## RESEARCH PROPOSAL

In this section, you must present your proposal by developing the following aspects: a) Theoretical-conceptual foundations and state of the art that support the proposal, b) Hypothesis or research questions and objectives, c) Scientific or technological novelty of your proposal, d) Methodology and e) Work plan or Gantt chart.

Remember:

- Strictly comply with the terms and conditions of the competition.
- **Applicants must not include in this file information that reveals their identity, nor information that reveals the sponsoring institution of the proposal, such as name(s), last name(s) or initials. This applies indistinctly to the content of the file or its name.**
- To report bibliographic references belonging to the applicant of the proposal, the citation must be neutral, using the third person or a neutral wording when referring to these works.
- If applying to the Interdisciplinary and Transdisciplinary Evaluation Group, we strongly recommend reviewing the definition of the Evaluation Group provided within the Online Application System while selecting the primary field. Please note that this group will evaluate interdisciplinary and transdisciplinary relevance and the justification that supports your selection. For further information please refer to the Fondecyt Project Evaluation Guide, Chapter 5: "Guidelines for Interdisciplinary and Transdisciplinary Evaluation".
- Aspects related to the proposed research included in annexes will not be considered in the evaluation process.
- Maximum length of this section: 10 pages (use letter size format, Verdana font size 10 or similar).

# 1 Proposal description

Healthcare systems are constantly exposed to disruptive events such as natural disasters (earthquakes, hurricanes, floods, and wildfires), human-made incidents (chemical releases, wars, or nuclear attacks), societal shifts, and the emergence of diseases like the recent SARS-CoV-2 pandemic [1, 2]. These events can significantly impair their functionality [3] by disrupting supply chains and infrastructure, compromising the distribution of medical supplies and equipment [4], complicating human resource activities, including the onset of adverse psychological conditions [5], increasing patient waiting times, and reducing the quality of care [6]. Additionally, they intensify the risk of disease transmission and the emergence of comorbidities [7], decreasing access to and opportunities for healthcare services, escalating demand and strain resource allocation [8], and challenging effective communication and coordination [9]. In response, governments worldwide are implementing different strategies to mitigate these effects, such as increasing funding for emergency preparedness, establishing pertinent policies and regulations, enhancing the distribution of healthcare resources, and improving the collaboration and coordination between healthcare systems and other sectors [10].

The **efficiency evaluation** has been crucial to assessing the success rate of government measures addressing disruptive events [11–13]. It identifies critical areas that need improvement regarding preparation, response, and recovery, helping to prioritise and allocate resources effectively and evaluate the quality of care [14–16]. Efficiency evaluation is an interdisciplinary problem that requires synthesising and integrating links between disciplines, such as **health economics**, healthcare administration, medicine, and **computer science**. Each offers different viewpoints, expertise, techniques, methodologies and abilities to evaluate healthcare system efficiency, enhance the overall understanding, and provide information for decision-making [17].

The healthcare **technical efficiency** (TE) quantifies how institutions, hospitals, centres, and health systems use inputs (infrastructure, weighted service mix, time or operational expenses) to generate outputs (outpatient visits, discharges or surgeries), utilising the institution with the highest score to evaluate the remaining ones [18]. Different authors have recommended weighting healthcare production using the Diagnosis-Related Groups (DRG) to avoid bias when institutions with different types of patients are contrasted (**case mix**) [11, 13, 18–20]. The DRG is a patient classification system that standardises prospective payment by grouping patients with similar features, such as costs, sex, age, procedures, diagnosis, and comorbidities [21]. However, in developing countries, the DRGs system has **null or partial implementation considering heterogeneous criteria, times, coding, coverage, and registration quality**. All these factors limit the evaluation of TE [22].

Technical Efficiency (TE) is traditionally measured using optimisation methods like Data Envelopment Analysis (DEA) or Stochastic Frontier Analysis (SFA), which are based on linear mathematical programming [23–25]. However, these methods face limitations dividing the main problem into *NP-hard* sub-problems, requiring substantial time and resources to handle large datasets across institutions with multiple types of data (clinical, financial, quality, healthcare resource) [26, 27]. Additionally, frequent updates, sensitivity to outliers, and constraints on predicting beyond the initial data highlight the need for more dynamic and robust analytical approaches [28].

The technological advances of the digital era have converged to the integration between healthcare and technology, resulting in new applications for data management, processing, and storage [29]. Every day, health information systems increase the volume of available data: medical records, prescriptions, clinical reports, data associated with drugs, insurance, financial, and laboratory [30]. Hence, a growing number of features can be considered daily in TE analysis [31].

**Machine learning** (ML), a key area of Artificial Intelligence, has contributed to addressing healthcare administration (HA) problems which involve large volumes of data, including imbalanced, noisy, heterogeneous, incomplete and inconsistent data [30, 32]. The applications cover clinical decision-making problems, privacy and fraud detection, mental health management, healthcare demographics analysis, patient care and managing surgeries, pharmacovigilance, and **technical efficiency analysis.** [29, 33, 34]. Advanced ML algorithms like metaheuristics, deep learning, and reinforcement learning have recently been applied to improve predictive accuracy and operational efficiency. Integrating ML with DEA researchers has improved variable selection and efficiency classifications, allowing DEA models to manage complex and dynamic datasets [28, 35, 36]. **However, these advances are just beginning to be explored in the healthcare context** [31].

This project evaluates the efficiency of Chile's public healthcare system using an interdisciplinary approach, combining expertise from healthcare administration, economics, computer science, and public policy. We aim to understand the impact of disruptive events **like the COVID-19 pandemic** on healthcare efficiency. Novel ML strategies will be proposed to assess technical efficiency (TE), overcoming the limitations of traditional methods. We will also compare hospital performance before and during the pandemic by proposing an alternative to using DRGs. We aim to provide insights into Chile's healthcare system, identify improvement opportunities, and recommend strategies to enhance resource management and the system's resilience to disruptive events. This approach could potentially be applied to healthcare TE analysis in other countries

The following section provides an overview of the main concepts related to evaluating TE in healthcare, focusing on the work performed in Chile. We also present the project's hypothesis, goals, methodology, work plan, and expected contribution.

## 1.1 Background

Although it is expected that increasing healthcare resources should imply an improvement in healthcare services, the World Health Organization (WHO) has found that in developing countries, **up to 40% of resources are not utilised** [14]. This can be attributed to problems with human resource management, inadequate use of economic resources, medications, drugs, and medical devices, overestimating hospital stays, surgical re-interventions, and clinical practice errors. Furthermore, corruption and fraud can also lead to additional events [14, 37]. The current annual loss of health resources due to hospital-related inefficiency is estimated worldwide at USD 300 billion [38].

In healthcare, **efficiency** measures how effectively resources like supplies and expenses are used to reach the system's goals. Inefficiency can affect healthcare access, timeliness, and quality, leading to issues such as (1) limited access to necessary medications due to economic constraints, (2) the choice of less effective treatments because of cost, (3) continuous resource reallocation impacting other sectors like education and other public services, (4) increased dependence on private healthcare, (5) reduced public willingness to fund health services, which undermine performance and social welfare [39].

In developing countries, **public hospitals** consume most of the public healthcare resources [37], **covering more than 50% of the total annual expenditure**. According to [22], public hospital spending in Chile has increased over time, while health production has remained relatively constant in the last decade. A preliminary diagnosis suggests that health budgets do not reflect operational expenses, resulting in public debt. However, how much of this is due to economic underestimation and changes in the case mix, and how much is due to inefficiencies in health systems and institutions?

