Sapienza
Università di Roma

Department of Computer, Control, and Management Engineering

The Role of Digital Twins in Modernizing Healthcare Systems

Productivity and Efficiency Analysis

Project by Group O:

Buzzichini Davide, Di Pietro Luca, Mazza Manuel, Perla Lorenzo, Rori Matteo

Abstract

Digital Twin (DT) technology is a promising evolving innovation that holds the potential to revolutionize healthcare by enabling a broad range of useful applications. This report presents a systematic review of the literature on DTs to explore their applications and potential benefit on healthcare sector. Using a structured PRISMA framework, 26 high-quality studies were selected, demonstrating how DTs can enhance care delivery and operational efficiency. In addition to the literature review, the report conducts a comprehensive efficiency analysis of healthcare systems in Europe, focusing on the more general adoption of technologies in the sector. Data from OECD and EUROSTAT were used to evaluate the relationships between technological investments, healthcare performance, and socio-economic factors, leading to a Data Envelopment Analysis (DEA) enhanced by Bootstrap methods and revealing that healthcare systems could achieve an 18% increase in efficiency on average by optimizing the use of existing technologies. The findings highlight the transformative role of technology in healthcare while acknowledging the challenges of its adoption, as high costs, data security concerns, and organizational resistance. In conclusion, while DTs represent a promising frontier, the optimized utilization of healthcare technologies remains essential for achieving sustainable and efficient healthcare systems across Europe.

Academic Year 2024/2025

Contents

1 Introduction

Healthcare has always reasonably been one of the most important sectors in any existing society and, over the span of the last centuries, it has experienced a massive evolution in medical practices, knowledge and digital transformation, driven by emerging technologies that promise to enhance patient outcomes and operational efficiency. Despite its numerous changes, the main objective has never mutated: making healthcare more efficient. For this reason, it is important to consider how healthcare is managed differently across countries: it can be either private or public while financial, technological and human resources can be available (or not) and allocated following a number of logics which have huge impact on its efficiency. Moreover, part of the efficiency can depend on exogenous factors which, in a scientific analysis, would be considered as "context variables", such as GDP per capita. One of the most transformative solutions derived by the digital innovation in this sector is **Digital Twins (DTs)**, considered to be able to *"revolutionize healthcare systems by leveraging real-time data integration, advanced analytics, and virtual simulations to enhance patient care, enable predictive analytics, optimize clinical operations, and facilitate training and simulation"[...]"Its implementation has the potential to significantly improve patient outcomes, enhance patient safety"* **(A. Vallée, 2023)**. DTs create virtual replicas of physical systems, enabling a broad range of applications, improving already existing medical practices, for instance *"[...] a true digital twin expands on previous surgery simulators by including not only generic anatomic models, but patient-specific digital information with real-time calibration of simulations using intraoperative data"[...]"and, in its optimum form, the digital twin could serve as a virtual environment where novel surgical procedures can be explored, and new ones can be designed"* **(Bjelland et al. ,2022)**. This report aims to analyze the impact of DTs on the **healthcare system efficiency** in European countries, presenting findings and studies from the literature on the applications of DTs, and gathering data able to demonstrate how technological innovations bring improvements in the healthcare sector. After producing a systematic review analyzing the existing literature on the topic, we gathered all the pertinent data from public databases in order to do a sound efficiency analysis, being careful to choose the most relevant variables that suited our research. Through a careful evaluation of our area of analysis, we formulated the following questions to guide our research:

Research Questions
Q1: How can Digital Twin technology improve healthcare system efficiency in European countries?
Q2: What impact does the implementation of Digital Twins have on the cost-effectiveness of healthcare delivery?
Q3:What are the most effective methods for assessing healthcare system efficiency, and when is the use of traditional analytical techniques sufficient for evaluating performance?
Q4: How much do European countries invest in the adoption and implementation of Digital Twin technologies for improving healthcare outcomes?

Table 1: Research Questions and Objectives

2 Systematic Review

A systematic review is a type of approach to answer a very specific research question using a clearly defined and predetermined protocol. The procedures include clearly stated objectives, well-defined eligibility criteria, and an extensive search for relevant studies among the existing literature about the investigated topic. Systematic reviews will be particularly useful in drawing robust, evidence-based conclusions from medicine and social sciences. These could form conclusions to guide practical decisions, or suggest new experiments.

2.1 Framework and Query

The **CIMO framework** *(Figure 16, in the appendix)* provides a structured approach to analyze the impact of Digital Twins in healthcare. It will help identify the contexts in which Digital Twins are most effective, the interventions they enable, the mechanisms behind their success, such as predictive analytics, and the outcomes they deliver which ranges from better patient care to improved efficiency and cost savings.

This approach supports both evaluating current applications and guiding future developments in healthcare. It guided the refinement of the search process using the primary framework through specific exploration in the publication of research related to Digital Twin and efficiency analysis, using **Scopus** as our database. This involved the development of a narrow search query to focus on pertinent studies, as described in *Figure 17* of the *Appendix*.

