thresholds for the Neural Network, Random Forest, KNN, Logistic Regression, AdaBoost, Stochastic Gradient Descent, and SVM are 0.996, 0.983, 1, 0.99, 1, 1, and 0.931, respectively. In this scenario, the AUC for the Neural Network reaches 0.99, demonstrating its superiority in accurately recognizing low efficiency instances, compared to Random Forest (0.579), KNN (0.4), Logistic Regression (0.979), AdaBoost (0.978), Stochastic Gradient Descent (0.2), and SVM (0.962).

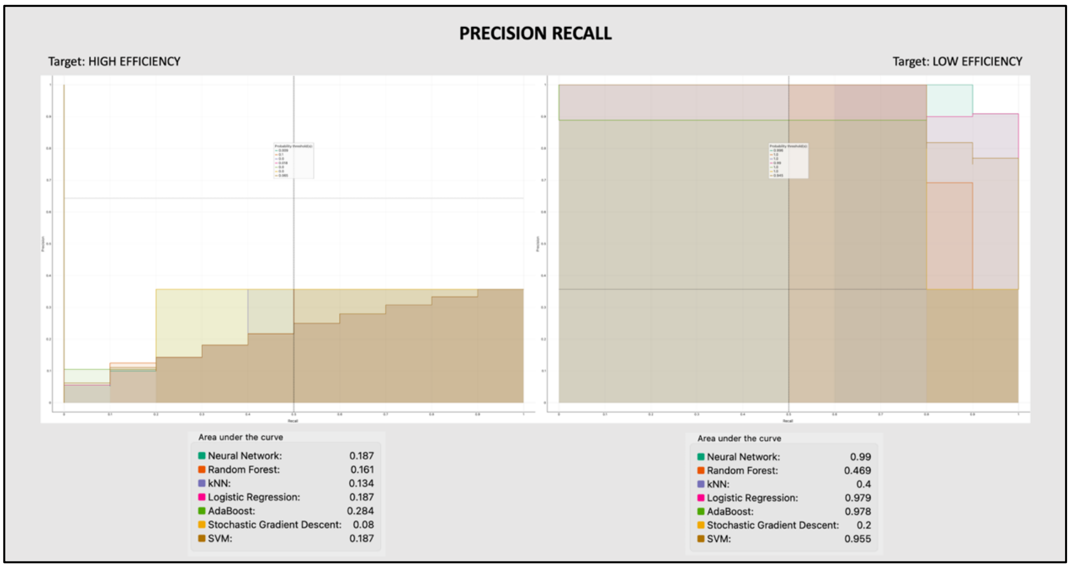


Figure 65 Precision-Recall Curves for the analyzed machine learning models.

The choice of the Neural Network as the primary model was supported by this analysis, highlighting its superiority in balancing sensitivity and specificity, as well as its overall robustness in predictions.

4.2 DECISION-MAKING PROCESS IN SELECTING THE MACHINE LEARNING MODEL

In our study, we embarked on a comprehensive comparative analysis of various machine learning models, including Neural Networks, Logistic Regression, SVM, Random Forest, Stochastic Gradient Descent (SGD), KNN, and AdaBoost. This analysis was grounded on key metrics such as AUC, Accuracy, F1 Score, Precision, Recall, and MCC.

Rigorous Evaluation Methodology: We adopted a stringent evaluation methodology, employing the 10-fold Cross Validation technique for each tested model. This approach ensured a robust and reliable assessment of each model's performance, minimizing the risk of overfitting and providing a more accurate estimation of their generalization capabilities.

Training on Full Dataset: Given the relatively limited data available (28 instances), we chose to train the models on the entire dataset. This decision was aimed at maximizing the use of available information, ensuring that each model had access to the broadest possible variety of data during the training phase.

Exceptional Performance of Neural Network: Among the analyzed models, the Neural Network stood out for its exceptional performance, as evidenced by an AUC of 0.994 and an Accuracy of 0.929. Its F1 Score of 0.926 highlighted an excellent balance between precision and recall. While other models also demonstrated good performance, they exhibited some limitations in terms of AUC and handling data complexity.

Graphical Analysis and Interpretation of Results: Figures 8, 9, and 10 provided crucial graphical representations for our analysis, illustrating the ROC curves, Cumulative Gain Curves, and Precision-Recall Curves for each model, respectively. These visualizations highlighted the discriminative capacity and overall performance of the models, particularly the superiority of the Neural Network in recognizing both high and low efficiency instances.

In conclusion, the choice of the Neural Network as the primary model was underpinned by this in-depth analysis. Its superiority in balancing sensitivity and specificity, coupled with its overall robustness in predictions, made it clear that it was the most suitable model for our study. This decision reflects a methodical, data-driven approach, focused on achieving the best possible performance.

4.3 CONFIGURATION OF THE NEURAL NETWORK MODEL PARAMETERS

For our analysis, we adopted a Neural Network as the primary model. The configuration of the Neural Network parameters was carefully selected to maximize performance and adaptability to our specific needs.

Key parameters of the Neural Network include:

1. Number of neurons in hidden layers: We chose to use 100 neurons in the hidden layers to effectively capture the complexity of the data and enable accurate learning of underlying relationships.
2. Activation function: We utilized the "RELU" (Rectified Linear Unit) activation function to introduce non-linearity into the model, which is particularly important for learning complex patterns.
3. Solver: We adopted the "adam" optimizer, known for its effectiveness in training neural networks on moderate-sized datasets.
4. Regularization: To prevent overfitting and enhance the model's generalization ability, we applied regularization with an alpha parameter set to 0.0001.
5. Maximum number of iterations: We set the maximum number of iterations during training to 200, ensuring that the model has sufficient opportunities to converge to optimal weight values.
6. Training reproducibility: We ensured the reproducibility of training results to guarantee consistency and reliability in our experiments.

All these parameters were carefully chosen to maximize the effectiveness of our Neural Network model in analyzing hospital efficiency within the available data. Their configuration was based on best practices and experimentation to ensure optimal performance.

4.4 SHAP ANALYSIS OF HOSPITAL SCALE EFFICIENCY

To deepen our understanding of efficiency in hospital management, we implemented SHAP (SHapley Additive exPlanations) analysis. This method allowed us to examine in detail the impact of various cost factors - energy consumption, costs of medical and non-medical staff, and the number of medical devices - on the efficiency predictions of our neural network model. Drawing on approaches and findings from recent studies in urban energy analysis and efficiency

forecasting (Gu et al., 2022), we identified energy consumption as a key factor in efficiency predictions, highlighting the importance of considering energy costs in any predictive model.
The analysis of staff costs and medical devices, inspired by studies with similar approaches (Rzychoń et al., 2021), clarified how each cost factor distinctly influences efficiency predictions. We also adopted feature analysis techniques from studies on interpreting predictions in complex machine learning models (Panda et al., 2023), which proved extremely useful in identifying the specific contribution of each variable.
The choice to adopt SHAP was motivated by its ability to provide clear and understandable explanations for model predictions, along with detailed visualizations of the influences of variables on predictive outcomes. This algorithm has been demonstrated to be effective across a wide range of contexts and industries, making it an ideal choice for our study on hospital efficiency.
Furthermore, we considered alternatives in interpreting neural network results. Some studies have proposed approaches based on variational autoencoders and multi-scale perceptual convolutional neural networks (Fang et al., 2022; Van De Leur et al., 2022), but these methods were more specific to applications and less adaptable to our context. Additionally, we explored other techniques for interpreting machine learning models (Zhao et al., 2022), but we chose SHAP due to its suitability for our research needs and its proven effectiveness.
Figure 66 visually illustrates, through SHAP value ribbons, how these cost factors modulate the predictive probabilities for efficiency goals. The chart highlights the features with the greatest influence on the prediction, where a longer ribbon length indicates a more significant influence. Features colored red increase the probability for a selected class, while those in blue decrease it.

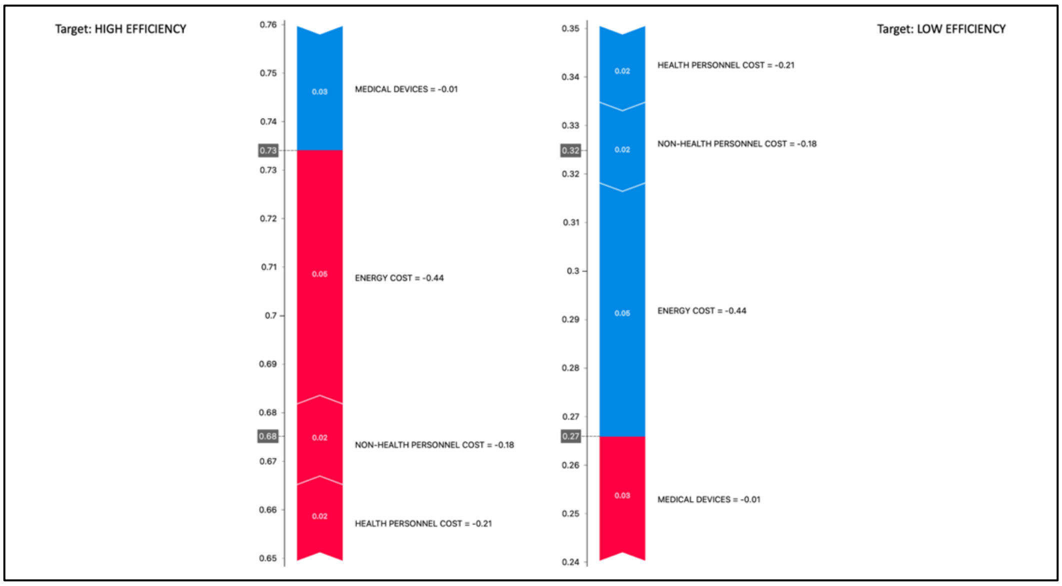


Figure 66 Visual Representation of SHAP Values for Hospital Scale Efficiency Analysis.

High Efficiency Objective Analysis:
The SHAP analysis revealed that, starting from a baseline probability of 0.68, the cost variables increased the predictive probability of high efficiency to 0.73. Energy cost (SHAP Value 0.05, impact -0.44) stands out as the most influential factor,

despite representing a significant burden. The costs of both medical (SHAP Value 0.02, impact -0.21) and non-medical staff (SHAP Value 0.02, impact -0.18) positively influence the prediction, while remaining significant expenditure factors. The number of medical devices (SHAP Value -0.03, impact -0.01) showed a lesser influence, slightly reducing the probability of high efficiency.

Low Efficiency Objective Analysis:

For the low efficiency objective, the predictive probability decreased from 0.32 to 0.27. Here, energy cost (SHAP Value -0.05, impact -0.44) played a significant negative role, reducing the probability of low efficiency. Similarly, the costs of medical (SHAP Value -0.02, impact -0.21) and non-medical staff (SHAP Value -0.02, impact -0.18) negatively influenced the prediction. In contrast, a higher number of medical devices (SHAP Value 0.03, impact -0.01) slightly increased the probability of low efficiency.

The SHAP analysis confirms that all considered cost factors significantly impact hospital efficiency, in both high and low efficiency contexts. Energy costs emerge as the most critical factor, strongly influencing both high and low efficiency predictions. These findings emphasize the importance of efficient and sustainable resource management to optimize efficiency and reduce operational costs. The reduction of energy costs, particularly relevant in low efficiency scenarios, underscores the importance of targeted strategies to enhance hospital efficiency. In this context, intelligent and sustainable resource management becomes an imperative not only economically, but also ethically and socially, to ensure efficient and responsible healthcare service. In addition to energy costs, the analysis also sheds light on the significant role of personnel costs in hospital efficiency. Both medical and non-medical staff costs, though less impactful than energy costs, are crucial elements in the overall efficiency equation. The SHAP values indicate a subtle yet positive influence of these costs on high efficiency predictions, suggesting that investment in skilled personnel may contribute to better operational outcomes. However, these costs also represent a substantial part of the hospital's budget, necessitating a balanced approach. Efficient management of personnel resources, therefore, emerges as a key factor in maintaining optimal operational efficiency. This involves not only controlling expenses but also ensuring that staff allocation aligns with the hospital's efficiency goals. The strategic deployment of personnel, coupled with effective cost management, can lead to significant improvements in service quality and patient care, ultimately enhancing the hospital's overall efficiency.

4.5 ANALYSIS OF THE IMPACT OF ENERGY COSTS, PERSONNEL COSTS, AND MEDICAL DEVICES ON HOSPITAL ALLOCATIVE EFFICIENCY

In our investigation into the impact of energy costs, healthcare and non-healthcare personnel costs, and the number of medical devices on hospital allocative efficiency, we adopted an innovative approach using the "Explain Model" widget in the Orange software. This tool proved essential for interpreting our model using the SHAP library.

The "Explain Model" widget received our already trained neural network model and reference data as input. Using this data, the widget calculated the contribution of each feature to the prediction for the classes of interest, in this case, "High Efficiency" and "Low Efficiency". This allowed us to obtain a detailed and

quantifiable understanding of the impact of each variable on the efficiency predictions of the model.
Figure 66 presents a table detailing the combined impacts of energy costs, personnel (healthcare and non-healthcare) costs, and medical devices on the allocative efficiency for each of the 28 public hospitals in the Apulia region, with a focus on the target variable "High Efficiency". The table offers a direct comparison between hospitals, highlighting how different costs affect efficiency in variable ways across different healthcare facilities.
Similarly, Figure 67 presents the results for the target variable "Low Efficiency". The table in this figure allows us to observe how the same costs influence efficiency in a low-efficiency scenario, providing an informative contrast to the results of Figure 68.
Subsequently, we integrated these results with the non-parametric Kruskal-Wallis ANOVA analysis. This statistical approach was used to further assess the significance and variability of energy costs, healthcare and non-healthcare personnel costs, and the number of medical devices in relation to hospital allocative efficiency, also considering differences between various hospital levels.
To enhance our analysis, we incorporated insights from existing literature that utilized the Kruskal-Wallis ANOVA in healthcare research. These studies effectively employed this statistical method to analyze complex healthcare data (Alqarni et al., 2022; Vassilaki et al., 2021). They demonstrate the versatility and robustness of the Kruskal-Wallis ANOVA in handling non-parametric data across diverse healthcare scenarios, reinforcing the validity of our approach in assessing the impact of various factors on hospital allocative efficiency. This integration of methodologies from broader healthcare research further substantiates our findings and offers a comprehensive perspective on the multifaceted nature of hospital efficiency.
To further enrich the analysis, we used scatterplots to explore the correlation and impact of these variables on the model's output in various hospital contexts.