To address hospital efficiency, clinicians, managers, and computer scientists globally have concentrated on using the concept of technical efficiency to maximise output from given inputs and technology [19]. This study is crucial as healthcare systems must adapt to disruptive events like epidemiological shifts, social changes, technological advances, and natural disasters. The COVID-19 pandemic, for instance, has highlighted these challenges by significantly impacting global health service efficiency, with features such as hospital infrastructure, technology, staff training, and location being analysed in developed countries through regression models [40–43]. **However, such studies remain limited in developing countries.**

### 1.1.1 Technical efficiency in healthcare administration

Efficiency measures the relationship between an institution's inputs and its outputs. Institutions at the Production- or Pareto-frontier are considered highly efficient, setting benchmarks for others. Efficiency is categorised into allocation and technical [39]. The former assesses how resources are combined to produce outputs. TE determines if an institution maximises output with minimal resources (input-oriented) or minimises resource use for a given output (output-oriented), commonly having a scores range from 0 (highly inefficient) to 1 (maximum efficiency or Pareto solution) [44]. For instance, a hospital with a TE score of 0.8 produces 20% less than a Pareto-frontier institution using identical resources [18].

According to [12], TE analysis has primarily been applied in sectors like finance, agriculture, transportation, and education, with a particular focus on healthcare due to its significant resource use and social impact. Specifically, much of the research has concentrated on hospital efficiency [19], although studies have also covered primary healthcare, thoroughly reviewed in [45]. The review realised by [19] analysed over 250 articles published from 2005 to 2016, finding that 37% focused on European hospitals, 25% on Asian, 24% on North American, 8% on African, and **only 6% on hospitals from Oceania, Central, and South America**. The majority employed Data Envelopment Analysis (DEA) for TE evaluation, with 44% using BCC (Banker, Charnes, and Cooper) models with variable returns to scale (VRS) and 34% applying CCR (Charnes, Cooper, and Rhodes) models with constant returns to scale (CRS) [46]. Other studies have utilised different algorithms like additive models (ADD), slack-based measures (SBM), Assurance regions, and Congestion methods [47]. Historically, input orientation dominated, but there has been a shift towards output-oriented models since 2015 [19]. Alternative methods like Stochastic Frontier Analysis (SFA) and Malmquist-DEA for different time periods have also been explored [37, 48].

Using DEA or SFA, employing many inputs, outputs, and institutions can increase the chances of achieving Pareto solutions but may reduce the discrimination power of these methods. Thus, the literature often recommends using parsimonious models, adhering to the guideline that the number of hospitals should not exceed twice the sum of the inputs and outputs [49]. However, the growing volume of data and variables presents a significant challenge in assessing TE.

The authors in [13] recommend using standard DEA-based models with inputs like beds, service combinations, work units, and operating expenses, adjusting for case mix in discharges and ambulatory care. A systematic review by [19] showed variations in the input data used across studies: 20% included all staff, 18% only medical personnel, 24% specifically nurses or other clinicians, and 7% non-clinical personnel. Additionally, 6% factored in medical devices and infrastructure, while 5% considered financial aspects. For outputs, 44% of the studies used case-mix-adjusted discharges, 22% surgeries and procedures, 14% admissions, 9% quality indices, 5% revenues, and 6% other functionalities.

Recent trends show an increasing use of regression analyses to identify **determinants** of TE [19]. If hospital discharges are considered outputs, a hospital where many patients quickly pass away would be defined as well-efficient according to

TE models. Therefore, evaluating efficiency in terms of exogenous quality variables (determinants) associated with clinical production, care quality, mortality rates, and other relevant factors is essential. This approach ensures a more comprehensive hospital performance assessment beyond mere output efficiency.

Classical DEA strategies focus on efficiency objectives related to short-term projections along the Pareto-frontier, which can lead to over-performance by inefficient institutions [26, 50]. These models solve efficiency issues by decomposing them into combinatorial ***NP-hard* subproblems**. Classical strategies are outlined in [20, 44], including:

- **Data envelopment analysis** (DEA). It is a non-parametric method used in operations research and economics to measure the efficiency of different **decision-making units** (DMUs). It employs linear programming to construct a production frontier by fitting a piece-wise linear surface over observed data. Typically, DEA utilises data related to revenues, expenses, and production; however, it can also incorporate financial data to assess allocated efficiency. One of the main advantages of DEA is its ability to consider multiple resources and outputs without needing to assume a specific functional form for the efficiency frontier. However, DEA's susceptibility to noise in the data can be a disadvantage, often requiring at least bootstrap methods when dealing with large datasets.

  The calculation for each DMU involves solving the following linear programming model, where $X$ s is an input matrix of dimensions $(k \times n)$, $Y$ is a matrix with size $(m \times n)$, and $\theta \in [0, \infty]$. $\lambda$ is a vector with $(n \times 1)$ that assigns weights to calculate the position of each DMU. The variable $\theta$ is a scalar that scales the inputs. The model adheres to the constant return to scale (CRS) when the constraint $N_1\lambda \leq 1$ is omitted, implying proportional output changes relative to inputs. In contrast, including this constraint allows for a variable return to scale (VRS) model, indicating non-proportional output changes relative to inputs. The equations for the model are:

$$\begin{gathered} \theta, \lambda\theta \\ s.t - y_i + Y\lambda \geq 0, \\ \theta x_i - X\lambda \geq 0, \\ N_1\lambda \leq 1, \\ \lambda \geq 0 \end{gathered}$$

  Despite widespread use, DEA is sensitive to outliers and erroneous data and requires continuous updates with new DMUs. It also assumes a fixed production function, limiting its adaptability to real-world problems.

- **Approximation based on Stochastic Frontiers** (SFA). It is an econometric method that estimates production or cost frontiers, including distance functions like DEA. It utilises data related to economic inputs, expenses, production, input prices, and outputs, which are essential for estimating long-term frontiers. Short-run cost frontiers require data on variable costs, variable input prices, fixed input quantities, and outputs.

  One of the main advantages of SFA is its robustness against noise, which allows for incorporating environmental variables and conducting traditional hypothesis tests. This robustness also facilitates the identification of outliers. SFA can handle multiple inputs and outputs. However, due to the specified noise distribution and asymmetry, SFA can add a potential bias in results, and it generally requires large samples to produce robust estimates.

  The calculation of the frontier in SFA is defined by the equation $y = f(x) + v - u$, where $y$ represents the output, $v$ is the error term that captures random shocks, and $u$ represents inefficiency, capturing the shortfall from the frontier. This model structure allows SFA to separate the influence of inefficiency from random noise, providing a clear view of the underlying production or cost efficiency.

- **Price-Based Index Numbers** (PINs). A traditional index number method, employing prices as weights, measures the production frontier. This approach requires data on economic inputs and outputs across multiple units or periods. PINs are advantageous for their ability to handle multiple observations and provide reproducible, transparent results, facilitating the calculation of allocation efficiency. However, the method requires specific price information and cannot divide the Pareto frontier into sub-problems. Additionally, transformations are needed to maintain the transitivity of multilateral comparisons [44].

- **Malmquist-DEA**. It is a method used to assess changes in productivity across two different periods. It breaks down the analysis into two components: one that evaluates shifts in the technological frontier and another that examines TE changes. This approach requires data on different years of production units, inputs, outputs, and resources. One of its main advantages is simultaneously evaluating DMUs across two periods by analysing changes in production frontiers and accommodating multiple inputs and outputs. However, Malmquist-DEA is very sensitive to data noise, which can affect its accuracy. For a detailed review of this method, see [51].

#### 1.1.2 DEA and Machine learning hybridisation

Integrating Data envelopment analysis with Machine learning approaches (DEA-ML) represents a significant advancement in assessing TE and decision-making by improving both methodologies' strengths. It has proved to enhance predictive accuracy and operational insights across different applications. For example, authors in [52] explored using DEA-ML to analyse the success factors of Initial Coin Offerings (ICOs) in cryptocurrency, demonstrating how DEA can identify efficient ICOs while ML pinpoints key success factors. Similarly, authors in [28] applied DEA-ML algorithms to evaluate the efficiency of Chinese manufacturing firms, showing that the integration significantly reduces the computational load and enhances predictive reliability. Additionally, different applications have been performed using DEA-ML, emphasising its adaptability in the public and private sectors for several efficiency assessments [53]. Authors in [34] and [54] have focused on proving the TE predictive capabilities of DEA-ML over ICOs and industrial applications, respectively.