2.2 PRISMA

PRISMA is a method for ensuring that a **systematic review** is transparent and rigorous, reducing bias and thereby enhancing reproducibility. It follows a structured approach through four stages: identification, screening, eligibility, and inclusion. This work identified the studies through a comprehensive literature search using the Scopus database, covering a wide literature on topics related to technology and health. The detailed search string was developed incorporating words such as ”digital twin”, ”health systems”, performance indicators like ”efficiency” and ”effectiveness”, among others. The total initial number of identified studies obtained was 319, thereby giving the review a wide basis yet to be shrank on the most relevant material. The screening stage involved the removal of duplicates, followed by a filter on the titles and abstracts. Further limitation, including only articles published in 2018 and beyond, focusing on only journal articles or reviews, reduced this number to 170. During this stage, only works that were clearly irrelevant, such as those not about health or Digital Twins, would be excluded. Finally, the eligibility and inclusion stages were combined into one detailed review of the full texts. This was in line with the specific criteria: it needed to deal with Digital Twins used in healthcare and must also demonstrate empirical evidence or, alternatively, strong theoretical insights. Works focused on other domains besides healthcare, chapters of books, and non-English publications have been excluded. This elaborate process finally provided 26 high-quality studies[1] that were pointed out as the most relevant and impactful ones for our

[1] A descriptive table about all the selected papers is attached at the end of the *Appendix*

study. This systematic approach allows for a reliable and comprehensive evaluation of the potential of Digital Twins in improving healthcare efficiency and effectiveness.

Our specific PRISMA diagram is displayed at *(Figure 18, in the appendix)*

2.3 Bibliometrix

Finally, the results of the analysis after its processing were exported into a ".csv" file used as input in **Bibliometrix** to carry out different graphical analyses. We preferred keyword analysis to determine the most relevant concepts and their importance order. Further, with the help of the built-in conceptual framework tool in Bibliometrix, we created a "**co-occurrence network**" of concepts in Figure 1. It shows three clear clusters represented in green, red, and blue. The green cluster groups keywords associated with "human" and "digital twin," showing strong connections to such themes as "personalized medicine," "artificial intelligence," and "health care system." These terms emerge as pivotal, showing their growing importance as drivers of innovation and personalization within the healthcare domain. The red cluster centers around words and phrases like "health care" and "medical services," further linked to concepts like "blockchain" and "augmented reality", highlighting the growth in interest around integrating leading-edge technologies into health service provision. The blue cluster, in contrast, brings in words like "automation" and "digital devices," reflecting an upsurge in the application of digital technologies to health operations and patient monitoring. According to the co-occurrence network visualization, some keywords appear to be peripheral to the core themes yet remain highly interconnected, like those about methodologies and theoretical approaches as "algorithm" and "systematic review". These terms, although more distant from the center, represent singular perspectives that provide valuable insights into the topic under study.

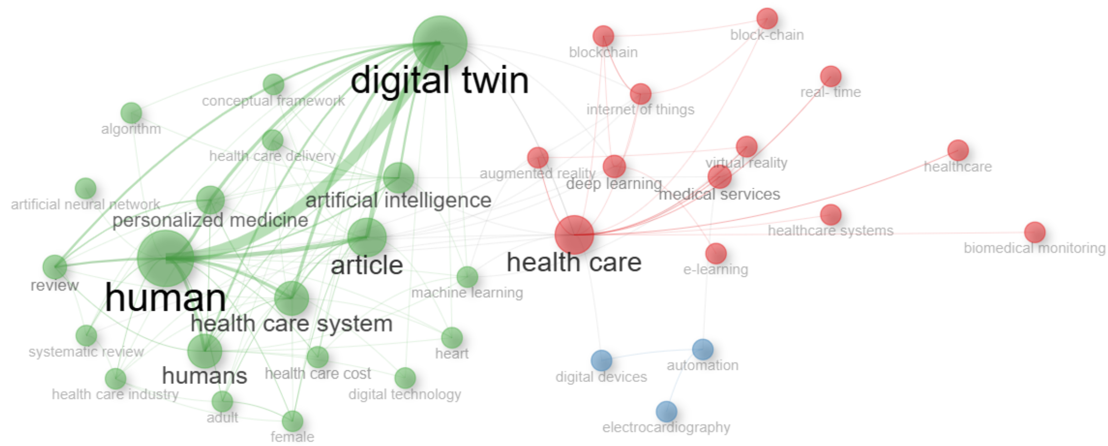

Figure 1: Co-occurence Network

In Bibliometrix, the "**Thematic Map**" provides a graphical representation of the distribution of topics and clusters of documents within a two-dimensional framework. This tool shows the main research areas and dominant themes within a dataset of scientific documents.

Clustering techniques or network analysis can be used to group documents in thematic maps, showing shared characteristics. Clusters or themes are placed on the map, where their position corresponds with the relation and degree of similarity between them. Characteristics of a Thematic Map include:

- **Node size:** Reflects the significance or occurrence rate of a cluster in a particular dataset. Larger nodes represent broader themes or those that have been more discussed.
- **Link among nodes:** Thematic relationship or similarities; the closer, the stronger. Node Labels: Keywords or concepts summarizing the central idea of each cluster.
- **Node labels:** Keywords or concepts summarizing the central idea of each cluster.
- **Colors:** Visual distinction among clusters, enabling one to recognize distinct thematic areas.