Data Table Sat Nov 25 23, 16:30:40

Data instances: 28
Features: 4
Meta attributes: 7
Target: Class 'SCALE EFFICIENCY (Categorical)'

	SCALE EFFICIENCY (Categorical)	HOSPITAL	NETWORK	ASL	HOSPITAL LEVEL		(HIGH EFFICIENCY)	(LOW EFFICIENCY)	I(ENERGY COST)	I(NON-HEALTH PERSONNEL COST)	I(HEALTH PERSONNEL COST)	I(MEDICAL DEVICES)
1	LOW EFFICIENCY	Ospedale Della Murgia - Perinei Altamura (BA)	PUBLIC	ASL BA	FIRST LEVEL	HIGH EFFICIENCY	0.734071	0.265929	0.0522295	0.0166376	0.0157819	-0.0257111
2	LOW EFFICIENCY	Consorziale Policlinico Bari Bari (BA)	PUBLIC	ASL BA	SECOND LEVEL	LOW EFFICIENCY	7.2807e-05	0.999927	-0.207257	-0.163065	-0.154152	-0.150587
3	HIGH EFFICIENCY	Istituto Tumori Giovanni Paolo II Bari (BA)	PUBLIC	ASL BA	IRCCS	HIGH EFFICIENCY	0.917316	0.0826837	0.110214	0.123199	0.0758422	-0.067073
4	LOW EFFICIENCY	Ospedale Di Venere Bari (BA)	PUBLIC	ASL BA	FIRST LEVEL	LOW EFFICIENCY	0.0686638	0.931336	-0.0246769	-0.241993	-0.174157	-0.165643
5	LOW EFFICIENCY	Ospedale San Paolo Bari (BA)	PUBLIC	ASL BA	FIRST LEVEL	LOW EFFICIENCY	0.00336031	0.99664	-0.124652	-0.227984	-0.187874	-0.131263
6	HIGH EFFICIENCY	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	PUBLIC	ASL BA	IRCCS	HIGH EFFICIENCY	0.987403	0.0125973	0.0960107	0.101557	0.0755027	0.0391985
7	HIGH EFFICIENCY	Ospedale Monopoli Monopoli (BA)	PUBLIC	ASL BA	BASE LEVEL	HIGH EFFICIENCY	0.973897	0.0261032	0.0911539	0.104123	0.044408	0.0590787
8	HIGH EFFICIENCY	Ospedale Putignano Putignano (BA)	PUBLIC	ASL BA	BASE LEVEL	HIGH EFFICIENCY	0.978399	0.0216007	0.0657521	0.104804	0.0608645	0.0718449
9	LOW EFFICIENCY	Ospedale Perrino Brindisi (BR)	PUBLIC	ASL BR	SECOND LEVEL	LOW EFFICIENCY	0.00456603	0.995434	-0.273275	-0.0871845	-0.156034	-0.154074
10	HIGH EFFICIENCY	Ospedale Francavilla Fontana Francavilla Fontana (BR)	PUBLIC	ASL BR	FIRST LEVEL	HIGH EFFICIENCY	0.966696	0.0333041	0.0881392	0.096046	0.0633756	0.0440017
11	HIGH EFFICIENCY	Ospedale Ostuni Ostuni (BR)	PUBLIC	ASL BR	BASE LEVEL	HIGH EFFICIENCY	0.973454	0.0265459	0.0693651	0.0547689	0.101608	0.0725782
12	LOW EFFICIENCY	Ospedale Andria Andria (BT)	PUBLIC	ASL BT	FIRST LEVEL	LOW EFFICIENCY	0.241817	0.758183	-0.121488	-0.133612	-0.0780515	-0.100165
13	LOW EFFICIENCY	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	PUBLIC	ASL BT	FIRST LEVEL	HIGH EFFICIENCY	0.543395	0.456605	-0.152235	-0.0477301	-0.0219725	0.0901994
14	HIGH EFFICIENCY	Ospedale Bisceglie Bisceglie (BT)	PUBLIC	ASL BT	BASE LEVEL	HIGH EFFICIENCY	0.907711	0.0922894	-0.0182016	0.0956867	0.0659781	0.0891139
15	HIGH EFFICIENCY	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	PUBLIC	ASL FG	FIRST LEVEL	HIGH EFFICIENCY	0.925202	0.0747983	0.102546	0.0733506	0.026562	0.0476098
16	LOW EFFICIENCY	Ospedali Riuniti Di Foggia Foggia (FG)	PUBLIC	ASL FG	SECOND LEVEL	LOW EFFICIENCY	0.00866342	0.991337	-0.223368	-0.220669	-0.16905	-0.0533835
17	HIGH EFFICIENCY	Ospedale Manfredonia Manfredonia (FG)	PUBLIC	ASL FG	BASE LEVEL	HIGH EFFICIENCY	0.967087	0.0329133	0.104318	0.0597088	0.0667155	0.0612112
18	HIGH EFFICIENCY	Ospedale San Severo - Teresa Masselli San Severo (FG)	PUBLIC	ASL FG	FIRST LEVEL	HIGH EFFICIENCY	0.854682	0.145318	0.0835317	-0.0126381	0.00896877	0.0996857
19	HIGH EFFICIENCY	Ospedale Casarano Casarano (LE)	PUBLIC	ASL LE	BASE LEVEL	HIGH EFFICIENCY	0.926835	0.0731654	0.074013	0.0670519	0.0641138	0.0465224
20	HIGH EFFICIENCY	Ospedale Copertino Copertino (LE)	PUBLIC	ASL LE	BASE LEVEL	HIGH EFFICIENCY	0.966959	0.0330407	0.0826368	0.0787871	0.0696553	0.0607466
21	HIGH EFFICIENCY	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	PUBLIC	ASL LE	FIRST LEVEL	HIGH EFFICIENCY	0.887745	0.112255	0.0617999	0.0475401	0.0570516	0.0462196
22	LOW EFFICIENCY	Ospedale Lecce 'V. Fazzi' Lecce (LE)	PUBLIC	ASL LE	SECOND LEVEL	LOW EFFICIENCY	0.000738076	0.999262	-0.199634	-0.166769	-0.147607	-0.160385
23	HIGH EFFICIENCY	Ospedale Scorrano Scorrano (LE)	PUBLIC	ASL LE	FIRST LEVEL	HIGH EFFICIENCY	0.904068	0.0959321	0.0689755	0.0583946	0.0549539	0.0466104
24	HIGH EFFICIENCY	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	PUBLIC	ASL LE	BASE LEVEL	HIGH EFFICIENCY	0.944707	0.0552932	0.0553348	0.067078	0.0635355	0.0836251
25	HIGH EFFICIENCY	Ospedale Castellaneta Castellaneta (TA)	PUBLIC	ASL TA	FIRST LEVEL	HIGH EFFICIENCY	0.957515	0.042485	0.0933635	0.0767486	0.0669373	0.0453322
26	HIGH EFFICIENCY	Ospedale Manduria 'Giannuzzi' Manduria (TA)	PUBLIC	ASL TA	BASE LEVEL	HIGH EFFICIENCY	0.970298	0.029702	0.0796227	0.103464	0.0687791	0.0432988
27	HIGH EFFICIENCY	Ospedale Civile Martina Franca (TA)	PUBLIC	ASL TA	FIRST LEVEL	HIGH EFFICIENCY	0.973634	0.0263664	0.0863844	0.0926089	0.0604223	0.0590846
28	LOW EFFICIENCY	Presidio Ospedaliero centrale Taranto (TA)	PUBLIC	ASL TA	SECOND LEVEL	LOW EFFICIENCY	0.0024824	0.997518	-0.213355	-0.142959	-0.145342	-0.170994

Figure 67 Impact of energy and non-health personnel costs on the efficiency of 28 Apulian hospitals (Target: High Efficiency).

Data Table

Sat Nov 25 23, 16:30:58

Data instances: 28
Features: 4
Meta attributes: 7
Target: Class 'SCALE EFFICIENCY (Categorical)'

	SCALE EFFICIENCY (Categorical)	HOSPITAL	NETWORK	ASL	HOSPITAL LEVEL		(HIGH EFFICIENCY)	(LOW EFFICIENCY)	I(ENERGY COST)	I(NON-HEALTH PERSONNEL COST)	I(HEALTH PERSONNEL COST)	I(MEDICAL DEVICES)
1	LOW EFFICIENCY	Ospedale Della Murgia - Perinei Altamura (BA)	PUBLIC	ASL BA	FIRST LEVEL	HIGH EFFICIENCY	0.734071	0.265929	-0.0522295	-0.0166376	-0.0157819	0.0257111
2	LOW EFFICIENCY	Consorziale Policlinico Bari Bari (BA)	PUBLIC	ASL BA	SECOND LEVEL	LOW EFFICIENCY	7.2807e-05	0.999927	0.207257	0.163065	0.154152	0.150587
3	HIGH EFFICIENCY	Istituto Tumori Giovanni Paolo II Bari (BA)	PUBLIC	ASL BA	IRCCS	HIGH EFFICIENCY	0.917316	0.0826837	-0.110214	-0.123199	-0.0758422	0.067073
4	LOW EFFICIENCY	Ospedale Di Venere Bari (BA)	PUBLIC	ASL BA	FIRST LEVEL	LOW EFFICIENCY	0.0686638	0.931336	0.0246769	0.241993	0.174157	0.165643
5	LOW EFFICIENCY	Ospedale San Paolo Bari (BA)	PUBLIC	ASL BA	FIRST LEVEL	LOW EFFICIENCY	0.00336031	0.99664	0.124652	0.227984	0.187874	0.131263
6	HIGH EFFICIENCY	IRCCS 'Saverio De Bellis' Castellana Grotte (BA)	PUBLIC	ASL BA	IRCCS	HIGH EFFICIENCY	0.987403	0.0125973	-0.0960107	-0.101557	-0.0755027	-0.0391985
7	HIGH EFFICIENCY	Ospedale Monopoli Monopoli (BA)	PUBLIC	ASL BA	BASE LEVEL	HIGH EFFICIENCY	0.973897	0.0261032	-0.0911539	-0.104123	-0.044408	-0.0590787
8	HIGH EFFICIENCY	Ospedale Putignano Putignano (BA)	PUBLIC	ASL BA	BASE LEVEL	HIGH EFFICIENCY	0.978399	0.0216007	-0.0657521	-0.104804	-0.0608645	-0.0718449
9	LOW EFFICIENCY	Ospedale Perrino Brindisi (BR)	PUBLIC	ASL BR	SECOND LEVEL	LOW EFFICIENCY	0.00456603	0.995434	0.273275	0.0871845	0.156034	0.154074
10	HIGH EFFICIENCY	Ospedale Francavilla Fontana Francavilla Fontana (BR)	PUBLIC	ASL BR	FIRST LEVEL	HIGH EFFICIENCY	0.966696	0.0333041	-0.0881392	-0.096046	-0.0633756	-0.0440017
11	HIGH EFFICIENCY	Ospedale Ostuni Ostuni (BR)	PUBLIC	ASL BR	BASE LEVEL	HIGH EFFICIENCY	0.973454	0.0265459	-0.0693651	-0.0547689	-0.101608	-0.0725782
12	LOW EFFICIENCY	Ospedale Andria Andria (BT)	PUBLIC	ASL BT	FIRST LEVEL	LOW EFFICIENCY	0.241817	0.758183	0.121488	0.133612	0.0780515	0.100165
13	LOW EFFICIENCY	Ospedale Barletta - 'Mons. R. Dimiccoli' Barletta (BT)	PUBLIC	ASL BT	FIRST LEVEL	HIGH EFFICIENCY	0.543395	0.456605	0.152235	0.0477301	0.0219725	-0.0901994
14	HIGH EFFICIENCY	Ospedale Bisceglie Bisceglie (BT)	PUBLIC	ASL BT	BASE LEVEL	HIGH EFFICIENCY	0.907711	0.0922894	0.0182016	-0.0956867	-0.0659781	-0.0891139
15	HIGH EFFICIENCY	Ospedale Cerignola 'S.Tatarella' Cerignola (FG)	PUBLIC	ASL FG	FIRST LEVEL	HIGH EFFICIENCY	0.925202	0.0747983	-0.102546	-0.0733506	-0.026562	-0.0476098
16	LOW EFFICIENCY	Ospedali Riuniti Di Foggia Foggia (FG)	PUBLIC	ASL FG	SECOND LEVEL	LOW EFFICIENCY	0.00866342	0.991337	0.223368	0.220669	0.16905	0.0533835
17	HIGH EFFICIENCY	Ospedale Manfredonia Manfredonia (FG)	PUBLIC	ASL FG	BASE LEVEL	HIGH EFFICIENCY	0.967087	0.0329133	-0.104318	-0.0597088	-0.0667155	-0.0612112
18	HIGH EFFICIENCY	Ospedale San Severo - Teresa Masselli San Severo (FG)	PUBLIC	ASL FG	FIRST LEVEL	HIGH EFFICIENCY	0.854682	0.145318	-0.0835317	0.0126381	-0.00896877	-0.0996857
19	HIGH EFFICIENCY	Ospedale Casarano Casarano (LE)	PUBLIC	ASL LE	BASE LEVEL	HIGH EFFICIENCY	0.926835	0.0731654	-0.074013	-0.0670519	-0.0641138	-0.0465224
20	HIGH EFFICIENCY	Ospedale Copertino Copertino (LE)	PUBLIC	ASL LE	BASE LEVEL	HIGH EFFICIENCY	0.966959	0.0330407	-0.0826368	-0.0787871	-0.0696553	-0.0607466
21	HIGH EFFICIENCY	Ospedale Gallipoli 'Sacro Cuore Di Gesu' Gallipoli (LE)	PUBLIC	ASL LE	FIRST LEVEL	HIGH EFFICIENCY	0.887745	0.112255	-0.0617999	-0.0475401	-0.0570516	-0.0462196
22	LOW EFFICIENCY	Ospedale Lecce 'V. Fazzi' Lecce (LE)	PUBLIC	ASL LE	SECOND LEVEL	LOW EFFICIENCY	0.000738076	0.999262	0.199634	0.166769	0.147607	0.160385
23	HIGH EFFICIENCY	Ospedale Scorrano Scorrano (LE)	PUBLIC	ASL LE	FIRST LEVEL	HIGH EFFICIENCY	0.904068	0.0959321	-0.0689755	-0.0583946	-0.0549539	-0.0466104
24	HIGH EFFICIENCY	Ospedale Galatina 'S. Caterina Novella' Galatina (LE)	PUBLIC	ASL LE	BASE LEVEL	HIGH EFFICIENCY	0.944707	0.0552932	-0.0553348	-0.067078	-0.0635355	-0.0836251
25	HIGH EFFICIENCY	Ospedale Castellaneta Castellaneta (TA)	PUBLIC	ASL TA	FIRST LEVEL	HIGH EFFICIENCY	0.957515	0.042485	-0.0933635	-0.0767486	-0.0669373	-0.0453322
26	HIGH EFFICIENCY	Ospedale Manduria 'Giannuzzi' Manduria (TA)	PUBLIC	ASL TA	BASE LEVEL	HIGH EFFICIENCY	0.970298	0.029702	-0.0796227	-0.103464	-0.0687791	-0.0432988
27	HIGH EFFICIENCY	Ospedale Civile Martina Franca (TA)	PUBLIC	ASL TA	FIRST LEVEL	HIGH EFFICIENCY	0.973634	0.0263664	-0.0863844	-0.0926089	-0.0604223	-0.0590846
28	LOW EFFICIENCY	Presidio Ospedaliero centrale Taranto (TA)	PUBLIC	ASL TA	SECOND LEVEL	LOW EFFICIENCY	0.0024824	0.997518	0.213355	0.142959	0.145342	0.170994

Figure 68 Impact of energy and non-health personnel costs on the efficiency of 28 Apulian hospitals (Target: Low Efficiency).