These studies highlight specific ML algorithms like **metaheuristics**, **artificial neural networks**, **random forest**, and **support vector machines**, which, when combined with DEA, offer robust models able of handling complex datasets and variable inputs. Further studies have elaborated methodological enhancements, incorporating ML algorithms to refine the efficiency scores derived from DEA, thus providing a deeper understanding of DMUs performance [35]. Despite these advances in integrating DEA-ML across sectors like finance, manufacturing, and public administration, their potential for enhancing efficiency in broader areas such as healthcare remains largely unexplored [31].

#### 1.1.3 Technical efficiency determinants

Identifying key exogenous features related to TE is essential for enhancing the quality of healthcare service. To this end, different quantitative methods such as regression analysis, statistical models, and machine learning techniques, including Random Forest, Gradient Boosting, and Artificial Neural Networks, have been employed to analyse extensive data from clinical, financial, quality, and healthcare resource sectors [20, 38, 55–58]. These approaches help pinpoint crucial relationships between hospital efficiency and factors such as staffing levels, staff training, case mix, hospital size and type, availability of resources and technology, and financial and quality management practices [20].

Understanding these determinant features enables healthcare managers to develop strategies to enhance institutional performance, optimise resource allocation, and improve care quality. This data-driven approach supports strategic investment and resource utilisation to boost TE.

#### 1.1.4 Technical efficiency in Chilean public hospitals

Since 2004, different studies have assessed TE in Chile's healthcare at both primary and tertiary-hospital levels [59–62] (See Table 1), finding that around 20% to 40% of resources are underutilised. Correlational research by [63], using data from [55], identified critical TE determinants such as human resources, resource management and cost savings, administrative management, and operational processes. The Fondo Nacional de Salud (FONASA) recently launched a web tool featuring expenditure and production data adjusted by DRGs for about 60 high-complexity hospitals[1]. However, although it offers descriptive statistics on DRGs and financial data, among other features, it lacks a methodology for evaluating TE [22].

The study by [64], which adopted the methodology proposed by [13], was constrained by the limited implementation and coverage of DRGs in Chile, analysing 28 high-complexity hospitals. If this methodology were reapplied today, fewer than 80 of over 190 hospitals in the public health network could be included due to the absence of DRGs in the rest, as noted by [19]. Despite these 80 hospitals handling nearly 80% of the annual hospital expenditures, this **study would not adequately represent the national hospital landscape concerning case mix.**

Authors such as [65, 66] have suggested using the complexity classification proposed by the Ministry of Health of Chile (MINSAL) to make fair efficiency comparisons. However, the few studies that have evaluated this relationship have obtained inconclusive results [67]. Other works, like the one developed by [18], have proposed models based on machine learning to group hospitals according to their production instead DRGs, including all public hospitals. However, it has not been studied whether such models respond to the methodology proposed by [13], representing the concept of case mix, much less how TE has been affected by disruptive events, such as the current pandemic scenario [68–74].

#### 1.1.5 Key background insights

The key points concerning the background can be summarised as follows:

1. **Resource utilisation and context**. In developing countries, up to 40% of healthcare resources are not utilised effectively. Public hospitals, consuming over 50% of the total annual healthcare expenditure, urgently need TE

[1] https://www.fonasa.cl/sites/fonasa/datos-abiertos

studies, especially when dealing with disruptive events like the COVID-19 pandemic. However, research in these regions is still limited.

2. **DEA application and limitations**. TE measures whether an institution maximises output with minimal resources (input-oriented) or minimises resource use for a given output (output-oriented). DEA and SFA are commonly used to assess TE, but they have problems dealing with large data volumes, noise, and extreme values. Additionally, dynamic analyses which integrate new DMUs are challenging. Hospital outputs **must be adjusted using a case mix** to ensure fair comparisons.

3. **DEA-ML integration**. DEA-ML integration improves data handling and analysis, enhancing TE assessments by filtering noise and effectively managing large datasets considering different application contexts. DEA-ML dynamically adapts to variable changes, enabling precise comparisons across different settings and new DMUs. Despite these advantages, applying these strategies in healthcare is null or lacking according to the state-of-the-art.

4. **Determinant analysis.** Identifying determinant features impacting TE in healthcare is critical for enhancing service quality. Quantitative methods such as **regression** and machine learning-based **feature selection** allow data analysis to identify efficiency-feature relationships.

5. **Chilean healthcare context.** Different studies have assessed TE in Chile's healthcare system (see Table 1), revealing that approximately 20-40% of resources are underutilised. The main determinants of TE include human resources, resource management, cost savings, administrative management, and operational processes. However, limited data availability and challenges in implementing and covering DRGs restrict a comprehensive analysis of TE in Chilean public hospitals. Some authors have suggested alternative strategies for TE comparison, yet results are inconclusive, and the influence of disruptive events on TE remains understudied.

Table 1: Technical efficiency evaluation in Chilean public hospitals.

| Reference | Income | Outcome | Method | # DMU | Determinants |
|---|---|---|---|---|---|
| [59, 68] | Spending, #beds | Average length of stay medical and other clinical appointments | DEA CRTS-O, VRST-O Scale-O | Hospitals divided in types 1, 2, 3, 4 and 5 | Tobit model |
| [66] | Medical, nurse and midwife staff and #available beds | Discharges | DEA-CCR-I | 190 hospitals grouped by geographical region | — |
| [64] | Medical and administrative staff, #beds, spending on goods and services | Discharges adjusted by IR-GRD | DEA, CRTS-O, VRST-O and Scale-O | 28 hospitals with high complexity | Tobit model published in [55] |
| [18] | Spending on goods and services | Discharges, occupied beds and medical appointments | 2-stages: clustering + CRTS-O, VRTS-O | 193 hospitals + clustering based on production | Correlation analysis |

## 2 Hypothesis

The emergence of disruptive events like the current pandemic underscores the critical importance of healthcare TE globally. However, the impact of such events on TE in developing countries like Chile remains unstudied. This project aims to analyse how the pandemic has affected TE in Chilean public hospitals, dealing with DEA and its limitations. Key areas of analysis include identifying affected institutions, evolving case mix, efficiency determinants, and resource utilisation to give recommendations for future disruptive events.

The research hypothesis of this project proposes that **”an interdisciplinary approach that integrates computer science, healthcare administration, and expertise from different domains can provide a comprehensive strategy for assessing the technical efficiency of Chilean public hospitals during disruptive events using DEA-ML hybrid algorithms to propose robust and dynamic models that allow the study of the pre-and pandemic periods, identifying key determinant features for resource management, and enhancing the healthcare system’s response to future disruptive events”.** To further investigate this hypothesis, different research questions are proposed:

1. How does the performance of the healthcare system differ when utilising DEA-ML compared to DEA, and what are the implications for decision-making, TE and resource allocation?
2. How has the case mix of public hospitals in Chile evolved considering the pre- and pandemic periods, and how does it influence technical efficiency?
3. How has Chilean public hospitals’ technical efficiency (TE) been affected by the pre- and pandemic periods, and what factors contributed to any observed changes?
4. Are healthcare production data related to TE determinants (or inefficiency), and how do they compare with other hospital performance indicators?
5. Which hospitals were the most and least efficient during the pandemic, and what characteristics determine their performance?
6. What other TE-related clinical and administrative processes have been affected by the pandemic, and how do these changes impact overall hospital efficiency?

By answering these questions, this project seeks to contribute to developing a more resilient and efficient healthcare system in Chile that is better prepared to face future disruptive events.

# 3 Goals

**General goal.** This project aims to develop an innovative interdisciplinary approach by integrating computer science and healthcare administration, introducing a hybrid DEA-Machine Learning model to enhance classical TE assessments’ predictive accuracy and operational efficiency considering Chilean public hospitals. This approach is designed to address the computational limitations of traditional algorithms, identify key factors impacting TE, and offer strategic recommendations for effective resource allocation and management. Additionally, it aims to develop robust strategies that enhance the responsiveness of the healthcare system to future disruptive events.