In this case, the map (Figure 2) has been used to assess clusters using "Development degree" (Density) and "Relevance degree" (Centrality) as its axes. This procedure identified six clusters distinguished by color, as the figure shows. For example, the brown cluster, located in the top-right quadrant, highlights "human", "digital twin", and "health care" as predominant themes, therefore addressed as "motor themes", while the pink and the orange clusters emphasizes new emerging relevant concepts such as "augmented reality", "cost-effectiveness" and "biomedical monitoring". The Thematic Map provides insight into the major thematic clusters of this research area and gives a proper understanding of the trends and relationships in the bibliometric landscape.

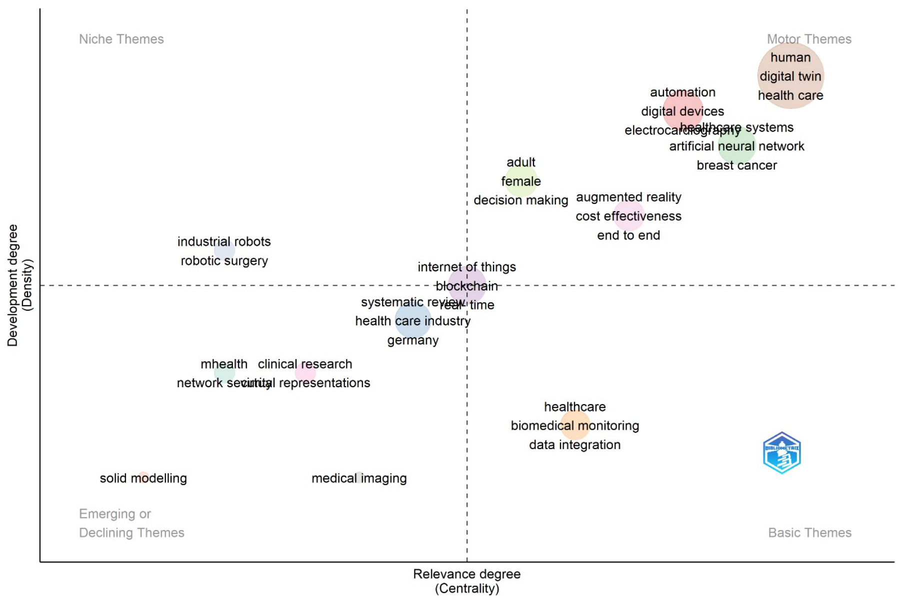

Figure 2: Thematic Map

2.4 Literature Observations

DTs are revolutionizing the face of healthcare on grounds of diagnosis, precision treatment, operational efficiency, and cost optimization. In regard to COVID-19, digital twin technologies have optimized resource distribution within ICUs, thus reducing queues while fairly distributing resources **(De Benedictis et al., 2023)**. A first study **(Ahmed et al., 2022)** reported a reduction of 30% in diagnosis related to COVID-19, while another **(Shen et al., 2024)** showed savings of $5,000 annually per patient because of a decrease in unnecessary procedures. In preventive care, on the other hand, DT-based monitoring decreased the rate of emergency admissions for elderly patients by 20%, shifting toward proactive health management rather than being reactive **(Wu et al., 2024)**. Other operational benefits include financial ones, as predictive maintenance of DT systems has reduced medical device repair costs by 20% **(Manocha et al., 2023)**. While many accomplishments in these areas have been attained, there are several limiting factors to the broad proliferation of DTs. Based on the insights gained from the selected sources, we have identified some of the main obstacles to their expansion. Smaller organizations typically struggle with the computational burden required for high-fidelity simulations while organizational and cultural resistance to change will continue to be barriers for adoption, as well as concerns regarding data security, ethical and regulatory issues to ensure that such private and sensible information doesn't end up in the wrong hands. Summing up, Digital Twins are much more than a technological development; they are a concept that means predictive, personalized, and efficient healthcare. Though interoperability, privacy, and computational demands are issues that must be dealt with, remarkable progress achieved by interdisciplinary and international cooperation illustrates the potential of DTs to transform healthcare systems. Overcoming these barriers could allow DTs to redefine the way care is provided and create a future that is not only more technological but also more humane and efficient.