We adopted a non-parametric analytical approach, using the Kruskal-Wallis ANOVA, to examine differences in the impacts of energy costs and non-medical personnel costs. This methodological choice was guided by the nature of our data, which did not adhere to the assumptions of normality required by traditional parametric ANOVA and the non-normal variability of the inputs and outputs, associated with the categorical nature of our output variable.

The Kruskal-Wallis ANOVA proved particularly effective in assessing whether there were statistically significant differences between the various hospital levels with respect to these cost variables. This non-parametric test is less sensitive to violations of the assumptions of homogeneity of variances and adapts well to data with asymmetric or unequal distributions. After identifying significant differences with the Kruskal-Wallis ANOVA, we implemented DSCF post hoc tests (Dunn's Test for Stochastic Dominance in Categorical Variables) to analyze specific differences more closely between hospital groups. These tests allowed us to isolate and examine in greater depth the incidence of energy costs and non-medical personnel costs on allocative efficiency, offering a more nuanced and detailed understanding of the dynamics at play. Figure 69 includes ten distinct tables: two showcasing the results of the Kruskal-Wallis test and eight tables detailing the outcomes of the subsequent post hoc analyses, respectively for High and Low target variables.

TARGET: HIGH EFFICIENCY

Kruskal-Wallis

	χ^2	df	p
I(ENERGY COST)	15.2	3	0.002
I(NON-HEALTH PERSONNEL COST)	15.4	3	0.002
I(HEALTH PERSONNEL COST)	18.1	3	<.001
I(MEDICAL DEVICES)	13.7	3	0.003

Dwass-Steel-Critchlow-Fligner pairwise comparisons

Pairwise comparisons - I(ENERGY COST)

		W	p
BASE LEVEL	FIRST LEVEL	-1.11	0.863
BASE LEVEL	IRCCS	2.67	0.234
BASE LEVEL	SECOND LEVEL	-4.24	0.014
FIRST LEVEL	IRCCS	2.84	0.185
FIRST LEVEL	SECOND LEVEL	-4.47	0.009
IRCCS	SECOND LEVEL	-2.74	0.213

Pairwise comparisons - I(NON-HEALTH PERSONNEL COST)

		W	p
BASE LEVEL	FIRST LEVEL	-3.42	0.074
BASE LEVEL	IRCCS	2.00	0.491
BASE LEVEL	SECOND LEVEL	-4.24	0.014
FIRST LEVEL	IRCCS	3.10	0.126
FIRST LEVEL	SECOND LEVEL	-2.83	0.187
IRCCS	SECOND LEVEL	-2.74	0.213

Pairwise comparisons - I(HEALTH PERSONNEL COST)

		W	p
BASE LEVEL	FIRST LEVEL	-4.32	0.012
BASE LEVEL	IRCCS	2.33	0.351
BASE LEVEL	SECOND LEVEL	-4.24	0.014
FIRST LEVEL	IRCCS	3.10	0.126
FIRST LEVEL	SECOND LEVEL	-2.98	0.151
IRCCS	SECOND LEVEL	-2.74	0.213

Pairwise comparisons - I(MEDICAL DEVICES)

		W	p
BASE LEVEL	FIRST LEVEL	-2.61	0.251
BASE LEVEL	IRCCS	-3.00	0.146
BASE LEVEL	SECOND LEVEL	-4.24	0.014
FIRST LEVEL	IRCCS	-1.29	0.798
FIRST LEVEL	SECOND LEVEL	-3.58	0.055
IRCCS	SECOND LEVEL	-2.19	0.408

TARGET: LOW EFFICIENCY

Kruskal-Wallis

	χ^2	df	p
I(ENERGY COST)	15.2	3	0.002
I(NON-HEALTH PERSONNEL COST)	15.4	3	0.002
I(HEALTH PERSONNEL COST)	18.1	3	<.001
I(MEDICAL DEVICES)	13.7	3	0.003

Dwass-Steel-Critchlow-Fligner pairwise comparisons

Pairwise comparisons - I(ENERGY COST)

		W	p
BASE LEVEL	FIRST LEVEL	1.11	0.863
BASE LEVEL	IRCCS	-2.67	0.234
BASE LEVEL	SECOND LEVEL	4.24	0.014
FIRST LEVEL	IRCCS	-2.84	0.185
FIRST LEVEL	SECOND LEVEL	4.47	0.009
IRCCS	SECOND LEVEL	2.74	0.213

Pairwise comparisons - I(NON-HEALTH PERSONNEL COST)

		W	p
BASE LEVEL	FIRST LEVEL	3.42	0.074
BASE LEVEL	IRCCS	-2.00	0.491
BASE LEVEL	SECOND LEVEL	4.24	0.014
FIRST LEVEL	IRCCS	-3.10	0.126
FIRST LEVEL	SECOND LEVEL	2.83	0.187
IRCCS	SECOND LEVEL	2.74	0.213

Pairwise comparisons - I(HEALTH PERSONNEL COST)

		W	p
BASE LEVEL	FIRST LEVEL	4.32	0.012
BASE LEVEL	IRCCS	-2.33	0.351
BASE LEVEL	SECOND LEVEL	4.24	0.014
FIRST LEVEL	IRCCS	-3.10	0.126
FIRST LEVEL	SECOND LEVEL	2.98	0.151
IRCCS	SECOND LEVEL	2.74	0.213

Pairwise comparisons - I(MEDICAL DEVICES)

		W	p
BASE LEVEL	FIRST LEVEL	2.61	0.251
BASE LEVEL	IRCCS	3.00	0.146
BASE LEVEL	SECOND LEVEL	4.24	0.014
FIRST LEVEL	IRCCS	1.29	0.798
FIRST LEVEL	SECOND LEVEL	3.58	0.055
IRCCS	SECOND LEVEL	2.19	0.408

Figure 69 Tables of Kruskal-Wallis ANOVA and Post Hoc Analysis for Energy and Non-Medical Personnel Cost Impacts on Hospital Efficiency.

For the "High Efficiency" variable, the Kruskal-Wallis ANOVA analysis revealed statistically significant differences in the impact values of energy costs ($\chi^2 = 15.2$, $p = 0.002$), non-health personnel costs ($\chi^2 = 15.4$, $p = 0.002$), health personnel costs ($\chi^2 = 18.1$, $p < 0.001$), and medical devices ($\chi^2 = 13.7$, $p = 0.003$). In the post hoc pairwise comparisons, significant differences were observed, for instance, between the base level and the second level for energy costs ($W = -4.24$, $p = 0.014$) and non-health personnel costs ($W = -4.24$, $p = 0.014$).
Similarly, for the "Low Efficiency" variable, the Kruskal-Wallis ANOVA analysis confirmed similar results, with statistically significant differences in the impact values of the same cost variables and medical devices. In the post hoc pairwise

comparisons, significant differences were again noted, such as between the base level and the second level for energy costs (W = 4.24, p = 0.014) and non-health personnel costs (W = 4.24, p = 0.014).
These findings underscore how the impact values of energy costs, personnel, and the use of medical devices significantly influence hospital allocative efficiency, with notable variations across different hospital levels. Understanding these dynamics is crucial for optimizing resource allocation and improving efficiency in the healthcare sector.
In this context, Figure 70 is presented, which consists of four separate scatterplots. Each scatterplot visually represents the predictive values of a neural network for four different feature variables, categorized by hospital levels.

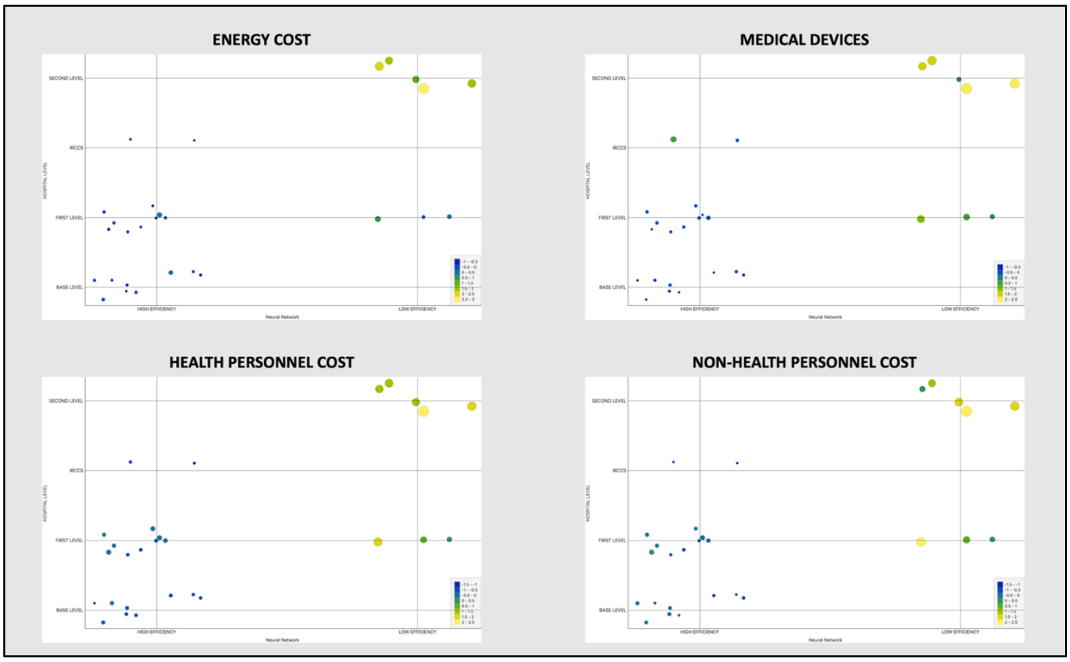


Figure 70 Scatter plots of hospital efficiency, depicting the impact of considered costs and medical devices.

The scatterplot displays the predictive values of a neural network on the x-axis, divided into High Efficiency and Low Efficiency categories, while the y-axis represents the four hospital levels: Base, First, IRCCS, and Second. The data points are color-coded with a gradient from blue to yellow to indicate different values, and their size varies in proportion to the magnitude of the value. A notable pattern emerges at the intersection of the Low Efficiency and Second Level categories, where five larger points with a yellow tint are observed, signifying a higher value in the considered category.
In all four scatter plots, the values tending towards green-yellow are predominantly concentrated in the Low Efficiency column, especially at the intersection with the Second Level row.
These scatterplots provide a clear and intuitive overview of the influence of energy costs, healthcare and non-healthcare staff costs, and the number of medical devices on hospital efficiency, offering a detailed and nuanced understanding of cost-efficiency dynamics in various hospital contexts. To further investigate this aspect, we applied the Kruskal-Wallis ANOVA to the predictive values, cost variables, and the medical device variable across the different hospital levels.

Figure 71 presents the results of the Kruskal-Wallis analysis applied to the predictive values obtained from the neural network analysis of hospitals, focusing on four feature variables. The Kruskal-Wallis analysis was conducted to examine the differences in predictive values obtained from the neural network analysis of hospitals, considering variables such as energy cost, non-health personnel cost, health personnel cost, and the number of medical devices. The results showed statistically significant differences across different hospital levels. For the Neural Network, the χ^2 value was 16.4 (df = 3, $p < .001$), for Energy Cost 15.5 (df = 3, $p = 0.001$), for Non-Health Personnel Cost 17.2 (df = 3, $p < .001$), for Health Personnel Cost 18.6 (df = 3, $p < .001$), and for Medical Devices 15.9 (df = 3, $p = 0.001$).
In pairwise comparisons, significant differences were observed between various hospital levels for each variable. For example, in the Neural Network, significant differences were found between the Base Level and the Second Level (W = 5.10, p = 0.002) and between the First Level and the Second Level (W = 3.87, p = 0.031). For Energy Cost, significant differences were observed between the Base Level and the Second Level (W = 4.24, p = 0.014) and between the First Level and the Second Level (W = 4.47, p = 0.009). Similarly, for Non-Health Personnel Cost, significant differences were found between the Base Level and the Second Level (W = 4.24, p = 0.014) and between the First Level and the Second Level (W = 3.73, p = 0.042). For Health Personnel Cost, significant differences were observed between the Base Level and the Second Level (W = 4.24, p = 0.014) and between the First Level and the Second Level (W = 4.02, p = 0.023). Finally, for Medical Devices, significant differences were found between the Base Level and the Second Level (W = 4.314, p = 0.012) and between the First Level and the Second Level (W = 4.156, p = 0.017). These results provide a detailed and nuanced understanding of cost-efficiency dynamics in different hospital contexts, highlighting how various costs and the use of medical devices influence hospital efficiency at different levels.

One-Way ANOVA (Non-parametric)

Kruskal-Wallis

	χ^2	df	p
ENERGY COST	15.5	3	0.001
NON-HEALTH PERSONNEL COST	17.2	3	<.001
HEALTH PERSONNEL COST	18.6	3	<.001
MEDICAL DEVICES	15.9	3	0.001
Neural Network	16.4	3	<.001

Dwass-Steel-Critchlow-Fligner pairwise comparisons

Pairwise comparisons - ENERGY COST

		W	p
BASE LEVEL	FIRST LEVEL	1.11	0.863
BASE LEVEL	IRCCS	-2.67	0.234
BASE LEVEL	SECOND LEVEL	4.24	0.014
FIRST LEVEL	IRCCS	-3.10	0.126
FIRST LEVEL	SECOND LEVEL	4.47	0.009
IRCCS	SECOND LEVEL	2.74	0.213

Pairwise comparisons - NON-HEALTH PERSONNEL COST

		W	p
BASE LEVEL	FIRST LEVEL	3.42	0.074
BASE LEVEL	IRCCS	-2.67	0.234
BASE LEVEL	SECOND LEVEL	4.24	0.014
FIRST LEVEL	IRCCS	-3.10	0.126
FIRST LEVEL	SECOND LEVEL	3.73	0.042
IRCCS	SECOND LEVEL	2.74	0.213

Pairwise comparisons - HEALTH PERSONNEL COST

		W	p
BASE LEVEL	FIRST LEVEL	4.02	0.023
BASE LEVEL	IRCCS	-1.67	0.641
BASE LEVEL	SECOND LEVEL	4.24	0.014
FIRST LEVEL	IRCCS	-3.10	0.126
FIRST LEVEL	SECOND LEVEL	4.02	0.023
IRCCS	SECOND LEVEL	2.74	0.213

Pairwise comparisons - MEDICAL DEVICES

		W	p
BASE LEVEL	FIRST LEVEL	3.275	0.094
BASE LEVEL	IRCCS	2.783	0.200
BASE LEVEL	SECOND LEVEL	4.314	0.012
FIRST LEVEL	IRCCS	0.942	0.910
FIRST LEVEL	SECOND LEVEL	4.156	0.017
IRCCS	SECOND LEVEL	2.191	0.408

Pairwise comparisons - Neural Network

		W	p
BASE LEVEL	FIRST LEVEL	2.24	0.390
BASE LEVEL	IRCCS	NaN	NaN
BASE LEVEL	SECOND LEVEL	5.10	0.002
FIRST LEVEL	IRCCS	-1.09	0.869
FIRST LEVEL	SECOND LEVEL	3.87	0.031
IRCCS	SECOND LEVEL	3.46	0.068

Figure 71 Kruskal-Wallis Analysis and Pairwise Comparisons Across Hospital Levels.