**Specific goals.** To successfully reach the general goal of this project, it is separated into the next specific goals:

1. Develop a methodology to group and characterize Chilean public hospitals by case mix without using DRGs, aiming to reduce bias in TE assessments and enhance hospital comparison accuracy.
2. Propose a novel hybrid DEA-ML approach to address the limitations of traditional TE assessment methods following the standards defined by [13].
3. Use descriptive, predictive, and prescriptive analytics to identify and analyse determinants of TE, applying ML techniques to uncover factors significantly impacting hospital efficiency: determinant features.
4. Conduct a comprehensive interdisciplinary analysis combining computer science, healthcare expertise, and insights from biomedical, economics, and public policy to provide strategic recommendations for resource management and enhance the healthcare system’s adaptability to disruptive events.
5. Disseminate the findings and methodologies of this interdisciplinary project across computer science, operations research, and healthcare administration communities through technical reports, publications in recognised journals, and presentations at conferences.

In addition to these objectives, this project attempts to promote technology and knowledge transfer, train under and graduate students, and improve the research process through collaboration between researchers from different disciplines.

# 4 Scientific or technological novelty of your proposal

The proposed project adopts an interdisciplinary approach, integrating healthcare administration, health economics, medicine, and computer science to improve the efficiency assessment of healthcare systems during disruptive events such as natural disasters, social changes, and pandemics. This strategy aims to develop a comprehensive method for evaluating TE in Chilean public hospitals. It will power robust, hybrid strategies for exhaustive analysis, offer resource allocation and management recommendations, identify crucial determinants that influence TE, and devise strategies to improve the healthcare system's responsiveness to future disruptive events using data from the current pandemic. The novelty of the proposal can be summarised as:

1. **Integration of disciplines and regional innovation**: This project is a pioneering effort to integrate healthcare administration, health economics, medicine, and computer science to evaluate hospital TE during disruptive events in developing countries, especially in pandemic scenarios. Moreover, it will be the first comprehensive analysis of hospital efficiency encompassing all public hospitals in Chile, incorporating state-of-the-art recommendations associated with case-mix.

2. **Development of innovative strategies for case mix modelling**. The project aims to employ ML algorithms to predict hospital case mix (DRG) using recorded healthcare production variables. This approach will help expand the coverage of efficiency analyses and provide data-driven insights to support prospective payment decisions.

3. **Enhancement of TE evaluation using ML**. This project will use hybrid DEA-ML algorithms to enhance the evaluation of TE in healthcare. By addressing the limitations of traditional TE approaches and using ML to identify key TE determinants, this initiative will be one of the pioneers in applying these advanced techniques in the healthcare sector.

4. **Informatics tool development**. The project will create a freely accessible informatics tool that enables health managers to evaluate TE effectively using the data and algorithms developed. This tool will support ongoing TE assessment and strategic planning.

5. **Knowledge and technology transfer**. The project aims to improve research capacity across disciplines by promoting technology transfer and providing training to undergraduate and graduate students. All findings will be formalised and disseminated through scientific articles, conference presentations, and workshops, extending the developed strategies to other regions and potentially globally.

# 5 Methodology

ML has made significant advances within artificial intelligence over recent decades, merging data analytics with the scientific method to enhance its capabilities in generalisation, adaptability, robustness, and flexibility [75, 76]. The Cross-Industry Standard Process for Data Science (CRISP-DM) is the predominant methodology for data science and machine learning, involving six essential stages: problem definition, data understanding, data preparation, modelling, evaluation, and deployment [77]. Despite the inherent integration of statistics, mathematics, and computer science within data science and ML, supplementing these with domain-specific expertise is essential for comprehensive problem-solving. Hence, we propose a methodological fusion of CRISP-DS with Participatory Action Research (PAR). PAR methodology underscores active involvement by individuals with lived experiences or co-researchers in generating insights and strategies for change [78]. PAR contains four core activities: planning, action, reflection, and evaluation [79]. This integrated approach called **Cognitive Project Management for Artificial Intelligence** (CPMAI), intricately incorporates artificial intelligence into each phase through a comprehensive framework [80].

In implementing this methodology, a diverse team of specialised researchers from different fields will collaborate, including experts in optimisation and ML (R1 and R2), healthcare operational research (R3), Chilean public health policies (R4), Diagnosis Related Groups (DRGs) (R5), and biomedical/clinical engineering (R6). While the Principal Investigator (PI) possesses expertise across these disciplines, the project's success is motivated by collaborative teamwork. The proposed methodology covers six stages (Figure 1A), as elaborated below, which are specified below, contextualised to this project:

1. **Problem definition.** This stage has been partially covered with the development of this project. It involves working with the researchers: (1) conducting a literature review and discussing the determining knowledge gaps (R1-R6), (2) identifying technological deficiencies in the national public health system in the application of recommended methodologies for evaluating TE, considering the theoretical background and practice (R4-R5), and (3) studying alternative solutions used in the current pandemic and disruptive events in other countries (R1-R6). This stage is associated with specific goals 1 to 5 and includes the following sub-stages: (1) goal definition, (2) assessment, (3) refinement of objectives, and (4) working plan.

2. **Data understanding**. In this stage, data and indicators will be downloaded from the public health databases available on Department of Health Statistics and Information (DEIS): REM, SIS-Q, and discharge databases (DB), among others [81]. These data do not include private information. Previous works [67] have indicated that this data is stored in different structures, formats, and dictionaries: PDF, CSV, MDB, and XLS. Hence, we will create a repository with a normalised DB, including data associated with hospital expenditure and DRG weights. In addition, we will conduct participatory activities with the research team (R3-R6) to better understand the data and their alignment with the project. This will involve collaborative analysis and interpretation of the data and its implications for the study's goals (R1-R6). This stage is associated with specific goals 1 to 4 and includes the following sub-stages: (1) data collection, (2) data description, (3) data exploration, and (4) data quality verification.

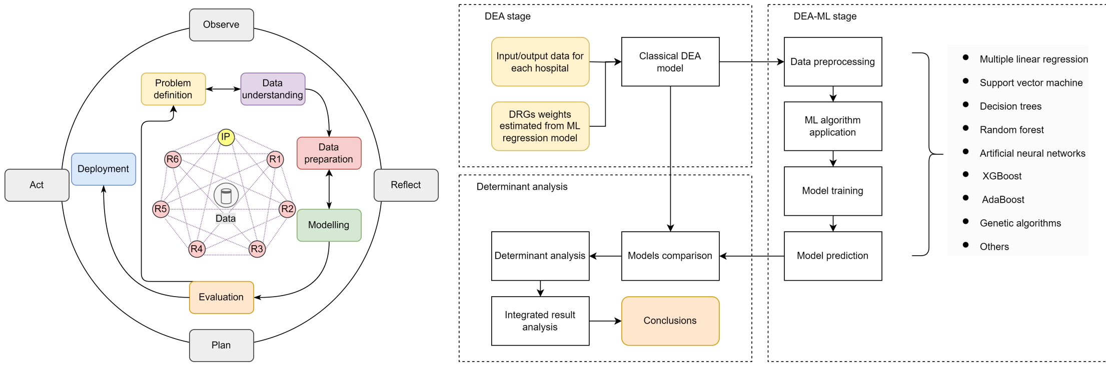
3. **Data preparation.** It involves the construction of the structured DB designed in the previous stage, applying recommendations from the literature, DB design techniques, and current governance regulations. This DB will be the core of the proposed strategy for evaluating TE. We will conduct participatory activities with the research team (R3-R6) to discuss and agree on selecting and prioritising candidates' features as resources, products, and determinants. The data for each selected feature will be reviewed for quality, removing features with incomplete data. Outliers and features with low variance should also be eliminated. Production and case mix features will group hospitals according to complexity, considering a homogeneous case mix using machine learning and optimisation strategies. This stage is associated with specific goals 1 to 4 and includes the following sub-stages: (1) selection, (2) data cleaning, (3) DB construction, (4) integration, and (5) format assignment.