3 Data

3.1 Data Collection Process

Collecting data, especially the numeric kind, about Digital Twin in healthcare has proved itself to be a really tough challenge. Numerous studies indeed discuss on its usefulness, such as its many applications and benefits, but finding enough actual data about these topics still represents an hard task, logically due to privacy and sensitivity concerns. Given this issue, it has been decided to pursue an efficiency analysis on the broader of the impact of technology on the efficiency of the health sector, rather than finding ourselves stuck with an insufficient amount of data focusing exclusively on Digital Twin. The said data has been gathered manually over the course of a few days, from a number of databases extracted from the **OECD** (Organization for Economic Co-operation and Development) website merged with other databases from the **EUROSTAT** (Statistical Office of the European Union) website to have a more complete dataset to work with. It appears extremely important to note that both open-sources databases consulted offer absolute transparency presenting reliable data and the combination of the two is what makes the analysis sound. For what concerns the databases in the EUROSTAT website, it provided us with some of the more descriptive data, such as: GDP per capita and total population. On the other hand, the OECD website supplied all the other data, especially those regarding the "Healthcare" and "Healthcare Technology" category. Before the data could be merged into a unique dataset, the need to pre-process all of it arose: firstly it has been quite common to encounter irrelevant data for the analysis which has been logically ignored, as well as missing values for some years or some countries that led, when it was possible and viable, to their estimation simply following the trend that actual data remarked in the observations. For example, if data about Italy in 2020 is hypothetically missing, we consider the Italian trend from 2010 to 2019 and the average variation observed in every other country from 2019 to 2020. It's important to notice that this interested only approximately 10 observations which constitutes a minimal amount, therefore not compromising the robustness of our dataset. Moreover the databases needed to be scaled and normalized. Regarding the scaling, we have considered observations going from 2010 to 2020 in order to have a sufficient amount and to be coherent with the technological aspect we wanted to initially give to the analysis; country-wise we kept only data observed in 20 European countries (United Kingdom included for simplicity) to have a broad range of informations. It's crucial to notice that our initial will was to include databases from the US as well, but we eventually avoided to do so due to missing data related to some variables. Finally the databases gathered showed any observation already normalized per capita (or by some other metric) which made the values comparable between different size countries. Data merging has been the last needed step of the process: Excel was the only tool utilized, as databases were manually transferred on the sheet to merge every observation from any gathered variable (column), country and year (row).

3.2 Data Description and Distribution

The variable selection was driven by the will of representing togheter environmental factors, healthcare system indicators and technology adoption measurements, making them crucial for a comprehensive analysis of **technological impact on healthcare efficiency**. They indeed offer a balanced view of financial resources, technological capability and the healthcare workforce in relation to the population's needs and outcomes. Some alternatives were considered but they often lack meaningfulness for our study or, in some cases, a too large amount of data for our selected countries. The chosen variables are widely recognized in healthcare economics and performance analysis, making them robust and comparable metrics. A brief description of all the selected variables can be seen in the appendix, in Table 3.

Analyzing the average distribution of the health status variable *(shown in Figure 3)* we can observe that Ireland and Sweden lead with the highest average health status, indicating exceptional healthcare quality and favorable living conditions, whereas countries like Poland, Hungary, and Czechia have the lowest average health statuses, reflecting challenges in healthcare provision and broader socio-economic factors. Germany, despite being a major economic leader in Europe, lies below average in this statistic. *Figure 4* Countries like France, Spain, and Italy show the highest average inverse mortality rates, indicating strong healthcare performance relative to mortality. Once again, Germany falls shortly below the average, suggesting inefficiencies in translating resources into health outcomes. The lowest values are observed in countries like Hungary, Slovakia and Poland, highlighting significant room for improvement in reducing mortality. Furthermore, all the other variables' average distributions are shown in the appendix, from Figure 19 to Figure 23.

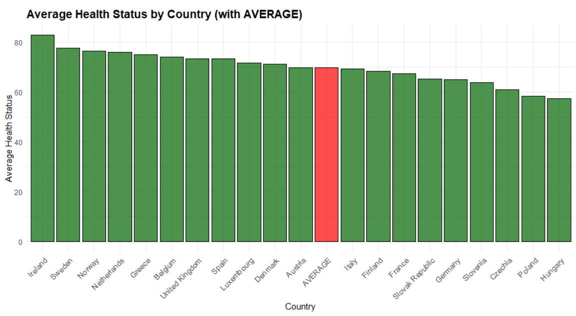

Figure 3: Average health status by country

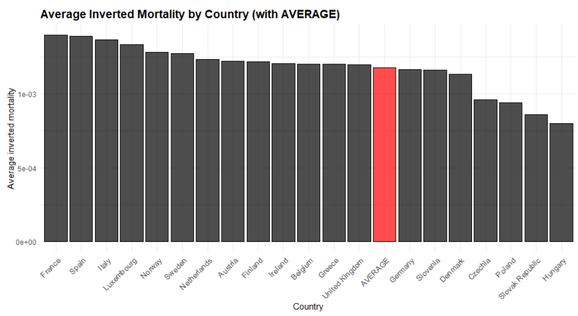

Figure 4: Average inverse mortality by country

4 Model and Method

For what concerns the preparation of data prior to its analysis, we started by loading the data from an Excel file, using the "readxl" library, and consequently utilized the "**FEAR**", "**benchmarking**" and "**frontier**" libraries for the efficiency analysis in R. Once the data was loaded, we created a new variable by combining the elderly population (Over65_Pop) with the total population (Population) which was necessary due to the high correlation between the two variables, as highlighted by *(Figure 5)*.