Following this, Figure 72 illustrates a boxplot of the distributions of the neural network's predictions. In this boxplot, Low Efficiency is highlighted in red and High Efficiency in blue, providing an immediate visual representation of the differences in predictions for these two efficiency levels. Additionally, the Base Level in blue and the Second Level in red offer a clear perspective on how predictions differ between these specific hospital levels.

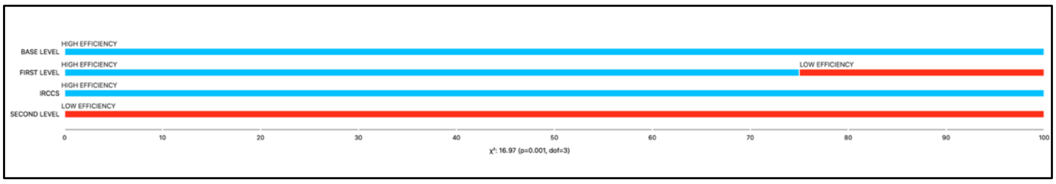


Figure 72 Box Plot of predictive values of scaling efficiency for different hospital levels.

Our comprehensive analysis of hospital efficiency, focusing on the hospital network of the Apulia region and based on neural network modeling and SHAP analysis, has provided significant insights regarding the interaction between various cost factors and their influence on the operational efficiency of hospitals:

1. **Energy Consumption**

In the Apulian hospital network, energy consumption has proven to be the most influential factor, underscoring the importance of efficient and sustainable energy management strategies. This aspect is particularly relevant in Apulia, where optimizing energy consumption can lead to significant reductions in operational costs and a lesser environmental impact. Recent studies have highlighted the importance of calculating the energy density of hospital equipment and assessing the efficiency in the use of equipment (Cakmak & Yol, 2019), as well as the importance of considering the determinants of electricity and thermal energy consumption in different hospital contexts, as demonstrated in a study conducted on Polish hospitals (Cygańska & Kludacz-alessandri, 2021).

2. **Personnel Costs**

Healthcare Personnel: In Apulia, although representing a significant expense, the costs of healthcare personnel have shown a positive influence on efficiency. Research from Romania highlights the importance of efficient management of healthcare personnel costs in maintaining hospital efficiency (Mleşnişe & Bocşan, 2016). Moreover, the expertise of medical professionals in systems like Diagnosis-Related Groups significantly contributes to reducing hospital costs and increasing profitability, further underscoring the role of healthcare personnel in improving efficiency (Iltchev et al., 2013). This indicates that targeted investments in healthcare personnel can improve operational efficiency and the quality of care.
Non-Healthcare Personnel: The costs of non-healthcare personnel have also had a positive impact on efficiency. This highlights the crucial role of support staff in optimizing hospital operations in the region.

3. **Number of Medical Devices**

In the Apulian hospital network, a greater number of medical devices slightly negatively influenced high efficiency and positively influenced low efficiency. This suggests that, although medical devices are essential, their optimal balance in management and use is crucial. The adoption of intelligent and integrated approaches for implementing energy efficiency concepts, such as the use of the

Internet of Things and other smart techniques, can be crucial in this context (Singh et al., 2022).

The integrated analysis of these factors in the Apulian hospital network highlights the importance of a holistic approach in hospital management. Efficient resource allocation, considering energy consumption, personnel costs, and the number of medical devices, is fundamental to ensuring not only operational efficiency but also the long-term sustainability of hospitals in the region. This approach allows for balancing immediate efficiency needs with goals of environmental, economic, and social sustainability, contributing to a more resilient and responsible healthcare system in Apulia.

Our findings underscore the need for integrated and sustainable management strategies specifically for the Apulian hospital network, where operational efficiency goes hand in hand with environmental sustainability and the quality of healthcare.

4.6 LIMITATIONS IN THE USE OF MACHINE LEARNING FOR ANALYZING HOSPITAL COSTS AND RESOURCES IN PUBLIC HOSPITALS

When applying machine learning to analyze energy costs, healthcare and non-healthcare personnel costs, and the number of medical devices in public hospitals, several limitations must be acknowledged. The quality of data is crucial: cost records and resource utilization data can vary in accuracy and completeness. Energy costs, for example, may be influenced by factors not evident in the data, such as infrastructure age and maintenance or seasonal energy consumption variations (Atalan et al., 2022). Similarly, healthcare personnel costs may not fully capture the complexities of staffing efficiency or the dynamics of healthcare labor markets, including contract variations and shift patterns (Rakshit et al., 2021).

The complexity of modeling direct relationships between these variables and hospital operations presents another challenge. Reductions in energy or personnel costs might result from practices that could negatively impact hospital operations, a nuance not always captured by machine learning models (Ball, 2021; Mazumdar et al., 2020). Additionally, the number of medical devices, while essential for hospital operations, may not directly correlate with efficiency, as their management and utilization are complex variables to model accurately (Philpott-Morgan et al., 2021).

The interpretability of machine learning models, particularly neural networks, is a significant limitation. The 'black box' nature of these models can hinder transparency and trust in the predictions they provide, which is critical in hospital decision-making environments (Egert et al., 2020).

Moreover, external factors such as changes in energy tariffs, labor regulations, or medical technology advancements can quickly make machine learning models outdated. This necessitates ongoing updates and reevaluations to maintain their relevance and accuracy (Hill et al., 2020).

In conclusion, while machine learning offers valuable insights into hospital cost and resource analysis, it is essential to consider these limitations. A nuanced understanding of the complex and dynamic nature of hospital environments is crucial for effectively leveraging machine learning in this context (Egert et al., 2020).

4.7 IMPLICATIONS, LIMITATIONS, AND FUTURE PERSPECTIVES OF HOSPITAL EFFICIENCY ANALYSIS

Our in-depth analysis has revealed crucial findings for the development of context-specific health policies, emphasizing the importance of considering a wide range of factors, including operational costs, staffing levels, and the number of advanced medical devices such as CT scanners and MRIs, in optimizing hospital efficiency. These results underscore the need for a holistic and data-driven approach in optimizing hospital efficiency, suggesting that adopting data-based approaches can lead to more informed decisions and greater allocative efficiency. However, it is essential to recognize the limitations of this analysis. In particular, the advanced use of machine learning techniques, such as Neural Networks, has highlighted the importance of considering a complex array of factors when evaluating hospital efficiency. Our findings show that efficiency is not influenced solely by isolated variables but rather by an interconnected network of factors, including operational costs, staff levels, energy consumption, and the use of advanced medical devices.

Specifically, SHAP analysis provided detailed insights into how various cost factors influence hospital efficiency. This approach revealed, for example, that energy costs have a greater impact compared to non-medical personnel costs in certain scenarios, while the efficient use of medical devices plays a significant role in operational efficiency. This information is crucial for developing targeted policies that consider the peculiarities of each individual hospital, rather than adopting a 'one-size-fits-all' approach.

Furthermore, the Kruskal-Wallis analysis and subsequent pairwise comparisons highlighted how allocative efficiency varies significantly across different hospital levels. This suggests that management and optimization policies should be tailored according to the specific level of each hospital, considering their unique needs and challenges.

The scatterplots further confirmed that there is no linear or simple relationship between operational costs and efficiency. This emphasizes the need for a more nuanced and contextualized analysis when developing strategies to improve efficiency. These visualizations help decision-makers quickly identify where interventions may be most effective.

The limited size of the dataset, which includes only 28 hospital instances from the Apulia region, might limit the generalizability of our results. Moreover, although machine learning models provide accurate predictions, interpreting the results requires caution, especially in decision-making contexts that impact patient health. Another limitation concerns the complexity and variability of factors not considered in the study, such as the quality of care and regional health policies.

Despite these limitations, the results of our study pave the way for future investigations and practical applications in healthcare management. Future research should expand the scope of the datasets used and explore additional influential variables. Integrating this quantitative approach with qualitative analysis could provide a more holistic understanding of challenges in the healthcare sector.

In conclusion, our findings underscore the importance of a holistic and data-driven approach in optimizing hospital efficiency. Health policies should be flexible and adaptable, capable of responding to the specificities of each hospital context. In this way, it is possible not only to improve efficiency but also to ensure that resources are allocated in a way that maximizes the positive impact on the quality of healthcare provided.

4.8 RESPONSE TO THE RESEARCH QUESTION

A_7: Our study has explored the contribution of machine learning models in analyzing and improving operational efficiency in public hospitals, with a particular focus on environmental sustainability and effective resource management. Focusing on the Apulia region, we identified and analyzed key variables such as energy costs, personnel costs (both medical and non-medical), and the number of medical devices, using neural network models.

The results demonstrated that operational efficiency in the hospital setting is influenced by a complex interplay of factors. We observed how the management of energy and personnel costs, along with the efficient use of medical devices, are crucial for promoting sustainable practices. However, it is important to emphasize that our study focused on a specific and limited sample, which could affect the generalizability of the results.

Despite these limitations, our findings offer significant insights for hospital management. They suggest that a data-based and sustainability-conscious approach can improve not only operational efficiency but also the environmental impact of hospitals. However, it is essential that further research expands the scope of these studies, including larger datasets and additional contextual variables, to confirm and deepen our findings.

In conclusion, our study responds to the research question by highlighting how machine learning models can be valuable tools in analyzing and improving the operational efficiency of hospitals, with an eye towards sustainability. These models provide a foundation for more informed decisions and management strategies that can be adapted to the specificities of each hospital context, promoting an efficiency that is sustainable both operationally and environmentally.

5. CONCLUSION OF IV SESSION

In conclusion, our study has provided significant insights into hospital efficiency, highlighting the critical importance of an integrated approach that combines advanced data analysis with contextual understanding. We demonstrated how energy costs, healthcare and non-healthcare personnel costs, and the number of medical devices such as CT scanners and MRIs, are determining factors in hospital allocative efficiency. The heterogeneity of these impacts underscores the need for differentiated and adaptable hospital management strategies, tailored to meet the specific needs of each context.

Using machine learning models, particularly neural networks, we were able to identify and quantify the impact of these operational costs and resources, offering a solid foundation for more informed strategic decisions. However, the main limitation of our study lies in the size of the dataset and its geographic specificity, which could affect the generalizability of the results. Therefore, it is essential for future research to broaden the scope of these studies, including larger datasets and additional contextual variables.

The implications of this research are wide-ranging and relevant. The findings emphasize the importance of data-driven, detail-informed hospital management that considers not only cost efficiency but also the quality of care provided and environmental sustainability. This approach can help hospitals navigate an increasingly complex healthcare landscape and continually improve their service delivery.

Moreover, our study highlights the importance of considering energy efficiency and environmental sustainability as key elements in hospital management. Integrating

sustainability into hospital management policies not only improves operational efficiency but also contributes to a more responsible environmental impact.
Ultimately, our study contributes to the emerging literature on the application of machine learning techniques in the healthcare sector, offering valuable insights for further research and hospital administrative practice. The conclusions drawn emphasize the urgency of reconsidering healthcare policies, encouraging the adoption of data-based and personalized approaches to address the unique challenges of each hospital facility. In doing so, we aspire not only to a more efficient healthcare system but also one that is sensitive to patient needs and well-being, thus promoting quality and accessible healthcare. In addition to our current findings, it is crucial to emphasize the importance of future investigations to expand and deepen the insights from our study. A key follow-up involves enlarging the dataset to include a greater number of hospitals and diversifying geographic regions. This expansion will allow for the validation of our results' generalizability and further exploration of hospital efficiency dynamics in varied contexts.
Another vital area for future research is the analysis of the impact of environmental sustainability policies on hospitals. Given the current focus on climate change and sustainability, it is imperative to assess how eco-friendly practices can be effectively integrated into hospital management, not only to reduce environmental impact but also to enhance operational efficiency. Furthermore, exploring the interplay between the quality of care and operational efficiency is important. Future studies should investigate how cost optimization affects the quality of healthcare services provided, with a keen focus on balancing economic efficiency with patient well-being.
Additionally, adopting qualitative approaches alongside quantitative ones could provide a more holistic understanding of challenges in the healthcare sector. Interviews, case studies, and ethnographic analyses could enrich our understanding of hospital dynamics, offering a more comprehensive view of the implications of our findings.

V SESSION

BUILDING THE FUTURE: DESIGNING A POLICY-CENTRIC DECISION SUPPORT SYSTEM FOR HEALTHCARE IN APULIA

1. INTRODUCTION

The evolution of Decision Support Systems (DSS) in healthcare marks a fundamental step towards more efficient and targeted care. In Tuscany, the implementation of such systems has led to significant improvements in hospital management, highlighting the potential of DSS in resource optimization and clinical decision-making (Iadanza et al., 2016). These tools, in fact, leverage advanced data analysis to provide timely and accurate decision support.

The growing relevance of DSS in the healthcare context is further underscored by their ability to integrate cutting-edge technologies for optimal patient management, as demonstrated in recent research (Reyana et al., 2021). This integration translates into personalized solutions that elevate the quality of care.

Concurrently, collaborative ontology engineering emerges as an effective approach in the development of DSS in healthcare. The semantics-based methodology, explored in 2021 publications, enhances the accuracy and relevance of the system's recommendations, emphasizing the importance of knowledge sharing in the field (Spoladore & Pessot, 2021).

Moreover, the analysis of the new frontiers of clinical decision support systems (CDSS), conducted in 2016, reveals how these tools can facilitate self-management of chronic diseases, leveraging social computing technologies to promote patient autonomy (Moon & Galea, 2016).

Beyond the healthcare context, the importance of Decision Support Systems (DSS) significantly extends to the allocation of business resources. A recent development in the IFS Applications ERP system has introduced a mathematical approach for optimizing resources across multiple projects, offering more efficient and strategic management (Fijas et al., 2023). Another study explored the use of human-centric networks to improve the allocation of human resources in information systems, emphasizing the importance of DSS in enhancing organizational efficiency (Yeon et al., 2022). In the field of cloud manufacturing, an innovative decision-making model based on the minority game has been proposed, demonstrating the effectiveness of DSS in the intelligent allocation of resources (Carlucci et al., 2020). Finally, the introduction of an intelligent information platform for human resource allocation based on fuzzy data mining algorithms has shown a significant increase in efficiency, greatly improving resource management in an organization (Peng, 2022).