4. **Modelling**. In this stage, we will implement the TE evaluation strategy, involving different disciplines in the modelling process (R1-R6). This includes (1) modelling the complexity of the combination of cases using production features and regression ML strategies as an alternative to the use of DRGs (R1-R6), (2) calculating TE using DEA-ML hybrid algorithms based on linear regression, support vector machine, decision trees, random forest, artificial neural networks, XGBoost, AdaBoost, among others (R1-R3), and (3) identifying determinants of efficiency using regression models (R4-R6) (This step is shown in Figure 1B). We will work as a team to determine the modelling strategy, test design, model construction, prototype construction, and validation (R1-R6). This stage is associated with specific goals 1 to 4. It includes the following sub-stages: (1) selection of modelling strategy, (2) test design, (3) model construction, (4) prototype construction, and (5) validation by the interdisciplinary group.

5. **Evaluation**. In this stage, the results obtained by ML and hybrid models will be evaluated, considering both the computational science and domain knowledge perspectives (R1-R6). Participatory activities with the team will be conducted to ensure that the results align with the needs and expectations defined in the project. The models and results will be formalised in technical reports and publications. The findings will be compared with previous literature on developed countries. This stage is associated with specific goals 1 to 4. It includes the following sub-stages: (1) results evaluation, (2) participatory review, and (3) discussion and determination of activities to be re-evaluated or repeated.

6. **Deployment**. This stage will synthesise findings from previous stages, consolidating results through participatory activities involving researchers (R1-R6) and utilising informatics tools integrated with the project's databases and outcomes. The hosting and administration of these tools will be managed by our institution throughout the project's duration, with the potential for transfer as a public good under the open science paradigm once the project is completed. Findings and conclusions will be disseminated through presentations, conference articles, journal publications, meetings, and workshops. This stage corresponds to specific goal 5 and encompasses the following sub-stages: (1) strategy implementation, (2) monitoring and maintenance, (3) reporting and dissemination of results, and (4) project review and consolidation.

   In parallel, we will use **SCRUM** [82] by considering ten development cycles (Sprints) to develop the informatics tool. Specifically, these are: *Sprint 0*. Initial planning and technological decision-making (Backlog); *Sprint 1*. Design and proposal of tool architecture; *Sprint 2*. Design and implementation of the database; *Sprint 3*. Development of the data management and maintenance module; *Sprint 4*. Implementation of a model for TE evaluation; *Sprint 5*. Implementation of a model for determinant analysis; *Sprint 6*. Generation of a reporting system; *Sprint 7*. Creating a module for results and Key Performance Indicators visualisation; *Sprint 8*. Adjustment of the graphical interface; *Sprint 9*. Integration module; *Sprint 10*. General validation of the web tool and deployment.




(a) CPMAI methodology. PI: principal investigator and R1-R6: research team.

(b) Research framework of the linkage between DEA method and ML algorithms associated with Stage 5, Methodology.

Figure 1: Schemes associated with CMPAI methodology and its 5 stage

.

## 5.1 Evaluation of the proposal

### 5.1.1 Datasets

To execute our project, we will collect data from all 190 public hospitals across Chile. The DEIS website will serve as our primary source, giving us access to a comprehensive dataset containing healthcare products, case mix, and efficiency determinants. This dataset includes over 10,000 features annually recorded between 2014 and 2023. Among these features are details on discharged patients, appointments with medical and clinical professionals, emergency and dental care, exam production, surgeries, clinical procedures, and healthcare quality indicators [18]. Financial data can be found on the FONASA website, while average DRG weights can be obtained through the Portal de Transparencia.

Studying causality considering different disruptive phenomena will be challenging, and it is essential to maintain the integrity of this aspect of the project. Given the characteristics of the current features, separating the effects associated with each disruptive event, such as social, epidemiological or geographical factors, is particularly difficult. Nonetheless, our focus will be on DEIS records, known as REM F, which include housing data related to the pandemic. These records could provide invaluable insights into the relationship between the pandemic and hospital TE.

### 5.1.2 Evaluation metrics

To study TE (Stage 5, Methodology), a range of data-driven quality metrics can be employed to evaluate:

- The regression models that study the healthcare features associated with the case mix (average DRG weight), residual distributions and statistical hypothesis tests can be used to select the most suitable models. The same approach could be used to evaluate the Tobit regression models, identifying TE determinants.
- The resultant clusters that compare hospitals with similar case mixes will be evaluated using quality indexes over the features associated with healthcare products, such as the Ball-Hall, Dunn, Silhouette, or Xie-Beni indexes [18]. Comparing clusters from different periods or strategies could be performed using the Folkes-Mallows, Rand index, or Jaccard indexes [69].
- The DEA-ML hybrid algorithms used in TE's feature selection and evaluation will be evaluated using a structured approach that starts with applying classical DEA to calculate the relative efficiency of hospitals. Subsequently, machine learning models are trained with these efficiency scores as outputs and additional relevant data as inputs. These models will be rigorously statistically tested for predictive accuracy using $R^2$, RMSE, and MAE metrics, allowing comparison under different feature combinations and orientations to identify the most effective predictive models (See Figure 1B).
- Although all the proposed models can be evaluated using quantitative ML metrics, their practical significance must be interpreted by experts, especially when considering recommendations. In addition to the **Principal Investigator** (PI), different specialists will contribute to this task in each methodology stage (see Section 5).
- The informatic tool proposed can be evaluated using heuristics, such as usability, functionality, performance, reliability, and maintainability, among others [83].

- Determinants, efficiency and case mix evolution during the pandemic could be contrasted with different works realised in developed countries [40, 42, 84].

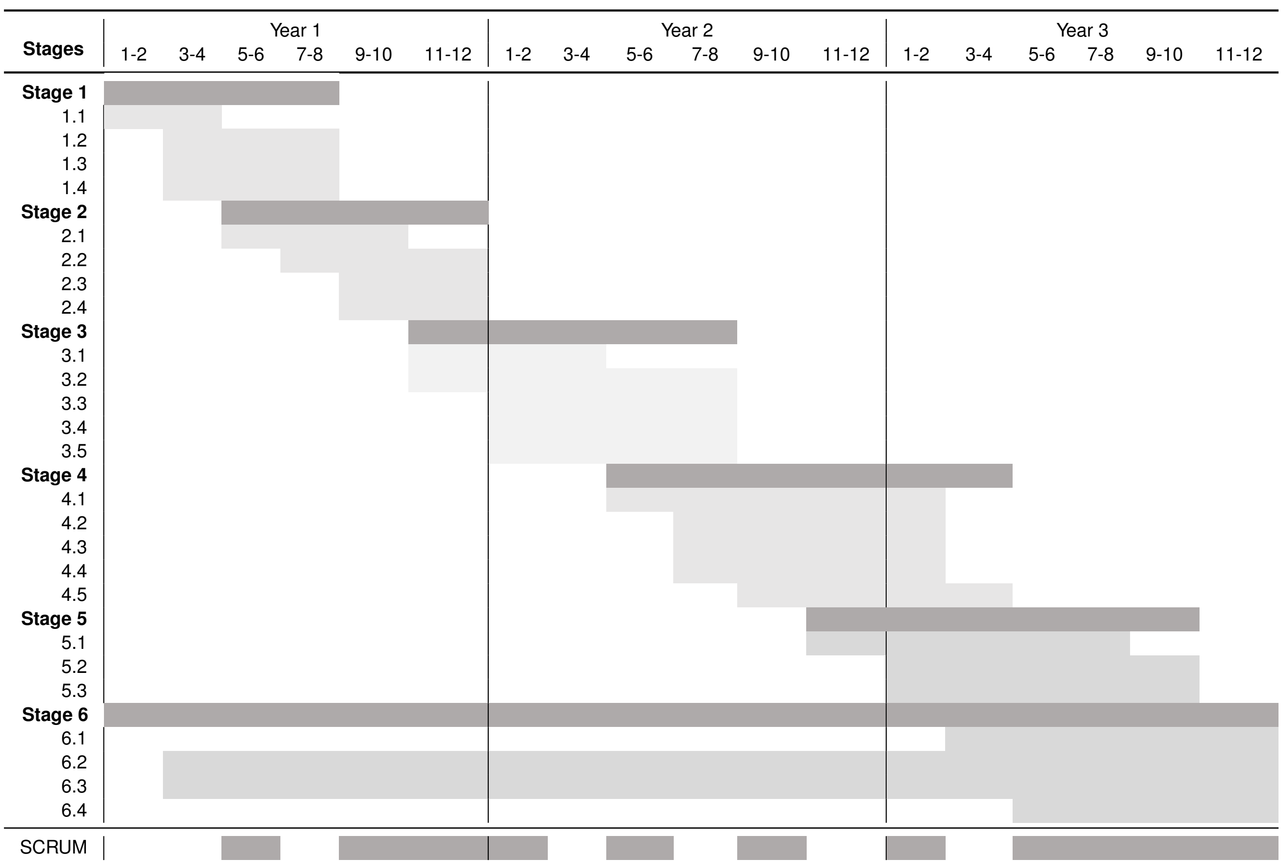
# 6 Work Plan

The project will be executed by the **PI** involving both **graduate** students (PhD and master) and **undergraduate** students. Additionally, the project will collaborate with **six researchers** at different stages (R1-R6, Section 5). A list of proposed research theses that will be integrated into the project can be found in the resources justification document.