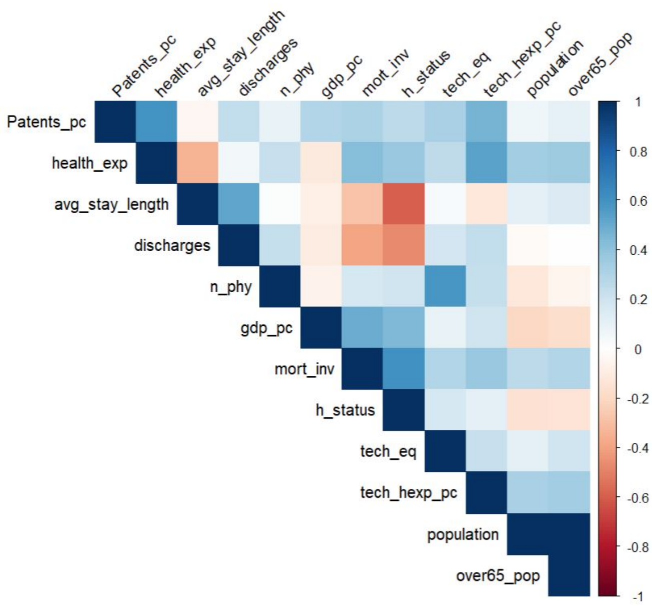

Figure 5: Initial variables correlation

Combining them simplified the model without losing any important information. Next, we defined a Min-Max normalization function to scale the numerical variables between 0 and 1, with 0 being the minimum value and 1 the maximum of each variable. This was meant to ensure that all the variables were comparable and to eliminate any scale difference. Once we defined the function, we applied it only to the numerical observations, standardizing all the variables so that they could be analyzed more effectively.

4.1 Model

In order to identify which variables will be considered in the model itself and select the most suitable inputs and outputs for the analysis, we needed to conduct a few preliminary tests. First order of business was to calculate the **Pearson coefficient** whose values range between -1 and 1: positive/negative values indicate positive/negative linear correlation between the variables, while 0 means that inputs are not correlated at all. Executing this procedure in R and always keeping in mind the logical idea of minimizing inputs and maximizing outputs, we selected as inputs "Patents in healthcare technology", "Health expenditure", "GDP per capita", "Medical Technology Availability", "ALOS", "Hospital discharges", "Population (Pop+Over65_pop)", "Active physicians" and "Technological appliances for healthcare's expenditure"; consequently, as outputs we chose "Inverse mortality" and "Perceived health status" since they provide important insights into the healthcare system's performance. To streamline the calculations, we organized the inputs and outputs into matrices.

Inputs	**Outputs**
Health expenditure	Health status
GDP per capita	Inverse mortality
Patents	
Medical tech equipment	
Tech health expenditure	

Table 2: Inputs and outputs

Finally we were able to perform a correlation test through R to examine relationships between inputs and outputs. The resulting correlation matrix provided valuable insights which guided our next step: whenever we spotted an inverse correlation (negative Pearson index value) or a weak correlation (positive but fairly low Pearson index value) between an input and an output we excluded the variable from the final list of inputs. As a result of such operation, we excluded Hospital discharges, ALOS and Population due to negative correlation and Active physicians due to weak correlation; it is important to notice that these observations will still be present in the analysis but from this point on will be considered as "context variables". The final scheme of inputs and outputs for our efficiency analysis and their correlation is thereby presented by the following matrix:

	mort_inv	h_status
Patents_pc	0.3141733	0.2618006
health_exp	0.4293827	0.3783542
gdp_pc	0.4999383	0.4479747
tech_eq	0.2918426	0.1746374
tech_hexp_pc	0.3739466	0.1155544

Figure 6: Correlation matrix between inputs and outputs

4.2 Method

4.2.1 DEA

The most suitable method to evaluate the healthcare sector efficiency in European countries, based on insights gathered during the literature inspection for our systematic review, is a non-parametric method known as **Data Envelopment Analysis (DEA)**. This is a deterministic frontier approach that calculates relative efficiency and allows to identify sources of inefficiency and compare subjects among themselves. As one of its most crucial advantages, DEA is able to simultaneously handle multiple inputs and outputs without specifying a predefined functional relationship between the variables. The relative efficiency for a subject with data (x0,y0)(x0,y0) is given by the solution to the following linear programming problem:

$$\hat{\lambda}_{\text{DEA}}(x_0, y_0) = \sup\left\{\lambda \mid (x_0, \lambda y_0) \in \hat{\psi}_{\text{DEA}}\right\} \tag{1}$$

More explicitly, this can be formulated as:

$$\hat{\lambda}_{\text{DEA}}(x_0, y_0) = \max\left\{\lambda \mid \lambda y_0 \leq \sum_{i=1}^{n} \gamma_i Y_i,\ x_0 \geq \sum_{i=1}^{n} \gamma_i X_i,\ \lambda > 0,\ \sum_{i=1}^{n} \gamma_i = 1,\ \gamma_i \geq 0 \text{ for } i = 1, \ldots, n\right\} \tag{2}$$

Since the outcomes of DEA are susceptible to bias, introduced by data variability, we needed to integrate the Bootstrap method in order to make the inferences more robust and solid: the Bootstrap consists of selecting multiple random samples from the original data set and then running DEA on each sampled set. The result is a distribution of efficiency scores that more accurately represents the said natural data variability. As a matter of fact, we can say that the efficiencies calculated through **DEA Bootstrap** represent the regular DEA efficiencies without the bias term. Such analysis will be further discussed in Chapter 5.2.