These examples vividly illustrate the potential of DSS not only in the healthcare sector but also in the business context, where they offer innovative solutions for more efficient and strategic resource management. In this context, sustainability in hospitals emerges as a crucial aspect, where Decision Support Systems (DSS) can play a fundamental role. Using intelligent systems based on ontologies, DSS can contribute to designing sustainable business models in the healthcare sector (Hamrouni et al., 2021). Moreover, the implementation of real-time monitoring systems in smart agriculture and construction logistics provides valuable insights for similar applications in hospitals, aiming for more efficient and sustainable

management of resources and healthcare infrastructures (Arshad et al., 2022; Guerlain et al., 2019).
These developments underscore the importance of integrating sustainability principles into decision support systems for hospitals, ensuring that healthcare management not only meets the immediate needs of patients but does so in a responsible and sustainable manner over the long term.
Within this framework, the DSS for Apulia aims to synthesize the results of previous research sessions, offering a system that not only analyzes existing data but also provides strategic recommendations for the improvement of regional healthcare services. The goal is to create a replicable and practical model that can guide decision-makers in the healthcare sector towards more efficient, effective, and sustainable management.

2. BACKGROUND

Modern healthcare is confronted with increasing challenges related to operational efficiency, resource sustainability, and quality of care. Against this backdrop, the Decision Support System (DSS) project for Apulia emerges as a pioneering initiative, aiming to revolutionize healthcare management by leveraging advanced technologies and data analytics. This study is poised to explore the effectiveness of the DSS in synthesizing and dissecting complex datasets, thereby providing evidence-based decision-making support to augment the efficiency and efficacy of healthcare services.
The DSS for Apulia is grounded in the optimized CPDA methodology, employing optimization algorithms to accurately calculate Pure Technical Efficiency (PTE) and Scale Efficiency (SE) scores across the entire Apulian hospital network. This approach, outlined in Session I of the thesis, represents an innovative method: it integrates detailed hospital healthcare variables such as the use of advanced medical diagnostic devices (CT scans and MRIs), and cost-related variables (energy consumption and personnel management), with machine learning tools, linear regression (Session III), neural networks, and the SHAP algorithm (Session IV), crafting a complex and multi-dimensional analytical framework.
The objectives of the DSS encompass the amalgamation of heterogeneous data to attain a holistic perspective of hospital efficiency, providing decision support to optimize healthcare resources, and fostering sustainable practices. This strategy aims to enhance the quality of care by reducing waiting times and boosting patient satisfaction and proposes a replicable model for other regions, aspiring to culminate in a reference policy for the sector.
Apulia, with its unique demographic and health profile, presents a significant context for healthcare innovation. The urgency for more responsive and sustainable health systems is acutely felt, given the rising demand for high-quality healthcare services and the constraints of available resources. The DSS project positions itself as an avant-garde endeavor, striving to overcome conventional barriers in healthcare management with a data-driven and in-depth analytical decision-making system.
In addition, Session II of the thesis served as a comparative benchmark, comparing the healthcare efficiency of Apulia with that of Emilia Romagna, a region with its own healthcare challenges and successes. This comparison allowed for the contextualization of the performance of the Apulian healthcare system within a broader national landscape, enabling a deeper understanding of regional disparities and opportunities for policy transfer and adaptation.

Furthermore, Session II rigorously tested the CPDA methodology with a broader and more diverse dataset from Emilia Romagna, validating its robustness and scalability. This step was essential to ensure that the developed models and algorithms could reliably handle larger and more variable data inputs, confirming the viability of the CPDA methodology for broader application beyond the initial regional focus.
The insights gained from this comparative analysis and methodological testing in Session II significantly enriched the DSS project for Apulia. They provided a comprehensive baseline for measuring improvements and gauging the effectiveness of the DSS in enhancing healthcare management practices. Moreover, they underscored the necessity for adaptable and scalable healthcare analytics capable of addressing the complexity of healthcare data across different regional healthcare systems.
The inclusion of these insights into the thesis highlights the rigorous nature of the research process and the commitment to developing a DSS that is not only region-specific but also adaptable and applicable to the wider context of healthcare management in Italy.
In conclusion, the goal of the Decision Support System (DSS) project for Apulia is the formulation of healthcare policy. The DSS, with its advanced methodologies and data-driven decision-making capabilities, is designed to provide valuable insights and evidence-based recommendations that can inform the creation of healthcare policies. These policies aim to enhance the efficiency, effectiveness, and sustainability of healthcare services in the Apulian region, ultimately contributing to improved patient care and satisfaction. Moreover, the DSS project aspires to serve as a replicable model that can inspire policy development in other regions, further advancing healthcare management practices across Italy.
It is important to note that Session V, following the background, will be structured into three main sections. The first section will focus on the methodology of the DSS, the second will address the discussion of results and policy formulation, while the third will conclude with an analysis of project limitations and prospects. This structure reflects the comprehensive and strategic approach of the DSS project for Apulia, which aims not only to generate data but also to translate it into concrete actions to improve the regional healthcare sector.

2.1 ACHIEVING EVIDENCE-BASED HEALTHCARE POLICIES: AN ALGORITHMIC APPROACH FOR APULIA

In the pursuit of evidence-based healthcare efficiency in the Apulia region, Chapter V of this thesis will outline the architecture of a methodological algorithm for a Decision Support System (DSS). This innovative tool is grounded in seven previously formulated research questions, which probe into the dynamics between operational efficiency, energy costs, and the quality of care in the hospital setting. The goal is to bridge the gap between research and operational decision-making through the application of advanced data methodologies and analytics.
The DSS algorithm represents a pivotal step in the project for Apulia, designed to synthesize complex datasets and provide evidence-based recommendations. These recommendations have the potential to influence healthcare policies, with a particular focus on sustainability and efficient resource management. The iterative process of data collection and analysis, coupled with model development in earlier phases, culminates in this critical stage, transforming research findings into concrete insights.

By translating these insights into a policy framework, the DSS aims to enhance the efficiency, effectiveness, and sustainability of healthcare services, leveraging machine learning models to refine analysis and promote operational improvement. This holistic, data-driven approach exemplifies how decision support can indeed shape policy formulation, ultimately improving the healthcare landscape and patient care quality in the Apulia region.

3. METHODOLOGY

The DSS project for healthcare in Apulia represents an innovative initiative aimed at revolutionizing the management of healthcare services through a multidisciplinary methodological approach, based on the CPDA model (Cluster-Principal Component-Data Envelopment-ANOVA Analysis). This model, chosen for its ability to integrate advanced statistical methods and data analysis, is aimed at providing a comprehensive evaluation of healthcare performance in the region.

At the heart of the project are key stakeholders, including healthcare decision-makers and political decision-makers in Apulia, who play a crucial role in shaping and being influenced by regional healthcare policies.

The project focuses on the Apulia region, with a comparative analysis that also includes the Emilia-Romagna region, thus offering a broader context and a deeper understanding of regional dynamics. The analysis is based on data collected in the year 2020, providing a current snapshot of healthcare performance and challenges in Apulia. The initiative was launched in response to the growing needs for operational efficiency and resource sustainability in the healthcare sector, with the goal of improving the quality of care and the management of healthcare resources.

The CPDA methodology was implemented to analyze and synthesize data, using techniques of cluster analysis, principal component analysis, data envelopment analysis, and ANOVA.

Table 31. Research Questions and Methodologies in the DSS Project for Healthcare in Apulia.

Who	What	Where	When	Why	Reference Session	DSS Project Output	How	Outcome
Stakeholders, healthcare management decision-makers, political decision-makers	Q_1: How does the efficiency of hospitals in the Apulia region, calculated in terms of efficiency scores based on hospital organization, influence the perceived quality of healthcare by resident patients?	Apulia	2020	To assess the impact of hospital efficiency on the perceived quality of healthcare in Apulia and influence patient choices.	Session I	The CPDA methodology allowed us to evaluate the efficiency of hospitals in the Apulia region, identifying standardized efficiency scores that influenced hospitalization propensity.	CPDA (Cluster-PCA-DEA-ANOVA) + Regression Analysis	Policy formulation
	Q_2: Can the differences in efficiency and inefficiency among different levels of hospitals in Apulia, based on calculated hospital efficiency scores, provide support for managerial decisions with policy implications?	Apulia	2020	To understand how differences in hospital efficiency can support managerial and policy decisions in Apulia.	Session I	The analysis highlighted higher efficiency in higher-level hospitals, providing data-based support for managerial and policy decisions.	CPDA (Cluster-PCA-DEA-ANOVA)	
	Q_3: How can the application of the CPDA method, based on the combination of analysis (Cluster-Pca-Dea-Anova) in a data mining environment, enhance the discriminant capability compared to traditional DEA analysis models for hospital efficiency evaluation?	Apulia	2020	To demonstrate how the CPDA method can provide a more accurate assessment of hospital efficiency compared to traditional DEA models.	Session I	It has been demonstrated that the CPDA methodology improves discriminatory capability in hospital efficiency evaluation compared to traditional DEA models, thus providing valuable insights for improving healthcare efficiency and quality.	CPDA (Cluster-PCA-DEA-ANOVA)	
	Q_4: What is the current state of hospital efficiency in the Apulia-Emilia Romagna macroregion and how does it impact the perceived quality of healthcare by resident patients?	Apulia and Emilia-Romagna	2020-2021	To analyze the current state of hospital efficiency and its impact on the perceived quality of healthcare in Apulia and Emilia-Romagna.	Session II	Variability in hospital efficiency between the public and private sectors in Apulia-Emilia Romagna. Some facilities have achieved optimal technical and scale efficiency. A negative relationship between scale efficiency and hospitalization propensity has been identified, affecting the perceived quality of healthcare.	CPDA (Cluster-PCA-DEA-ANOVA)	
	Q_5: How do hospital organizational variables and patients' propensity for hospitalization correlate with per capita energy costs in public healthcare facilities in Apulia?	Apulia	2020	To explore the correlation between hospital organization, hospitalization propensity, and energy costs in Apulia.	Session III	Interaction between hospital facilities, hospitalization propensity, and energy costs. Efficient managerial practices increase energy expenses, especially in facilities with a high number of admissions. The need to balance patient care and energy savings.	CPDA (Cluster-PCA-DEA-ANOVA) + Regression Analysis	
	Q_6: What is the relationship between pure technical efficiency (PTE) of public hospitals in Apulia and the number of medical devices in relation to per capita energy costs?	Apulia	2020	To investigate how technical efficiency and the number of medical devices influence energy costs in public hospitals in Apulia.	Session III	The analysis has highlighted significant differences in efficiency between hospital levels, implying that technical efficiency and the number of medical devices influence energy costs.	CPDA (Cluster-PCA-DEA-ANOVA) + Regression Analysis	
	Q_7: How can machine learning models contribute to the analysis and improvement of operational efficiency in public hospitals considering the need for sustainability and efficient resource management?	Apulia	2020	To examine the contribution of machine learning models to operational efficiency in hospitals, considering sustainability and efficient resource management.	Session IV	Machine learning models, applied through the CPDA methodology, have demonstrated their significant contribution to the analysis and improvement of operational efficiency, taking into account sustainability and efficient resource management.	CPDA (Cluster-PCA-DEA-ANOVA) + Neural Network + Linear Regression	

This approach allows for a thorough and systematic examination of every aspect of the healthcare system, from resource management to the quality of care, and to formulate recommendations based on concrete evidence.

Table 31 illustrates how each research question is linked to a specific component of the CPDA analysis, ensuring a thorough and systematic examination of every aspect of the healthcare system. This table represents a fundamental element of the project, providing a clear and detailed structure for data collection and analysis, and for the formulation of concrete conclusions.

To focus on the processes used in the thesis to achieve the DSS project for healthcare in Apulia, we can examine how each research session and its related project questions contributed to the development and implementation of the system. Here is a detailed overview:

1. **Session I - Hospital Efficiency and Quality of Care:**
Process: In this phase, the efficiency of hospitals in Apulia and its impact on the perceived quality of care was evaluated. Analysis methods such as CPDA were used to calculate efficiency scores and analyze their influence on patient choices.
Contribution to DSS: This session provided fundamental data on hospital efficiency, essential for formulating recommendations in the DSS regarding the improvement of hospital management and quality of care.
2. **Session II - Comparative Analysis with Emilia-Romagna:**
Process: A comparative analysis of hospital efficiency between Apulia and Emilia-Romagna was conducted, examining variations in efficiency between the public and private sectors and their impact on the perceived quality of care.
Contribution to DSS: This comparison enriched the DSS with a broader perspective, allowing the identification of best practices and areas for improvement.
3. **Session III - Correlation between Hospital Organization, Hospitalization, and Energy Costs:**
Process: The relationship between hospital organizational variables, propensity for hospitalization, and energy costs was explored. This included analyzing the interaction between hospital facilities, admission frequency, and energy expenses.
Contribution to DSS: The results highlighted the importance of balancing patient care with energy savings, providing vital indications for the DSS on efficient resource management.
4. **Session IV - Application of Machine Learning Models:**
Process: In this phase, the contribution of machine learning models to the analysis and improvement of operational efficiency was investigated, considering sustainability and efficient resource management.
Contribution to DSS: The use of machine learning models enriched the DSS with advanced analytical capabilities, allowing for the development of recommendations based on complex data and improving operational efficiency.

Each research session played a crucial role in defining the parameters, methodologies, and analyses that formed the basis of the DSS project. Through thorough data collection and analysis, the use of advanced statistical techniques, and the application of machine learning models, the DSS project was able to develop concrete recommendations for improving the management of healthcare services in Apulia. This process ensured that the DSS was well-founded on empirical evidence and detailed analysis, making it a valuable tool for decision-makers in the healthcare sector of the region.
The 'how' of our Decision Support System is materialized through a structured workflow (Figure 73) that guides the analysis and decision-making process. The DSS workflow is defined to illustrate the analytical process and the methodologies adopted. The system begins with the collection of data from nationally certified sources, which feed the statistical and machine learning analysis engine. This engine is composed of advanced software that processes health variables through an optimized CPDA analysis path, detailed in SESSION I and tested on a broader dataset including analogous data used for Apulia and also for Emilia-Romagna, as described in SESSION II of this thesis.
The CPDA workflow initially utilizes a cluster analysis, through which healthcare structures are classified and variables are divided into two distinct clusters of inputs and outputs. This step is fundamental for the correct application of subsequent

techniques, as it allows for the organization of variables into homogeneous groups that reflect their functions within the healthcare system.
Subsequently, the variables within these clusters are subjected to optimized PCA to reduce dimensionality and identify the main components that influence efficiency and perceived quality. Optimized DEA is then applied to assess technical and scale efficiencies. ANOVA analysis follows to compare and deepen the results among different hospital groups (Grey box "CPDA ANALYSIS OPTIMIZED – SESSION I-II" in Figure 73).
In parallel, the DSS implements a linear regression analysis, which leverages the identified variables to model and interpret the relationships with the target variables, providing answers to the research questions posed in SESSIONS II and III. This process is crucial for understanding the factors that influence health policies and resource management (White boxes "LINEAR REGRESSIONS SESSION III" in Figure 73).
Furthermore, in SESSION IV, the DSS adopts a neural network analysis to examine the complex non-linear relationships between cost variables, such as those related to hospital energy costs, personnel costs, and the number of medical devices, and the target defined by the scale efficiency identified by the CPDA process (Sand-colored box "NEURAL NETWORK ANALYSIS – SESSION IV" in Figure 73).