Each stage described in Section 5 has been schematised in Figure 1 and scheduled in Table 2. Two international research visits are proposed during the project. They are identified in the last row of Table 2, indicating the month in which it is expected to happen. The cells $JP$, $N$ and $I$ indicate the expected date for submitting papers or proceedings.

**Target conferences** for the publications expected in this project are ICHSM (International Conference on Healthcare Service Management), ICMMGH ( International Conference on Modern Medicine and Global Health), GECCO (Genetic and Evolutionary Computation Conference) and CEC (IEEE Congress on Evolutionary Computation). Also, the publication of papers at the International Conference of the Chilean Computing Society is expected to disseminate the work among the national community. **Target journals** for the publications expected are the Journal of Biomedical Informatics, the Journal of Medical Systems, the IEEE Journal of Biomedical and Health Informatics, the Health Information Management Journal and Health Information Management Journal and Applied Soft Computing.

Table 2: Scheduled activities during the development of the project. Three types of activities are considered during the project: **RV** shows the expected dates for the research visits, **IC** shows the expected date for international collaboration and **CP** shows the expected date for conference presentations (N: National and I: International) during the project.




RV: Research visit; IC: International collaboration; JP: Journal paper and CP: Conference presentation (N: national; I: international)

| | Y1 1-2 | Y1 3-4 | Y1 5-6 | Y1 7-8 | Y1 9-10 | Y1 11-12 | Y2 1-2 | Y2 3-4 | Y2 5-6 | Y2 7-8 | Y2 9-10 | Y2 11-12 | Y3 1-2 | Y3 3-4 | Y3 5-6 | Y3 7-8 | Y3 9-10 | Y3 11-12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IC | | | | | IC1 | | | | | | IC2 | | | | | | IC3 | |
| RV | | | | | | | | | RV1 | | | | | RV2 | | | | |
| JP | | | | | | | | | | JP1 | | | | | | | JP2 | |
| CP | | | | | N | | | | I | | N | | | | | I | N | |

**BIBLIOGRAPHIC REFERENCES:**
In this section, include the list of cited references in the Proposed Research section. **Length: 5 pages**.
(Must use letter size, Verdana size 10 or similar).

# References


[1] H. Mahmoud, T. Kirsch, D. O'Neil, and S. Anderson, "The resilience of health care systems following major disruptive events: Current practice and a path forward," Reliability Engineering & System Safety, vol. 235, p. 109264, 2023.

[2] S. K. Mangla, A. Chauhan, T. Kundu, and A. Mardani, "Emergency order allocation of e-medical supplies due to the disruptive events of the healthcare crisis," Journal of Business Research, vol. 155, p. 113398, 2023.

[3] J.-S. Wu, "Measuring efficiency of the global fight against the covid-19 pandemic," DIGITAL HEALTH, vol. 9, 2023.

[4] Y. Shi and Z. He, "Decision analysis of disturbance management in the process of medical supplies transportation after natural disasters," International Journal of Environmental Research and Public Health, vol. 15, no. 8, p. 1651, 2018.

[5] M. Niazi, M. R. Dehkordi, M. Eghbali, and D. Samadian, "Seismic resilience index evaluation for healthcare facilities: A case study of hospital in tehran," International Journal of Disaster Risk Reduction, vol. 65, p. 102639, Nov. 2021. [Online]. Available: https://doi.org/10.1016/j.ijdrr.2021.102639

[6] R. M. Wood, "Supporting COVID-19 elective recovery through scalable wait list modelling: Specialty-level application to all hospitals in england," Health Care Management Science, vol. 25, no. 4, pp. 521–525, 2022.

[7] C. C. Hammer, J. Brainard, and P. R. Hunter, "Risk factors and risk factor cascades for communicable disease outbreaks in complex humanitarian emergencies: a qualitative systematic review," BMJ Global Health, vol. 3, no. 4, p. e000647, 2018.

[8] H. Leite, C. Lindsay, and M. Kumar, "COVID-19 outbreak: implications on healthcare operations," The TQM Journal, vol. 33, no. 1, pp. 247–256, 2020.

[9] N. Mohammadzadeh, S. Saeedi, and S. Rezayi, "Telemedicine and natural disasters: various services, requirements, challenges, and general framework," Frontiers in Emergency Medicine, 2022.

[10] M. Mavrouli, S. Mavroulis, E. Lekkas, and A. Tsakris, "The impact of earthquakes on public health: A narrative review of infectious diseases in the post-disaster period aiming to disaster risk reduction," Microorganisms, vol. 11, no. 2, p. 419, 2023. [Online]. Available: https://doi.org/10.3390/microorganisms11020419

[11] H. Shirouyehzad, J. Jouzdani, and M. Khodadadi Karimvand, "Fight against covid-19: a global efficiency evaluation based on contagion control and medical treatment," Journal of Applied Research on Industrial Engineering, vol. 7, no. 2, 2020.

[12] A. Emrouznejad and G. liang Yang, "A survey and analysis of the first 40 years of scholarly literature in dea: 1978–2016," Socio-Economic Planning Sciences, vol. 61, pp. 4–8, 2018, recent developments on the use of DEA in the public sector.

[13] Y. A. Ozcan, Health Care Benchmarking and Performance Evaluation. Springer US, 2008.

[14] H. Nassar, H. Sakr, A. Ezzat, and P. Fikry, "Technical efficiency of health-care systems in selected middle-income countries: an empirical investigation," Review of Economics and Political Science, vol. 5, no. 4, pp. 267–287, 2020.

[15] I. Vrabková and S. Lee, "Approximating the influence of external factors on the technical efficiency score of hospital care: evidence from the federal states of germany," Health Economics Review, vol. 13, no. 1, 2023.

[16] J. Piedra-Peña and D. Prior, "Analyzing the effect of health reforms on the efficiency of ecuadorian public hospitals," International Journal of Health Economics and Management, 2023.

[17] E. Schot, L. Tummers, and M. Noordegraaf, "Working on working together. a systematic review on how healthcare professionals contribute to interprofessional collaboration," Journal of Interprofessional Care, vol. 34, no. 3, pp. 332–342, 2019.

[18] M. Villalobos-Cid, M. Chacón, P. Zitko, and M. Inostroza-Ponta, "A new strategy to evaluate technical efficiency in hospitals using homogeneous groups of casemix: How to evaluate when there is not DRGs?" Journal of medical systems, vol. 40, no. 4, p. 103, 2016.

[19] S. Kohl, J. Schoenfelder, A. Fügener, and J. O. Brunner, “The use of data envelopment analysis (DEA) in healthcare with a focus on hospitals,” Health Care Management Science, vol. 22, no. 2, pp. 245–286, 2018.

[20] E. Z. Asbu, M. D. Masri, and M. A. Naboulsi, “Determinants of hospital efficiency: A literature review,” International Journal of Healthcare, vol. 6, no. 2, p. 44, May 2020.

[21] N. Mihailovic, S. Kocic, and M. Jakovljevic, “Review of diagnosis-related group-based financing of hospital care,” Health Services Research and Managerial Epidemiology, vol. 3, p. 233339281664789, Jan. 2016.