Before we are able to apply DEA for our analysis through R, a few last steps were necessary in order to respect some crucial assumptions of the method, such as the free disposal hypothesis and the convexity of the production set. For the first one we could easily assume it was respected since any hospital or medical institution might figuratively discard resources without having any cost as a consequence; for the second one, on the other hand, it was not as easy since we needed to perform a convexity test to prove that the data distribution could be analyzed through the DEA approach (if we couldn't prove convexity we would have selected **Free Disposal Hull (FDH)**). Another important test we carried out was the Returns to Scale's which is aimed at recognizing the most fitting returns to scale that characterize our DEA, between constant and variable.

4.2.2 Convexity Test

This test evaluates the following hypotheses:

$$\begin{cases} H_0 : \mu \xrightarrow{n^k} \theta_{DEA} \\ H_1 : \mu \xrightarrow{n^k} \theta_{FDH} \end{cases}$$

The test statistic was $\tau = 1.686$, while the p-value was 0.7537. As the p-value exceeded 0.05, we did not have sufficient evidence to reject the null hypothesis H0. This confirmed that the production set was convex, allowing us to apply DEA. If the set turned out to be non-convex, we would have considered using FDH (Free Disposal Hull) instead.

4.2.3 Returns to Scale Test

To determine how the efficiency of the analyzed **DMUs** change when the scale of its operations is varied, we performed a return to scale (RTS) test. The hypotheses for this test were:

$$\begin{cases} H_0 : \mu \xrightarrow{n^k} \theta_{CRS} \\ H_1 : \mu \xrightarrow{n^k} \theta_{VRS} \end{cases}$$

The test statistic was $\tau = 0.3875$, with a p-value of 0.1281. The p-value was greater than 0.05, so we failed to reject the null hypothesis, suggesting that the **CRS** (Constant Returns to Scale) model should better fit the data compared to the **VRS** (Variable Returns to Scale) model. τ measures the impact of the difference between the two models (CRS and VRS). When its value is very small, the difference between the two approaches is negligible, indicating that the advantage of one model over the other is not substantial. Although CRS might appear statistically better, the small value of τ implies that the one model isn't statistically better than the other. Therefore, in the consequent analysis, we conducted both CRS and VRS DEA analysis to determine which one resulted in the best outcome.

4.2.4 Separability Test

Additionally, a third test was conducted to evaluate the separability assumption regarding the time variable. Separability implies that the considered environmental variables influence only the efficiency distribution without affecting production possibilities. A lack of separability undermines the meaningful interpretation of unconditional DEA estimators, potentially leading to inconsistent results within the analyzed dataset.

5 Analysis

As anticipated in the previous chapter, the first step of our analysis compares efficiency scores derived from DEA models of both constant returns to scales and variable returns to scale, as supplemented of the RTS test, to determine which model would best suit our dataset. The average efficiency scores under CRS and VRS were 1.3760 and 1.1830 respectively, indicating that scale inefficiencies play a significant role in this analysis. The VRS model provides a more flexible and realistic assessment of efficiency by accounting for scale effects, making it better suited for situations where Decision-Making Units (DMUs) operate at varying scales. Additionally, just 21% of DMUs were shown efficient under CRS against the better percentage obtained under VRS of almost 32% efficient DMUs. This smaller share of efficient cases can be justified by the flexibility of the model, since it also identifies a lot of DMUs to be locally efficient, but not conform to optimal operational conditions assumed under CRS. What we would like to demonstrate is that the applied CRS model seems to give stricter evaluation, since it severely punishes inefficient DMUs; while VRS instead is considered to be more accommodating in giving efficiency scores to weaker DMUs, thus reflecting its ability to accommodate different operational settings, variability and local efficiencies, making it useful in cases like ours where countries widely differ in dimension and scope.

In addition, the graph that illustrates the distribution of the efficiency scores of the DMUs under both DEA models *(Figure 7)* makes it possible to observe that the VRS model (represented by the blue bars) had an higher concentration of DMUs scoring 1.00 in efficiency, compared to the CRS model (represented by the red bar). However, in the range from 1.50 to 1.75, almost the same pattern of DMUs in both models could be observed, demonstrating that the comparison evaluated performances within that range. On the other hand, only a few DMUs can be found over efficiency levels of 2.00 in both models, indicating that most units hover just about fairly well. We concluded that the VRS would offer a more comprehensive and realistic evaluation of performance among DMUs.