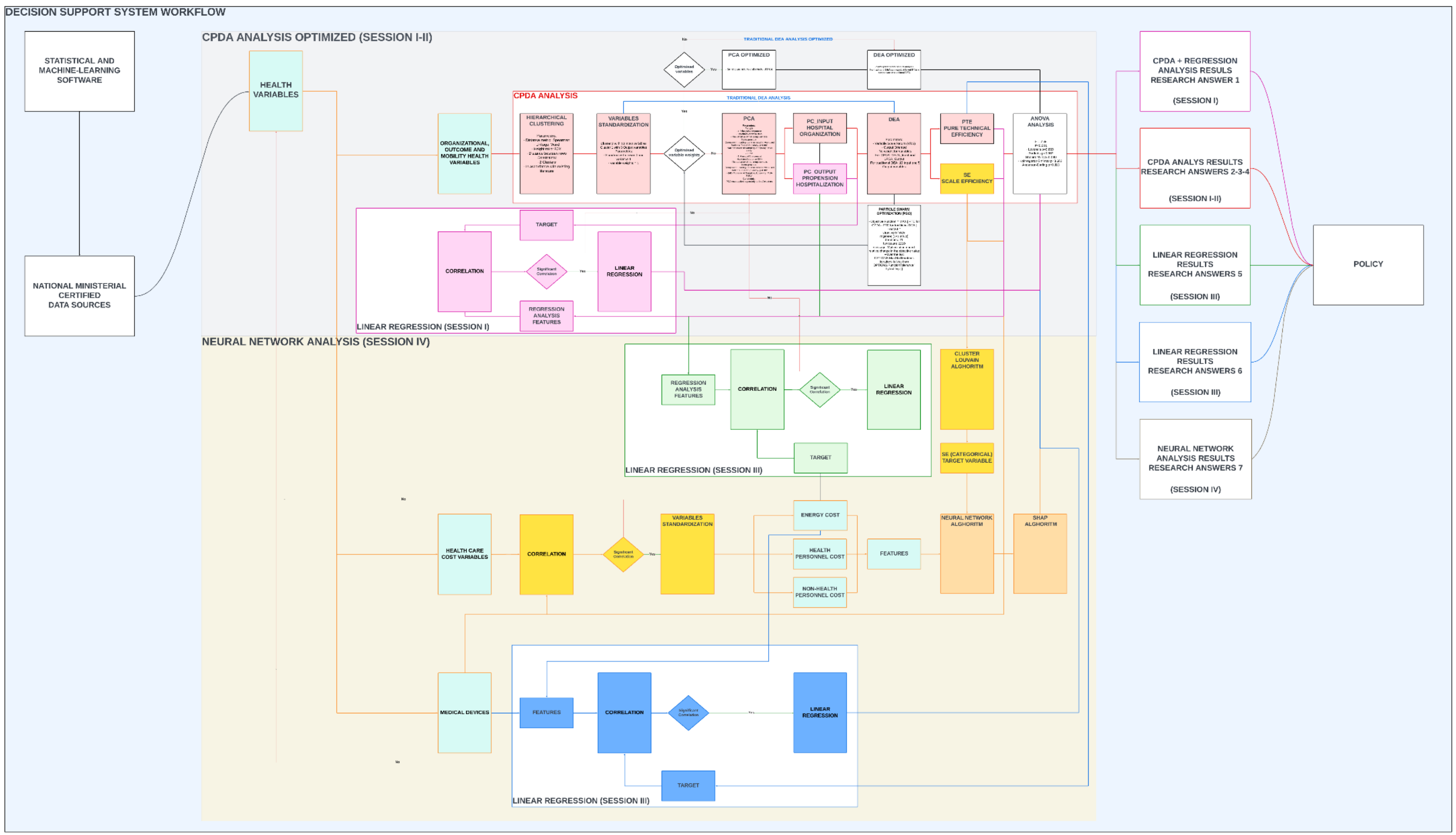


Figure 73 Workflow of the Decision Support System (DSS) for Healthcare Efficiency and Policy Development.

The Decision Support System described in this thesis represents an integrated analytical ecosystem that leverages advanced data mining techniques to draw significant insights from health data. Each component of the workflow, from cluster analysis to optimized PCA and DEA, to ANOVA analysis, plays a specific role in identifying efficiency levers and quality within the healthcare system. The linear regression and neural network analyses, detailed in SESSIONS III and IV, further enrich this framework, allowing for a detailed understanding of cost dynamics and hospital performance. Through the application of these advanced methodologies, the DSS provides decision-makers with support based on concrete data and rigorous analysis, facilitating the development of informed health policies and resource optimization in two distinct regional contexts. The synergistic work of these analytical techniques allows for a clear roadmap for continuous improvement in the healthcare field, emphasizing the importance of a data-driven approach to meet the challenges of the sector.
The workflow converges in the integration of the results of the various analyses to formulate policy recommendations. In doing so, the DSS not only provides an assessment of efficiency and perceived quality through CPDA analysis but expands its predictive and interpretive capacity with regression analysis and neural networks, ensuring that policy decisions are well-informed and based on a detailed data analysis.
In conclusion, the DSS workflow described in this session offers a complex and structured framework to guide healthcare policymakers in the Apulia region towards evidence-based decisions aimed at improving operational efficiency and patient satisfaction.

4. DISCUSSIONS AND POLICY FORMULATION FOR HOSPITAL EFFICIENCY IN THE APULIA REGION: A DSS-BASED ANALYSIS

This thesis has explored hospital efficiency and healthcare policies in the Apulia region, utilizing an advanced approach of data mining and machine learning implemented via the Decision Support System (DSS). The CPDA analysis, supported by linear regression and neural network analyses, has highlighted significant disparities in the performance of healthcare facilities, indicating key areas for targeted interventions and improvements.
The findings underscore the need for strategic reforms for more effective management of healthcare resources and enhancement of care quality. Policies should focus on optimizing resources, including training and retention of healthcare staff, and technological updating of equipment. A detailed analysis of operational costs underscores the necessity for a more rational and economical approach to resource utilization, without compromising the quality of care.
A crucial aspect is the sustainability of healthcare policies. Policies should encourage the adoption of energy-efficient technologies and personnel management practices that consider worker welfare and continuous training. This approach will contribute to a sustainable and resilient work environment, essential for maintaining high standards of care and fostering research and innovation.
Furthermore, differences between hospital levels within the healthcare network require differentiated policy approaches. While primary-level hospitals may benefit from greater integration with local services, secondary-level hospitals require targeted investments for specialization and research. Policies reflecting these

dynamics and facilitating access to quality information are essential for enabling patients to make informed health choices.
This thesis has conducted a thorough exploration of efficiency dynamics in hospitals at all levels, both private and public, in the Apulia region. The CPDA analysis, integrated with linear regression and neural network techniques, has revealed critical aspects of efficiency oriented towards the quality of services (PTE) perceived by patients. Regardless of the hospital's level or nature, there is a clear need to improve the perception of care quality through policies that emphasize effective communication, empathy, and a more patient-centered approach.
In terms of allocative efficiency (SE), the results show how resource management in Apulian hospitals, both public and private, can be optimized. Special attention must be paid to energy costs, personnel management, and the utilization of medical devices. The implications of these variables on allocative efficiency highlight the importance of judicious resource management, considering not just economic aspects but also environmental impact and sustainability.
Comparison with the Emilia Romagna region provides fundamental insights for Apulia. While Emilia Romagna stands out for more established management practices and overall higher efficiency, Apulia shows areas for improvement in cost management and resource efficiency. This interregional comparison underscores how adopting effective strategies, already successfully implemented in Emilia Romagna, could lead to significant improvements in efficiency and service quality in Apulian hospitals.
The thesis thus emphasizes the importance of a data-driven and AI-based approach for formulating effective healthcare policies. The DSS, with its capability to provide detailed analyses and interregional comparisons, emerges as an essential tool for guiding sustainable and effective strategies, aimed at improving care quality and patient experience across all levels and types of hospital networks in Apulia.
The integration of the DSS into decision-making processes can transform the way healthcare resources are allocated and managed in Apulia, shifting focus from reactive problem-solving to proactive prevention. The resulting policies should emphasize creating a healthcare system that not only efficiently responds to immediate crises but also continuously works towards improving the overall health of the population.
In conclusion, the DSS proposes a paradigmatic shift in healthcare policies in the Apulia region, steering them towards a model based on data analysis and machine learning. This allows for anticipating trends, optimizing responses, and personalizing care, ensuring that every policy decision is supported by detailed analysis and robust empirical evidence. This approach promotes a more equitable, resilient, and sustainable healthcare system, attentive to future needs and long-term sustainability.

5. CONCLUSION OF V SESSION

This thesis represented an exploratory and in-depth journey into hospital efficiency and healthcare policies in the Apulia region, highlighting the transformative potential of integrating advanced data mining and machine learning technologies. Using the Decision Support System (DSS), we identified significant disparities in the performance of healthcare facilities and outlined key areas for strategic interventions.

However, it is important to acknowledge the limitations of this study. The thesis is based on data specific to the Apulia region and, therefore, the results may not be fully generalizable to other regions or contexts with different demographic, economic, and healthcare characteristics. Additionally, the complexity of the machine learning models and data mining techniques employed could limit their applicability in contexts where such technical expertise is scarce. Another limitation concerns the availability and quality of data, which can influence the accuracy of the analyses and the conclusions drawn.
Despite these limitations, the findings emphasize the urgent need for strategic reforms aimed at more effective management of healthcare resources, with a particular focus on sustainability and resource optimization. The comparison with the Emilia-Romagna region provided valuable insights, showing how the adoption of proven management practices can improve both efficiency and service quality.
In conclusion, while taking into account the limitations, the thesis proposes a paradigm shift in healthcare policies in Apulia, underlining the vital role of a data-driven and artificial intelligence-based approach. The DSS emerges as a fundamental tool not only for detailed analysis and interregional comparisons but also as a guide for sustainable and effective strategies aimed at improving care quality and patient experience. The integration of the DSS into decision-making processes marks the beginning of a new era where healthcare resources are managed proactively, with a continuous focus on improving the health of the population. This approach promotes a fairer, more resilient, and sustainable healthcare system, ready to face future challenges and ensure the long-term sustainability of the healthcare sector in Apulia.

6. CONCLUSIONS

This thesis has explored hospital efficiency and healthcare policies in the Apulia region through an innovative approach that integrates advanced data mining and machine learning techniques. The CPDA methodology (Cluster Analysis, Principal Component Analysis, Data Envelopment Analysis, and Analysis of Variance), supported using linear regression and neural network algorithms, has enabled a detailed and robust evaluation of hospital efficiency. Optimization through Particle Swarm Optimization (PSO) has further enhanced the discriminatory power of the CPDA model, confirming its effectiveness compared to traditional methods.
The analysis of hospital efficiency in the Apulia-Emilia Romagna macroregion revealed significant differences between the two regions. Specifically, the prevalence of facilities with high technical efficiency was greater in Emilia Romagna. However, public facilities demonstrated higher scale efficiency compared to private ones, regardless of the region, suggesting a direct correlation between hospital efficiency and the perceived quality of care by patients.
The study highlighted the importance of managing energy and human resources in public hospitals, showing that greater organizational efficiency can lead to increased energy costs. This result underscores the need to balance managerial decisions between resource optimization and energy cost management. Adopting

ISO 50001 guidelines for energy management and acquiring renewable energy sources are crucial steps toward greater sustainability.
Integrating machine learning models enabled the identification and quantification of the impact of operational costs and resources on hospital allocative efficiency. However, the geographic specificity of the dataset and its limited size represent limitations that affect the generalizability of the results. Future research should include larger datasets and additional contextual variables to overcome these limitations.
The implications of this research are wide-ranging and significant. A data-driven, detailed approach to hospital management not only improves cost efficiency but also the quality of care provided and environmental sustainability. This study contributes to the emerging literature on the application of machine learning techniques in the healthcare sector, offering valuable insights for further research and hospital administrative practices. The conclusions emphasize the urgency of reconsidering healthcare policies, promoting the adoption of data-based and personalized approaches to address the unique challenges of each hospital facility.
In summary, the thesis proposes a paradigm shift in healthcare policies in Apulia, emphasizing the crucial role of an artificial intelligence-based approach. The implementation of the Decision Support System (DSS) emerges as a fundamental tool for detailed analysis and interregional comparisons, guiding sustainable and effective strategies to improve care quality and patient experience. This approach promotes a more equitable, resilient, and sustainable healthcare system, ready to face future challenges and ensure the long-term sustainability of the healthcare sector in Apulia.