[22] M. Henríquez, Eficiencia del gasto hospitalario en Chile. Fundación ObservatorioFiscal, 2020.

[23] J. Odeck, “Measuring technical efficiency and productivity growth: a comparison of sfa and dea on norwegian grain production data,” Applied Economics, vol. 39, no. 20, pp. 2617–2630, 2007.

[24] G. Katharakis, M. Katharaki, and T. Katostaras, “An empirical study of comparing dea and sfa methods to measure hospital units’ efficiency,” International Journal of Operational Research, vol. 21, no. 3, pp. 341–364, 2014.

[25] J. Dong and D. Wu, “An evaluation of the impact of ecological compensation on the cross-section efficiency using SFA and DEA: A case study of xin’an river basin,” Sustainability, vol. 12, no. 19, p. 7966, Sep. 2020.

[26] J. Aparicio, M. Gonzalez, J. J. Lopez-Espin, and J. T. Pastor, “A parameterized scheme of metaheuristics to solve NP-hard problems in data envelopment analysis,” in International Series in Operations Research Management Science. Springer International Publishing, 2016, pp. 195–224. [Online]. Available: https://doi.org/10.1007/978-3-319-48461-7_9

[27] M. González, J. J. López-Espín, J. Aparicio, D. Giménez, and E.-G. Talbi, “A parameterized scheme of metaheuristics with exact methods for determining the principle of least action in data envelopment analysis,” in 2017 IEEE Congress on Evolutionary Computation (CEC), 2017, pp. 588–595.

[28] N. Zhu, C. Zhu, and A. Emrouznejad, “A combined machine learning algorithms and dea method for measuring and predicting the efficiency of chinese manufacturing listed companies,” Journal of Management Science and Engineering, vol. 6, no. 4, p. 435–448, 2021.

[29] S. V. G. Subrahmanya, D. K. Shetty, V. Patil, B. M. Z. Hameed, R. Paul, K. Smriti, N. Naik, and B. K. Somani, “The role of data science in healthcare advancements: applications, benefits, and future prospects,” Irish Journal of Medical Science (1971 -), 2021. [Online]. Available: https://doi.org/10.1007/s11845-021-02730-z

[30] M. S. Islam, M. M. Hasan, X. Wang, H. D. Germack, and M. Noor-E-Alam, “A systematic review on healthcare analytics: Application and theoretical perspective of data mining,” Healthcare, vol. 6, no. 2, 2018. [Online]. Available: https://www.mdpi.com/2227-9032/6/2/54

[31] N. R. Vajjhala and P. Eappen, Data Envelopment Analysis in Healthcare Management: Overview of the Latest Trends. IGI Global, 2024, p. 245–260.

[32] K. Batko and A. Slezak, “The use of big data analytics in healthcare,” Journal of Big Data, vol. 9, no. 1, 2022. [Online]. Available: https://doi.org/10.1186/s40537-021-00553-4

[33] J. E. Kasten, “Big data applications in healthcare administration,” in Research Anthology on Big Data Analytics, Architectures, and Applications. IGI Global, 2022, pp. 1003–1034. [Online]. Available: https://doi.org/10.4018/978-1-6684-3662-2.ch048

[34] N. M. Guerrero, J. Aparicio, and D. Valero-Carreras, “Combining data envelopment analysis and machine learning,” Mathematics, vol. 10, no. 6, p. 909, Mar. 2022. [Online]. Available: https://doi.org/10.3390/math10060909

[35] S. Boubaker, T. D. Q. Le, T. Ngo, and R. Manita, “Predicting the performance of msmes: a hybrid dea-machine learning approach,” Annals of Operations Research, 2023.

[36] T. Duras, F. Javed, K. Månsson, P. Sjölander, and M. Söderberg, “Using machine learning to select variables in data envelopment analysis: Simulations and application using electricity distribution data,” Energy Economics, vol. 120, p. 106621, 2023.

[37] S. Hamidi and F. Akinci, “Measuring efficiency of health systems of the middle east and north africa (MENA) region using stochastic frontier analysis,” Applied Health Economics and Health Policy, vol. 14, no. 3, pp. 337–347, 2016.

[38] A. D. Alatawi, L. W. Niessen, and J. A. M. Khan, "Determinants of technical efficiency in public hospitals: The case of saudi arabia," Health Economics Review, vol. 10, no. 1, 2020. [Online]. Available: https://doi.org/10.1186/s13561-020-00282-z

[39] W. H. O. R. O. for Europe, E. O. on Health Systems, Policies, J. Cylus, I. Papanicolas, and P. C. Smith, How to make sense of health system efficiency comparisons?, ser. Health Systems and Policy Analysis: policy brief, 27. World Health Organization. Regional Office for Europe, 2017.

[40] M. C. Breitenbach, V. Ngobeni, and G. C. Aye, "Global healthcare resource efficiency in the management of covid-19 death and infection prevalence rates," Frontiers in Public Health, vol. 9, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpubh.2021.638481

[41] C. Henriques and M. Gouveia, "Assessing the impact of covid-19 on the efficiency of portuguese state-owned enterprise hospitals," Socio-Economic Planning Sciences, vol. 84, p. 101387, 2022.

[42] S. Chen, Y. Li, Y. Zheng, B. Wu, R. Bardhan, and L. Wu, "Technical efficiency evaluation of primary health care institutions in shenzhen, china, and its policy implications under the COVID-19 pandemic," International Journal of Environmental Research and Public Health, vol. 20, no. 5, p. 4453, 2023.

[43] S. Maity and A. Sinha, "Technical efficiency and its determinants in regulating adolescents' coronavirus infection across asian countries," Scientific Reports, vol. 13, no. 1, 2023. [Online]. Available: http://dx.doi.org/10.1038/s41598-023-45442-3

[44] T. Coelli, A. Estache, S. Perelman, and L. Trujillo, Una introducción a las medidas de eficiencia para reguladores de servicios públicos y de transporte, ser. Economía internacional. Banco Mundial, 2003. [Online]. Available: https://books.google.cl/books?id=FVjFAAAACAAJ

[45] I. Zakowska and M. Godycki-Cwirko, "Data envelopment analysis applications in primary health care: a systematic review," Family Practice, Oct. 2019. [Online]. Available: https://doi.org/10.1093/fampra/cmz057

[46] J. Zhu, "DEA under big data: data-enabled analytics and network data envelopment analysis," Annals of Operations Research, vol. 309, no. 2, pp. 761–783, 2020.

[47] S. S. Ganji and A. A. Rassafi, "DEA malmquist productivity index based on a double-frontier slacks-based model: Iranian road safety assessment," European Transport Research Review, vol. 11, no. 1, Jan. 2019. [Online]. Available: https://doi.org/10.1186/s12544-018-0339-z

[48] A. Alinezhad and M. Mirmozaffari, "Malmquist Productivity Index Using Two-Stage DEA Model in Heart Hospitals," Iranian Journal of Optimization, vol. 10, no. 2, pp. 81–92, 2018. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02277863

[49] R. Dyson, R. Allen, A. Camanho, V. Podinovski, C. Sarrico, and E. Shale, "Pitfalls and protocols in dea," European Journal of Operational Research, vol. 132, no. 2, pp. 245–259, 2001, data Envelopment Analysis.

[50] M. González, J. J. López-Espín, J. Aparicio, D. Giménez, and J. T. Pastor, "Using genetic algorithms for maximizing technical efficiency in data envelopment analysis," Procedia Computer Science, vol. 51, pp. 374–383, 2015. [Online]. Available: https://doi.org/10.1016/j.procs.2015.05.257

[51] M. Alimohammadlou and S. Mohammadi, "Evaluating the productivity using malmquist index based on double frontiers data," Procedia - Social and Behavioral Sciences, vol. 230, pp. 58–66, Sep. 2016. [Online]. Available: https://doi.org/10.1016/j.sbspro.2016.09.008

[52] R. Jaberian and H. Ahady Dolatsara, "A dea and machine learning based platform for investigating success of initial coin offerings (icos)," SSRN Electronic Journal, 2023.