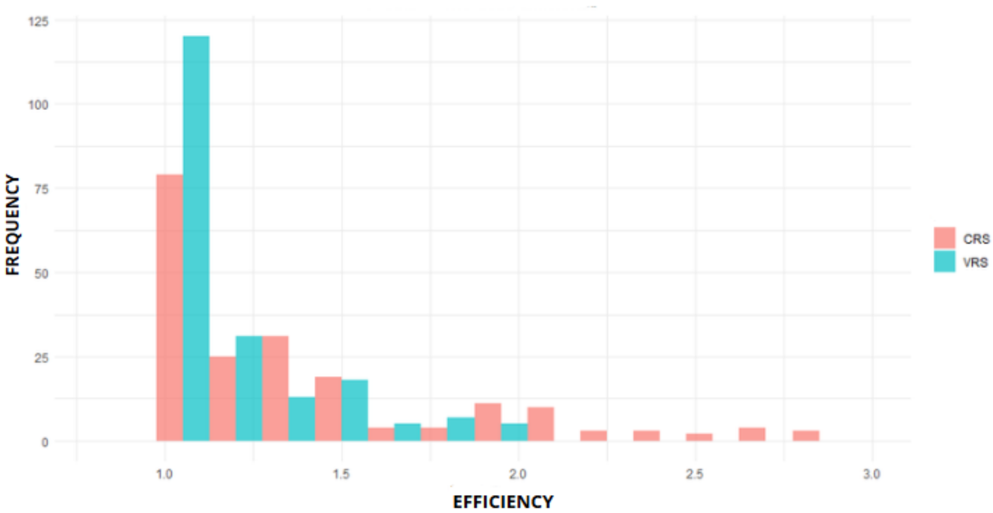

Figure 7: Efficiency comparison between CRS and VRS

5.1 Output-oriented VRS Model

The output-oriented approach, on this note, evaluates how much more outputs can be increased without altering the levels of inputs for the Decision Making Units. The method provides identification of inefficiencies and consequently shows possibilities for expanding outputs at fixed levels of inputs. Excluding the efficiency of one unit that reached infinity, using the VRS model we analyzed a total of 199 DMUs: the average efficiency score of the results stands at 1.18, indicating that inefficient DMUs can increase outputs by up to 18.2% at an average level if their input is maintained at their current level. Out of the efficient units, with a score of 1 defining the efficient frontier, a total of 63 DMUs (32% of the sample) acted as benchmarks for other units, and the distribution of efficiency scores indicates that 26% of DMUs attain scores within the interval of 1.0 to 1.1, thus most units function near the efficient frontier. Besides, none of the DMUs obtained efficiency scores greater than 2; this fact already implies no extreme inefficiency exists among the records. Overall, these results give a reasonably high level of efficiency as indeed several units have optimal performance, reflecting the balanced use of technological resources within the healthcare sector. However, inefficient DMUs with scores above 1 have quite clear potentials for improvement by increasing their outputs without needing additional inputs. The DEA thus points to numerous DMUs as making use of their resources well, with room for some units in excess of an efficiency score value over 1.2 to seek further improvement. For instance the application of the best practices from the efficient DMUs may yield more optimal resource allocation and better health outcomes. All these results are gathered in a table at *Figure 8*, presenting the distribution of the DMUs efficiencies.

Average Efficiency	E = 1	1 < E < 1.1	1 < E < 1.25	1.25 < E < 1.5	1.5 < E < 2	E > 2
1.18	32 %	26 %	18 %	13 %	10 %	0 %

Figure 8: Efficiency resume

From a dynamic perspective, the various dispersion of efficiency trends is illustrated in *Figure 9*, showing the average efficiency by country and year. Countries such as Norway, Luxembourg, and the Netherlands show promising trend lines in efficiency over the years, which suggests a relatively better use of technological resources in the healthcare sector. On the other hand, Greece and the Slovak Republic demonstrate good, but stable or slightly declining efficiencies, indicating areas for potential improvement. While no extreme inefficiencies are observed, the overall trend is positive, meaning many countries generally progress in healthcare efficiency, also due to the huge development in the tech sector over the last few years. This can also help us in deducing the best practices from countries that reflect positive trends, which could then be applied to the less efficient ones to get better scores in time. Besides, studying periods of stability or decline might point to external factors or policies affecting performance. Furthermore, it's crucial to highlight that Hungary's average efficiency is at a negative value in the first year of observation, 2010: this represents for us a reason to be careful with its management during the rest of the analysis because of the probable presence of particularly extreme observations in some years. When this issue will present itself again in the following work, we will take action in order to maintain a statistical robustness in the report.

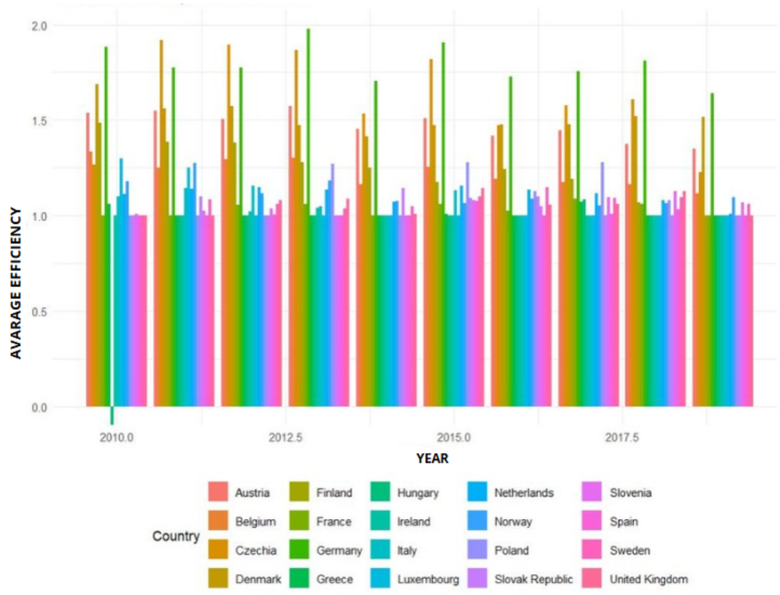

Figure 9: Average efficiency by country and year

At this point, before conducting any deeper analysis on our model, we decided to make it more robust by shifting from the simple DEA approach to the Bootstrap DEA, which allows us to filter out the bias from our results and therefore make more solid observations.