REFERENCES

A Rahim, A. I., Ibrahim, M. I., Musa, K. I., Chua, S.-L., & Yaacob, N. M. (2021). Assessing Patient-Perceived Hospital Service Quality and Sentiment in Malaysian Public Hospitals Using Machine Learning and Facebook Reviews. International Journal of Environmental Research and Public Health, 18(18), 9912. https://doi.org/10.3390/ijerph18189912

Abe, T. K., Beamon, B. M., Storch, R. L., & Agus, J. (2016). Operations research applications in hospital operations: Part I. IIE Transactions on Healthcare Systems Engineering, 6(1), 42–54. https://doi.org/10.1080/19488300.2015.1134727

Aloh, H. E., Onwujekwe, O. E., Aloh, O. G., & Nweke, C. J. (2020). Is bed turnover rate a good metric for hospital scale efficiency? A measure of resource utilization rate for hospitals in Southeast Nigeria. Cost Effectiveness and Resource Allocation: C/E, 18, 21. https://doi.org/10.1186/s12962-020-00216-w

Alomari, O. A., Elnagar, A., Afyouni, I., Shahin, I., Nassif, A. B., Hashem, I. A., & Tubishat, M. (2022). Hybrid Feature Selection Based on Principal Component Analysis and Grey Wolf Optimizer Algorithm for Arabic News Article Classification. IEEE Access, 10, 121816–121830. Scopus. https://doi.org/10.1109/ACCESS.2022.3222516

Alqarni, T., Alghamdi, A., Alzahrani, A., Abumelha, K., Alqurashi, Z., & Alsaleh, A. (2022). Prevalence of stress, burnout, and job satisfaction among mental healthcare professionals in Jeddah, Saudi Arabia. PLoS ONE, 17(4 April). https://doi.org/10.1371/journal.pone.0267578

Andrews, A. (2022). An application of PCA-DEA with the double-bootstrap approach to estimate the technical efficiency of New Zealand District Health Boards. Health Economics, Policy and Law, 17(2), 175–199. Scopus. https://doi.org/10.1017/S1744133120000420

Androutsou, L., Kokkinos, M., Latsou, D., & Geitona, M. (2022). Assessing the Efficiency and Productivity of the Hospital Clinics on the Island of Rhodes during the COVID-19 Pandemic. International Journal of Environmental Research and Public Health, 19(23). https://doi.org/10.3390/ijerph192315640

Arjaria, S. K., Rathore, A. S., Bisen, D., & Bhattacharyya, S. (2022). Performances of Machine Learning Models for Diagnosis of Alzheimer's Disease. Annals of Data Science. https://doi.org/10.1007/s40745-022-00452-2

Atalan, A., Şahin, H., & Atalan, Y. A. (2022). Integration of Machine Learning Algorithms and Discrete-Event Simulation for the Cost of Healthcare Resources. Healthcare (Switzerland), 10(10). https://doi.org/10.3390/healthcare10101920

Bağci, H., & Koçyiğit, S. Ç. (2022). EFFECTS OF RECONSTRUCTION OF HEALTH CARE ON SERVICE DELIVERY PERFORMANCE IN TURKEY: THE PUBLIC HOSPITAL UNIONS. Ankara Medical Journal, 1, 140–154. Scopus. https://doi.org/10.5505/amj.2022.98958

Balia, S., Brau, R., & Marrocu, E. (2018). Interregional patient mobility in a decentralized healthcare system. Regional Studies, 52(3), 388-402. https://doi.org/10.1080/00343404.2017.1307954

Ball, H. C. (2021). Improving Healthcare Cost, Quality, and Access Through Artificial Intelligence and Machine Learning Applications. Journal of Healthcare Management, 66(4), 271–279. https://doi.org/10.1097/JHM-D-21-00149

Barra, C., Lagravinese, R., & Zotti, R. (2022). Exploring hospital efficiency within and between Italian regions: New empirical evidence. Journal of Productivity Analysis, 57(3), 269–284. Scopus. https://doi.org/10.1007/s11123-022-00633-4

Berta, P., Guerriero, C., & Levaggi, R. (2021). Hospitals' strategic behaviours and patient mobility: Evidence from Italy. Socio-Economic Planning Sciences, 77. Scopus. https://doi.org/10.1016/j.seps.2021.101030

Berta, P., Martini, G., Moscone, F., & Vittadini, G. (2016). The association between asymmetric information, hospital competition and quality of healthcare: Evidence from Italy. Journal of the Royal Statistical Society. Series A: Statistics in Society, 179(4), 907–926. Scopus. https://doi.org/10.1111/rssa.12214

Bleich, S. N., Ozaltin, E., & Murray, C. K. L. (2009). How does satisfaction with the health-care system relate to patient experience? Bulletin of the World Health Organization, 87(4), 271–278. https://doi.org/10.2471/blt.07.050401

Borges de Oliveira, K., dos Santos, E. F., Neto, A. F., de Mello Santos, V. H., & de Oliveira, O. J. (2021). Guidelines for efficient and sustainable energy management in hospital buildings. Journal of Cleaner Production, 329, 129644. https://doi.org/10.1016/j.jclepro.2021.129644

Brenna, E., & Spandonaro, F. (2015). Regional Incentives and Patient Cross-Border Mobility: Evidence from the Italian Experience. International Journal of Health Policy and Management, 4(6), 363–372. https://doi.org/10.15171/ijhpm.2015.65

Briestensky, R., & Kljucnikov, A. (2021). "The impact of DRG-based management of healthcare facilities on amenable mortality in the European Union". Problems and Perspectives in Management, 19(2), 264–275. Scopus. https://doi.org/10.21511/ppm.19(2).2021.22

Cakmak, B., & Yol, S. (2019). Medical device energy consumption analysis. TIPTEKNO 2019 - Tip Teknolojileri Kongresi. https://doi.org/10.1109/TIPTEKNO.2019.8895048

Chang, I., Park, H., Hong, E., Lee, J., & Kwon, N. (2022). Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP. Accident Analysis and Prevention, 166. https://doi.org/10.1016/j.aap.2021.106545

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Chen, L., Liang, X., & Li, T. (2015). Collaborative Performance Research on Multi-level Hospital Management Based on Synergy Entropy-HoQ. Entropy, 17(4), Articolo 4. https://doi.org/10.3390/e17042409

Cheng, S., Quilodran-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., Xiao, D., Janjic, T., Ding, W., Guo, Y., Carrassi, A., Bocquet, M., & Arcucci, R. (2023). Machine Learning With Data Assimilation and Uncertainty Quantification for Dynamical Systems: A Review. IEEE/CAA Journal of Automatica Sinica, 10(6), 1361–1387. https://doi.org/10.1109/JAS.2023.123537

Chisari, G., & Lega, F. (2023). Impact of austerity programs: Evidence from the Italian national health service. Health Services Management Research, 36(2), 145–152. https://doi.org/10.1177/09514848221134473

Colombi, R., Martini, G., & Vittadini, G. (2017). Determinants of transient and persistent hospital efficiency: The case of Italy. Health Economics (United Kingdom), 26, 5–22. Scopus. https://doi.org/10.1002/hec.3557

Cordero, J. M., García-García, A., Lau-Cortés, E., & Polo, C. (2023). Assessing Panamanian hospitals' performance with alternative frontier methods. International Transactions in Operational Research, 30(1), 394–420. Scopus. https://doi.org/10.1111/itor.13013

Cygańska, M., & Kludacz-alessandri, M. (2021). Determinants of electrical and thermal energy consumption in hospitals according to climate zones in poland. Energies, 14(22). https://doi.org/10.3390/en14227585

Dansana, J., Kabat, M. R., & Pattnaik, P. K. (2023). Improved 3D Rotation-based Geometric Data Perturbation Based on Medical Data Preservation in Big Data. International Journal of Advanced Computer Science and Applications, 14(5), 867–879. Scopus. https://doi.org/10.14569/IJACSA.2023.0140592

Devi, R. M., Premkumar, M., Jangir, P., Elkotb, M. A., Elavarasan, R. M., & Nisar, K. S. (2022). IRKO: An improved Runge-Kutta optimization algorithm for global optimization problems. Computers, Materials and Continua, 70(3), 4803–4827. Scopus. https://doi.org/10.32604/cmc.2022.020847

Dihan, M. R., Abu Nayeem, S. M., Roy, H., Islam, M. S., Islam, A., Alsukaibi, A. K. D., & Awual, M. R. (2023). Healthcare waste in Bangladesh: Current status, the impact of Covid-19 and sustainable management with life cycle and circular economy framework. Science of the Total Environment, 871. https://doi.org/10.1016/j.scitotenv.2023.162083

Ding, J., Yang, C., Wang, Y., Li, P., Wang, F., Kang, Y., Wang, H., Liang, Z., Zhang, J., Han, P., Wang, Z., Chu, E., Li, S., & Zhang, L. (2023). Influential factors of intercity patient mobility and its network structure in China. Cities, 132, 103975. https://doi.org/10.1016/j.cities.2022.103975

Dion, H., & Evans, M. (2023). Strategic frameworks for sustainability and corporate governance in healthcare facilities; approaches to energy-efficient hospital management. Benchmarking. https://doi.org/10.1108/BIJ-04-2022-0219

Dion, H., Evans, M., & Farrell, P. (2023). Hospitals management transformative initiatives; towards energy efficiency and environmental sustainability in healthcare facilities. Journal of Engineering, Design and Technology, 21(2), 552–584. https://doi.org/10.1108/JEDT-04-2022-0200

Doyle, C., Lennox, L., & Bell, D. (2013). A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open, 3(1), e001570. https://doi.org/10.1136/bmjopen-2012-001570

Dubey, S., Tiwari, G., Singh, S., Goldberg, S., & Pinsky, E. (2023). Using machine learning for healthcare treatment planning. Frontiers in Artificial Intelligence, 6. https://doi.org/10.3389/frai.2023.1124182

Dulce-Chamorro, E., & Martinez-de-Pison, F. J. (2021). An advanced methodology to enhance energy efficiency in a hospital cooling-water system. Journal of Building Engineering, 43, 102839. https://doi.org/10.1016/j.jobe.2021.102839

Egert, M., Steward, J. E., & Sundaram, C. P. (2020). Machine Learning and Artificial Intelligence in Surgical Fields. Indian Journal of Surgical Oncology, 11(4), 573–577. https://doi.org/10.1007/s13193-020-01166-8

Ekwonwune, E. N., Ubochi, C. I., & Duroha, A. E. (2022). Data Mining as a Technique for Healthcare Approach. International Journal of Communications, Network and System Sciences, 15(9), Articolo 9. https://doi.org/10.4236/ijcns.2022.159011

Eiriz, V., & António Figueiredo, J. (2005). Quality evaluation in health care services based on customer-provider relationships. International Journal of Health Care Quality Assurance, 18(6), 404–412. https://doi.org/10.1108/09526860510619408

Emrouznejad, A., & Thanassoulis, E. (2013). Introduction to performance improvement management software (PIM-DEA). In Handbook of Research on Strategic Performance Management and Measurement Using Data Envelopment Analysis (pp. 256–275). Scopus. https://doi.org/10.4018/978-1-4666-4474-8.ch005

Esposito, D., & Abbattista, I. (2020). Dynamic network visualization of space use patterns to support agent-based modelling for spatial design. In Cooperative Design, Visualization, and Engineering: 17th International Conference, CDVE 2020, Bangkok, Thailand, October 25–28, 2020, Proceedings 17 (pp. 260-269). Springer International Publishing. http://dx.doi.org/10.1007/978-3-030-60816-3_29

Esposito, D., Schaumann, D., Camarda, D., & Kalay, Y. E. (2020). Decision support systems based on multi-agent simulation for spatial design and management of a built environment: the case study of hospitals. In Computational Science and Its Applications–ICCSA 2020: 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part III 20 (pp. 340-351). Springer International Publishing. http://dx.doi.org/10.1007/978-3-030-58808-3_25

Evangelista, V. (2016). The geographics of patients transfers: The case of an Italian Regional Health System. GeoJournal, 81(5), 771–778. Scopus. https://doi.org/10.1007/s10708-015-9662-2

Falavigna, G., & Ippoliti, R. (2013). The efficiency of regional health care systems in Italy: Public or private? Economia e Politica Industriale, 40(2), 29–49. https://doi.org/10.3280/poli2013-002002

Fang, C., Wang, Z., Song, X., Zhu, Z., Yang, D., & Liu, M. (2022). A Novel Cementing Quality Evaluation Method Based on Convolutional Neural Network. Applied Sciences (Switzerland), 12(21). https://doi.org/10.3390/app122110997

Fang, T., & Lahdelma, R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Applied Energy, 179, 544–552. https://doi.org/10.1016/j.apenergy.2016.06.133

Ferreira, D. C., Figueira, J. R., Greco, S., & Marques, R. C. (2023). Data Envelopment Analysis models with imperfect knowledge of input and output values: An application to Portuguese public hospitals. Expert Systems with Applications, 231, 120543. https://doi.org/10.1016/j.eswa.2023.120543

Fiedler, L., Shah, K., & Cangi, A. (2023). Machine-Learning for Static and Dynamic Electronic Structure Theory. Challenges and Advances in Computational Chemistry and Physics, 36, 113–160. https://doi.org/10.1007/978-3-031-37196-7_5

Fumbwe, F., Lihawa, R., Andrew, F., Kinyanjui, G., & Mkuna, E. (2021). Examination on level of scale efficiency in public hospitals in Tanzania. Cost Effectiveness and Resource Allocation, 19(1). https://doi.org/10.1186/s12962-021-00305-4

Fumo, N., & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. Renewable and Sustainable Energy Reviews, 47, 332–343. https://doi.org/10.1016/j.rser.2015.03.035

Ghasemi, M., Akbari, M.-A., Jun, C., Bateni, S. M., Zare, M., Zahedi, A., Pai, H.-T., Band, S. S., Moslehpour, M., & Chau, K.-W. (2022). Circulatory System Based Optimization (CSBO): An expert multilevel biologically inspired meta-heuristic algorithm. Engineering Applications of Computational Fluid Mechanics, 16(1), 1483–1525. https://doi.org/10.1080/19942060.2022.2098826

Ghosh, S., Halappanavar, M., Tumeo, A., & Kalyanarainan, A. (2019). Scaling and quality of modularity optimization methods for graph clustering. 2019 IEEE High Performance Extreme Computing Conference, HPEC 2019. https://doi.org/10.1109/HPEC.2019.8916299

Grennan, M., Kim, G. H., McConnell, K. J., & Swanson, A. (2022). Hospital management practices and medical device costs. Health Services Research, 57(2), 227–236. https://doi.org/10.1111/1475-6773.13898

Gu, C., Liu, H., & Ou, B. (2022). Analysis of Urban Social Electricity Consumption Based on Neural Network and SHAP Model. 213–218. https://doi.org/10.1109/ISPCEM57418.2022.00049

Guede-Cid, R., Rodas-Alfaya, L., Leguey-Galán, S., & Cid-Cid, A. I. (2021). Innovation efficiency in the spanish service sectors, and open innovation. Journal of Open Innovation: Technology, Market, and Complexity, 7(1), 1–18. Scopus. https://doi.org/10.3390/joitmc7010062

Gupta, D., Dhar, A. R., & Roy, S. S. (2021). A partition cum unification based genetic- firefly algorithm for single objective optimization. Sadhana - Academy Proceedings in Engineering Sciences, 46(3). Scopus. https://doi.org/10.1007/s12046-021-01641-0

Gutierrez-Romero, G., Blanco-Oliver, A., Montero-Romero, M. T., & Carbonero-Ruz, M. (2021). The impact of ceos' gender on organisational efficiency in the public sector: Evidence from the english nhs. Sustainability (Switzerland), 13(4), 1–15. Scopus. https://doi.org/10.3390/su13042188

Hajiagha, S. H. R., Amoozad Mahdiraji, H., Hashemi, S. S., Garza-Reyes, J. A., & Joshi, R. (2023). Public Hospitals Performance Measurement through a Three-Staged Data Envelopment Analysis Approach: Evidence from an Emerging Economy. Cybernetics and Systems, 54(1), 1–26. Scopus. https://doi.org/10.1080/01969722.2022.2055382

Hasni, M., Aissaoui, N., Layeb, S. B., & Manai, A. (2021). A Clustering-Based Data Envelopment Analysis for Learning Processes Performance Assessment in Teaching Hospitals. 2021 1st International Conference On Cyber Management and Engineering, CyMaEn 2021. Scopus. https://doi.org/10.1109/CyMaEn50288.2021.9497298

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer. https://doi.org/10.1007/978-0-387-84858-7

Henriques, C. O., & Gouveia, M. C. (2022). Assessing the impact of COVID-19 on the efficiency of Portuguese state-owned enterprise hospitals. Socio-Economic Planning Sciences, 84. Scopus. https://doi.org/10.1016/j.seps.2022.101387

Hill, N. R., Sandler, B., Mokgokong, R., Lister, S., Ward, T., Boyce, R., Farooqui, U., & Gordon, J. (2020). Cost-effectiveness of targeted screening for the identification of patients with atrial fibrillation: Evaluation of a machine learning risk prediction algorithm. Journal of Medical Economics, 23(4), 386–393. https://doi.org/10.1080/13696998.2019.1706543

Iltchev, P., Sierocka, A., Gierczyński, S., & Marczak, M. (2013). The knowledge of medical professionals from selected hospitals in the Lubelskie province about diagnosis-related groups systems. Studies in Logic, Grammar and Rhetoric, 35(48), 191–201. https://doi.org/10.2478/slgr-2013-0044

ISTAT Database (https://demo.istat.it/app/?i=POS).