[53] J. Wu and Y. Wu, "Machine learning in data envelopment analysis: A smart mechanism for indicator selection," JUSTC, vol. 52, no. 12, p. 5, 2022.

[54] Z. Zhang, Y. Xiao, and H. Niu, "Dea and machine learning for performance prediction," Mathematics, vol. 10, no. 10, p. 1776, 2022.

[55] E. Santelices C, H. Ormeño C, M. Delgado S, C. Luis M, R. Valdés V, and L. Durán, "Análisis de los determinantes de la eficiencia hospitalaria: el caso de Chile," Revista médica de Chile, vol. 141, pp. 457 – 463, 2013.

[56] R. Kalhor, S. Amini, M. Sokhanvar, F. Lotfi, M. Sharifi, and E. Kakemam, "Factors affecting the technical efficiency of general hospitals in iran," Journal of the Egyptian Public Health Association, vol. 91, no. 1, pp. 20–25, 2016.

[57] R. Ayiko, P. N. Mujasi, J. Abaliwano, D. Turyareeba, R. Enyaku, R. Anguyo, W. Odoch, P. Bakibinga, and T. Aliti, "Levels, trends and determinants of technical efficiency of general hospitals in uganda: data envelopment analysis and tobit regression analysis," BMC Health Services Research, vol. 20, no. 1, 2020.

[58] M. A. Subhi, "Technical efficiency, productivity, and determinants of technical inefficiency of local hospitals in oman: Using data envelopment analysis," International Journal of Social Sciences and Economic Review, pp. 10–17, 2022.

[59] R. Castro, "Midiendo la eficiencia municipal en salud," in Serie informe Social 97, 2004.

[60] M. T. Ramírez-Valdivia, S. Maturana, and S. Salvo-Garrido, "A multiple stage approach for performance improvement of primary healthcare practice," Journal of Medical Systems, vol. 35, no. 5, pp. 1015–1028, 2010.

[61] M. Ramírez-Valdivia, S. Maturana, J. Mendoza-Alonzo, and J. Bustos, "Measuring the efficiency of chilean primary healthcare centres," International Journal of Engineering Business Management, vol. 7, p. 15, 2015.

[62] M. Cornejo-Ovalle, R. Brignardello-Petersen, and G. Pérez, "Pay-for-performance and efficiency in primary oral health care practices in chile," Revista Clínica de Periodoncia, Implantología y Rehabilitación Oral, vol. 8, no. 1, pp. 60–66, 2015.

[63] A. Candia and M. Petitpa, "Principales determinantes de la ineficiencia hospitalaria: una revisión bibliográfica," in Serie informe social 158, 2007, pp. 1–20.

[64] E. Santelices, H. Ormeño, M. Delgado, C. Luis, R. Valdés, and L. Durán, "Análisis de la eficiencia técnica hospitalaria 2011," Revista médica de Chile, vol. 141, no. 3, pp. 332–337, 2013.

[65] R. Castro, "Midiendo la (in)eficiencia de los hospitales públicos en chile," in Serie informe Social 83, 2006, p. 97.

[66] P. Barahona-Urbina, "Análisis de eficiencia hospitalaria en Chile," Anales de la Facultad de Medicina, vol. 72, pp. 33 – 38, 2011.

[67] A. C. Gonzalez, J. Lillo, M. Inostroza-Ponta, and M. Villalobos-Cid, "Evaluating the categorisation of the public hospitals in chile according to case-mix complexity: a genetic algorithm approach," in 2020 39th International Conference of the Chilean Computer Science Society (SCCC). IEEE, Nov. 2020. [Online]. Available: https://doi.org/10.1109/sccc51225.2020.9281282

[68] R. Castro, "Midiendo la eficiencia de los hospitales públicos en chile," in Serie en Foco N°108, Expansiva, 2007.

[69] J. Giglio, G. Gaete-Lucero, and M. Villalobos-Cid, "Classification of chileans public hospitals based on healthcare production using clustering techniques," in 2021 40th International Conference of the Chilean Computer Science Society (SCCC), 2021, pp. 1–8.

[70] M. Dorn, B. I. Grisci, P. H. Narloch, B. C. Feltes, E. Avila, A. Kahmann, and C. S. Alho, "Comparison of machine learning techniques to handle imbalanced COVID-19 CBC datasets," PeerJ Computer Science, vol. 7, p. e670, 2021.

[71] B. C. Feltes, I. A. Vieira, J. Parraga-Alava, J. Meza, E. Portmann, L. Terán, and M. Dorn, "Feature selection reveal peripheral blood parameter's changes between covid-19 infections patients from brazil and ecuador," Infection, Genetics and Evolution, vol. 98, p. 105228, 2022.

[72] M. J. Prado, S. Singh, R. Ligabue-Braun, B. V. Meneghetti, T. Rispoli, C. Kopacek, K. Monteiro, A. Zaha, M. L. R. Rossetti, and A. V. Pandey, "Characterization of mutations causing CYP21a2 deficiency in brazilian and portuguese populations," International Journal of Molecular Sciences, vol. 23, no. 1, p. 296, 2021.

[73] ——, "Characterization of mutations causing steroid 21-hydroxylase deficiency in brazilian and portuguese populations," Int J Mol Sci, 2021.

[74] D. Deconte, C. B. L. dos Santos, C. O. de Morais, T. M. Yonamine, L. T. Nogueira, M. A. T. Ferreira, V. B. Franceschi, A. L. S. Longhi, R. A. R. Villacis, S. R. Rogatto, R. Ligabue-Braun, P. R. G. Zen, R. F. M. Rosa, and M. Fiegenbaum, "Unusual features in a child with marshall-smith syndrome due to a novel NFIX variant: Evidence for an abnormal protein function," Gene Reports, vol. 22, p. 100991, mar 2021.

[75] H. J. G. Palacios, R. A. J. Toledo, G. A. H. Pantoja, and Á. A. M. Navarro, "A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change," Advances in Science, Technology and Engineering Systems Journal, vol. 2, no. 3, pp. 598–604, Jun. 2017. [Online]. Available: https://doi.org/10.25046/aj020376

[76] J. S. Saltz and I. Krasteva, "Current approaches for executing big data science projects—a systematic literature review," PeerJ Computer Science, vol. 8, p. e862, Feb. 2022. [Online]. Available: https://doi.org/10.7717/peerj-cs.862

[77] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 01 2000.

[78] C. Lenette, "What is participatory action research?" in Participatory Action Research. Oxford University Press, 2022, pp. 1–20.

[79] S. Kindon, M. Kesby, and R. Pain, Eds., Participatory action research approaches and methods, ser. Routledge Studies in Human Geography. London, England: Routledge, Aug. 2010.

[80] V. Pontello, H. Beckmann, and C. Lanquillon, "Meta-learning approach for implementation of ai methods in the context of crisp-dm with case studies from master data management," in 2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC). IEEE, Jun. 2021.

[81] M. Villalobos-Cid, M. Chacon, P. Zitko, and M. Instroza-Ponta, "A New Strategy to Evaluate Technical Efficiency in Hospitals Using Homogeneous Groups of Casemix : How to Evaluate When There is Not DRGs?" Journal of Medical Systems, vol. 40, no. 4, p. 103, 2016.

[82] M. A. Subih, B. H. Malik, I. Mazhar, I. ul Hassan, U. Sabir, T. Wakeel, W. Ali, A. Yousaf, B. bin Ijaz, H. Nawaz, and M. Suleman, "Comparison of agile method and scrum method with software quality affecting factors," International Journal of Advanced Computer Science and Applications, vol. 10, no. 5, 2019.

[83] E. Caron, A. Chevalier, N. Baillon-Bachoc, and A.-L. Vion, "Heuristic for license-aware, performant and energy efficient deployment of multiple software in Cloud architecture," in ICICS 2021 - 12th International Conference on Information and Communication Systems, 2021. [Online]. Available: https://hal.science/hal-03572922

[84] D. Lupu and R. Tiganasu, "COVID-19 and the efficiency of health systems in europe," Health Economics Review, vol. 12, no. 1, 2022.