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065). https://doi.org/10.1098/rsta.2015.0202

Jung, S., Son, J., Kim, C., & Chung, K. (2023). Efficiency Measurement Using Data Envelopment Analysis (DEA) in Public Healthcare: Research Trends from 2017 to 2022. Processes, 11(3). Scopus. https://doi.org/10.3390/pr11030811

Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. International Journal of Electrical Power & Energy Systems, 67, 431–438. https://doi.org/10.1016/j.ijepes.2014.12.036

Kent, J. S., & Ménager, D. H. (2023). Indecision Trees: Learning Argument-Based Reasoning under Quantified Uncertainty. 12529. https://doi.org/10.1117/12.2652598

Klemen, M., Carreira-Perpiñán, M. Á., & Lopez-Garcia, P. (2023). Solving Recurrence Relations using Machine Learning, with Application to Cost Analysis. 385, 155–168. https://doi.org/10.4204/EPTCS.385.16

Koh, H. C., & Tan, G. (2005). Data mining applications in healthcare. Journal of Healthcare Information Management: JHIM, 19(2), 64–72.

Kucsma, D., & Varga, K. (2021). Exploring Effectiveness Reserves in Hospitals with the DEA Method. Public Finance Quarterly, 66(2), 75–87. Scopus. https://doi.org/10.35551/PFQ_2021_S_2_4

Kumar, M., Swaminathan, K., Rusli, A., & Thomas-Hy, A. (2022, ottobre 14). Applying Data Analytics & Machine Learning Methods for Recovery Factor Prediction and Uncertainty Modelling. SPE Asia Pacific Oil & Gas Conference and Exhibition. https://doi.org/10.2118/210769-MS

Lan, T., Chen, T., Hu, Y., Yang, Y., & Pan, J. (2021). Governmental Investments in Hospital Infrastructure Among Regions and Its Efficiency in China: An Assessment of Building Construction. Frontiers in Public Health, 9. https://www.frontiersin.org/articles/10.3389/fpubh.2021.719839

Liew, B. X. W., Kovacs, F. M., Rügamer, D., & Royuela, A. (2022). Machine learning versus logistic regression for prognostic modelling in individuals with non-specific neck pain. European Spine Journal, 31(8), 2082–2091. https://doi.org/10.1007/s00586-022-07188-w

Lippi Bruni, M., Ugolini, C., & Verzulli, R. (2021). Should I wait or should I go? Travelling versus waiting for better healthcare. Regional Science and Urban Economics, 89, 103697. https://doi.org/10.1016/j.regsciurbeco.2021.103697

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. 2017-December, 4766–4775.

Ma, Y., Tian, Y., Le, J., Guo, M., Zhang, H., & Zhang, C. (2023). High-Dimensional Multiobjective Optimization of an Aeroengine Combustor Based on Cubic Polynomial. Journal of Aerospace Engineering, 36(2). Scopus. https://doi.org/10.1061/JAEEEZ.ASENG-4434

Mazumdar, M., Lin, J.-Y. J., Zhang, W., Li, L., Liu, M., Dharmarajan, K., Sanderson, M., Isola, L., & Hu, L. (2020). Comparison of statistical and machine learning models for healthcare cost data: A simulation study motivated by Oncology Care Model (OCM) data. BMC Health Services Research, 20(1). https://doi.org/10.1186/s12913-020-05148-y

Mirmozaffari, M., Yazdani, R., Shadkam, E., Khalili, S. M., Mahjoob, M., & Boskabadi, A. (2022). An integrated artificial intelligence model for efficiency assessment in pharmaceutical companies during the COVID-19 pandemic. Sustainable Operations and Computers, 3, 156–167. Scopus. https://doi.org/10.1016/j.susoc.2022.01.003

Mleşnişe, M., & Bocşan, I. S. (2016). Cost-efficiency analysis of a multi-pavilion hospital in Cluj County. Clujul Medical, 89(1), 110–116. https://doi.org/10.15386/cjmed-606

Mohammed, M. A., Akawee, M. M., Saleh, Z. H., Hasan, R. A., Ali, A. H., & Sutikno, T. (2023). The effectiveness of big data classification control based on principal component analysis. Bulletin of Electrical Engineering and Informatics, 12(1), 427–434. Scopus. https://doi.org/10.11591/eei.v12i1.4405

Mohanta, K. K., Sharanappa, D. S., & Aggarwal, A. (2021). Efficiency analysis in the management of COVID-19 pandemic in India based on data envelopment analysis. Current Research in Behavioral Sciences, 2, 100063. https://doi.org/10.1016/j.crbeha.2021.100063

Morano, P., Tajani, F., & Anelli, D. (2021). Urban planning variants: A model for the division of the activated “plusvalue” between public and private subjects. Valori e Valutazioni, 2021(28), 31–47. Scopus.

Nante, N., Guarducci, G., Lorenzini, C., Messina, G., Carle, F., Carbone, S., & Urbani, A. (2021, September). Inter-Regional Hospital Patients’ Mobility in Italy. In Healthcare (Vol. 9, No. 9, p. 1182). MDPI. https://doi.org/10.3390/healthcare9091182

National Outcomes Plan (https://pne.agenas.it/home).

Nemani, V., Biggio, L., Huan, X., Hu, Z., Fink, O., Tran, A., Wang, Y., Zhang, X., & Hu, C. (2023). Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial. Mechanical Systems and Signal Processing, 205. https://doi.org/10.1016/j.ymssp.2023.110796

NHS Database (https://www.salute.gov.it/portale/documentazione/usldb/reguslDB.jsp?hname=hvalue&reg=160).

Ninamango Origuela, N. A., & Sovero Rivera, L. E. (2022). Efficiency of Peruvian regions in the management of hospital resources against COVID-19 in 2020: A two stage DEA model. 53–58. https://doi.org/10.1145/3524338.3524347

Noudeh, G. D., Asdaghi, M., Noudeh, N. D., Dolatabadi, M., & Ahmadzadeh, S. (2022). Response surface modeling of ceftriaxone removal from hospital wastewater. Environmental Monitoring and Assessment, 195(1), 217. https://doi.org/10.1007/s10661-022-10808-z

Nygren Zotterman, A., Skär, L., Olsson, M., & Söderberg, S. (2016). Being in togetherness: Meanings of encounters within primary healthcare setting for patients living with long-term illness. Journal of Clinical Nursing, 25(19–20), 2854–2862. Scopus. https://doi.org/10.1111/jocn.13333

Obeidat, S. M., Abdalla, S., & Al Bakri, A. A. K. (2023). Integrating green human resource management and circular economy to enhance sustainable performance: An empirical study from the Qatari service sector. Employee Relations, 45(2), 535–563. https://doi.org/10.1108/ER-01-2022-0041

Oliver, A., & Mossialos, E. (2005). European health systems reforms: Looking backward to see forward? Journal of Health Politics, Policy and Law, 30(1–2), 7–28. https://doi.org/10.1215/03616878-30-1-2-7

Panagiotou, D. K., & Dounis, A. I. (2022). Comparison of Hospital Building's Energy Consumption Prediction Using Artificial Neural Networks, ANFIS, and LSTM Network. Energies, 15(17), Articolo 17. https://doi.org/10.3390/en15176453

Panda, C., Mishra, A. K., Dash, A. K., & Nawab, H. (2023). Predicting and explaining severity of road accident using artificial intelligence techniques, SHAP and feature analysis. International Journal of Crashworthiness, 28(2), 186–201. https://doi.org/10.1080/13588265.2022.2074643

Pecoraro, F., Clemente, F., & Luzi, D. (2020). The efficiency in the ordinary hospital bed management in Italy: An in-depth analysis of intensive care unit in the areas affected by COVID-19 before the outbreak. PloS One, 15(9), e0239249. https://doi.org/10.1371/journal.pone.0239249

Philpott-Morgan, S., Thakrar, D. B., Symons, J., Ray, D., Ashrafian, H., & Darzi, A. (2021). Characterising the nationwide burden and predictors of unkept outpatient appointments in the National Health Service in England: A cohort study using a machine learning approach. PLoS Medicine, 18(10). https://doi.org/10.1371/journal.pmed.1003783

Proença, C. A. N., Neves, M. E. D., do Castelo Baptista Gouveia, M., & da Silva Madaleno, M. T. (2023). Technological, healthcare and consumer funds efficiency: Influence of COVID-19. Operational Research, 23(2). Scopus. https://doi.org/10.1007/s12351-023-00749-x

Psillaki, M., Apostolopoulos, N., Makris, I., Liargovas, P., Apostolopoulos, S., Dimitrakopoulos, P., & Sklias, G. (2023). Hospitals' Energy Efficiency in the Perspective of Saving Resources and Providing Quality Services through Technological Options: A Systematic Literature Review. Energies, 16(2), Articolo 2. https://doi.org/10.3390/en16020755

Rakshit, P., Zaballa, O., Pérez, A., Gómez-Inhiesto, E., Acaiturri-Ayesta, M. T., & Lozano, J. A. (2021). A machine learning approach to predict healthcare cost of breast cancer patients. Scientific Reports, 11(1), 12441. https://doi.org/10.1038/s41598-021-91580-x

Robinson, J. C., & Brown, T. T. (2014). Quantifying opportunities for hospital cost control: Medical device purchasing and patient discharge planning. The American Journal of Managed Care, 20(9), e418-424.

Rodriguez, S. D., Pascual, M. S., Oletto, A., Barnabas, S., Zuidewind, P., Dobbels, E., Danaviah, S., Behuhuma, O., Lain, M. G., Vaz, P., Luis, S. F., Nhampossa, T., Varela, E. L., Otwombe, K., Liberty, A., Violari, A., Maiga, A. I., Rossi, P., Giaquinto, C., … Tagarro, A. (2022). Machine learning outperformed logistic regression classification even with limit sample size: A model to predict pediatric HIV mortality and clinical progression to AIDS. PLoS ONE, 17(10 October). https://doi.org/10.1371/journal.pone.0276116

Rzychoń, M., Żogała, A., & Róg, L. (2021). SHAP-based interpretation of an XGBoost model in the prediction of grindability of coals and their blends. International Journal of Coal Preparation and Utilization. https://doi.org/10.1080/19392699.2021.1959324

Santamato, V., Esposito, D., Tricase, C., Faccilongo, N., Marengo, A., & Pange, J. (2023). Assessment of Public Health Performance in Relation to Hospital Energy Demand, Socio-Economic Efficiency and Quality of Services: An Italian Case Study. In O. Gervasi, B. Murgante, A. M. A. C. Rocha, C. Garau, F. Scorza, Y. Karaca, & C. M. Torre (A c. Di), Computational Science and Its Applications – ICCSA 2023 Workshops (pp. 505–522). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37111-0_35

Shay, L. A., & Lafata, J. E. (2015). Where is the evidence? A systematic review of shared decision making and patient outcomes. Medical Decision Making: An International Journal of the Society for Medical Decision Making, 35(1), 114–131. https://doi.org/10.1177/0272989X14551638

Singh, S., Srivastava, A., Rizwan, M., & Kumar, A. (2022). An Adaptive Intelligent Approach to Implement Energy Efficiency Concept in Built Infrastructure. 333–337. https://doi.org/10.1109/ICICICT54557.2022.9917766

Taghipour, F., Hamid, M., Aghakarimi, E., & Rabbani, M. (2023). An integrated framework to evaluate and improve the performance of emergency departments during the COVID-19 pandemic: A mathematical programing approach. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 237(6), 683–705. https://doi.org/10.1177/09544119231170303

Torra, V. (2023). A systematic construction of non-i.i.d. data sets from a single data set: Non-identically distributed data. Knowledge and Information Systems, 65(3), 991–1003. https://doi.org/10.1007/s10115-022-01785-3

Tosun, O. (2012). Using data envelopment analysis-neural network model to evaluate hospital efficiency. International Journal of Productivity and Quality Management, 9(2), 245–257. https://doi.org/10.1504/IJPQM.2012.045194

Van De Leur, R. R., Bos, M. N., Taha, K., Sammani, A., Yeung, M. W., Van Duijvenboden, S., Lambiase, P. D., Hassink, R. J., Van Der Harst, P., Doevendans, P. A., Gupta, D. K., & Van Es, R. (2022). Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders. European Heart Journal - Digital Health, 3(3), 390–404. https://doi.org/10.1093/ehjdh/ztac038

Vassilaki, N., Gargalionis, A. N., Bletsa, A., Papamichalopoulos, N., Kontou, E., Gkika, M., Patas, K., Theodoridis, D., Manolis, I., Ioannidis, A., Milona, R. S., Tsirogianni, A., Angelakis, E., & Chatzipanagiotou, S. (2021). Impact of age and sex on antibody response following the second dose of covid-19 bnt162b2 mrna vaccine in greek healthcare workers. Microorganisms, 9(8). https://doi.org/10.3390/microorganisms9081725

Wang, S. (2022). Application of Improved SOM Neural Network in Intelligent Auditing of Hospital Financial Vouchers. Proceedings of 2022 6th Asian Conference on Artificial Intelligence Technology, ACAIT 2022. https://doi.org/10.1109/ACAIT56212.2022.10137867

Wolf, J. A., Niederhauser, V., Marshburn, D., & LaVela, S. L. (2021). Reexamining “Defining Patient Experience”: The human experience in healthcare. Patient Experience Journal, 8(1), 16–29. Scopus. https://doi.org/10.35680/2372-0247.1594

Xu, G.-C., Zheng, J., Zhou, Z.-J., Zhou, C.-K., & Zhao, Y. (2015). Comparative study of three commonly used methods for hospital efficiency analysis in Beijing tertiary public hospitals, China. Chinese Medical Journal, 128(23), 3185–3190. Scopus. https://doi.org/10.4103/0366-6999.170279

Zaparoli Cunha, B., Droz, C., Zine, A.-M., Foulard, S., & Ichchou, M. (2023). A review of machine learning methods applied to structural dynamics and vibroacoustic. Mechanical Systems and Signal Processing, 200. Scopus. https://doi.org/10.1016/j.ymssp.2023.110535

Zhang, G., Shi, Y., Yin, P., Liu, F., Fang, Y., Li, X., Zhang, Q., & Zhang, Z. (2022). A machine learning model based on ultrasound image features to assess the risk of sentinel lymph node metastasis in breast cancer patients: Applications of scikit-learn and SHAP. Frontiers in Oncology, 12. https://doi.org/10.3389/fonc.2022.944569

Zhao, Y., Wang, S., & Chen, N. (2022). Thermal fault diagnosis of marine diesel engine based on LSTM neural network algorithm. 41, 198–203. https://doi.org/10.21595/vp.2022.22515