5.2 Bootstrap DEA

Since the outcomes of DEA are susceptible to **bias**, due to data variability, we needed to integrate the Bootstrap method in order to make the inferences more robust, solid and realistic: the Bootstrap consists of selecting multiple random samples from the original data set and then running DEA on each sampled set. The result is a distribution of efficiency scores that more accurately represents the said natural data variability, worsening indeed the efficiency values of every DMU due to its tendency to correct the bias estimation itself. This procedure produces multiple artificial frontiers from each of the resampled datasets, that act as an alternative benchmark for the evaluated subjects, offering many advantages in situations when, for instance, the samples are too small or when enhanced robustness of efficiency indicators is considered. However, random sampling introduces Bootstrap bias, in the form of systematic deviations. In order to deal with it, the central body of efficiency distribution has to be adjusted to provide more reliable and accurate estimates compared to those obtained by the conventional DEA method. Fortunately, the Bootstrap is able to introduce this correction iteratively recalculating efficiency scores across several samples so that the results are more stable and representative of true efficiency in the system we analyzed.

Practically, in our case, we applied this method using the FEAR library available on R to compare biased efficiency scores, obtained with standard DEA, with bias-corrected score, obtained by bootstrapping and observed how efficiency scores raised up from the value of 1. Despite this, in some cases results indicated the same exact value due to very already robust original scores or minimal correction of bias. The bootstrap results confirmed our initial hypothesis, formulated observing the traditional DEA's results: by observing the confidence intervals of each country in each year *(Figure 27)* and the bias distribution in *Figure 10*, it comes easy to understand why Hungary can no longer be part of the sample for the bootstrap DEA analysis. The significant variability of the observations' confidence intervals in some specific years, especially 2015, that we can attribute to the few extreme value visible in the bias distribution, led us to label Hungary as an outlier for our research and therefore remove it from the remaining analysis, illustrating results, tables and graphs without its contribution. Other than that, the other countries' confidence intervals remain reliable over the years.

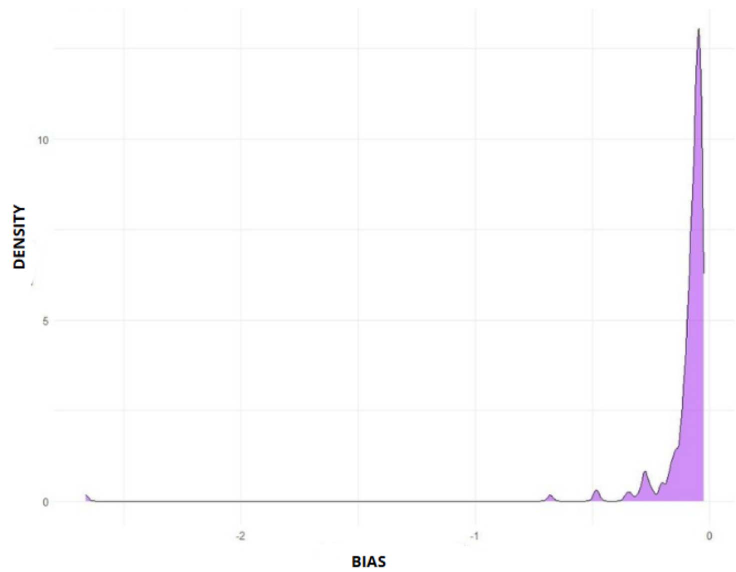

Figure 10: Bias distribution

The graph *Figure 11* demonstrates that, for most countries, the bias-corrected efficiency scores (depicted in red) are slightly higher than the original efficiency scores (depicted in blue) and the reason why has already been explained in the previous paragraph. However, the differences between the two measures remain relatively small, confirming that the original scores were already fairly robust. Still, it is helpful to note how some countries like Czechia, United Kingdom, Ireland, Poland, Spain and Greece show a more pronounced increase in bias-corrected scores, indicating even more inefficiency and that the original estimates for these countries were more influenced by bias. On the other hand, other countries such as Austria, Belgium, France, Italy, Netherlands, Slovenia and Sweden exhibit minimal differences between the two measures, suggesting that the original scores for these countries were less biased or well-calibrated at least. Eventually, by excluding Hungary from the comparison, we managed to have a more consistent perspective on the efficiency scores without distortion from extreme values, strengthening the validity of the analysis and highlighting the importance of dealing with outliers when interpreting data.

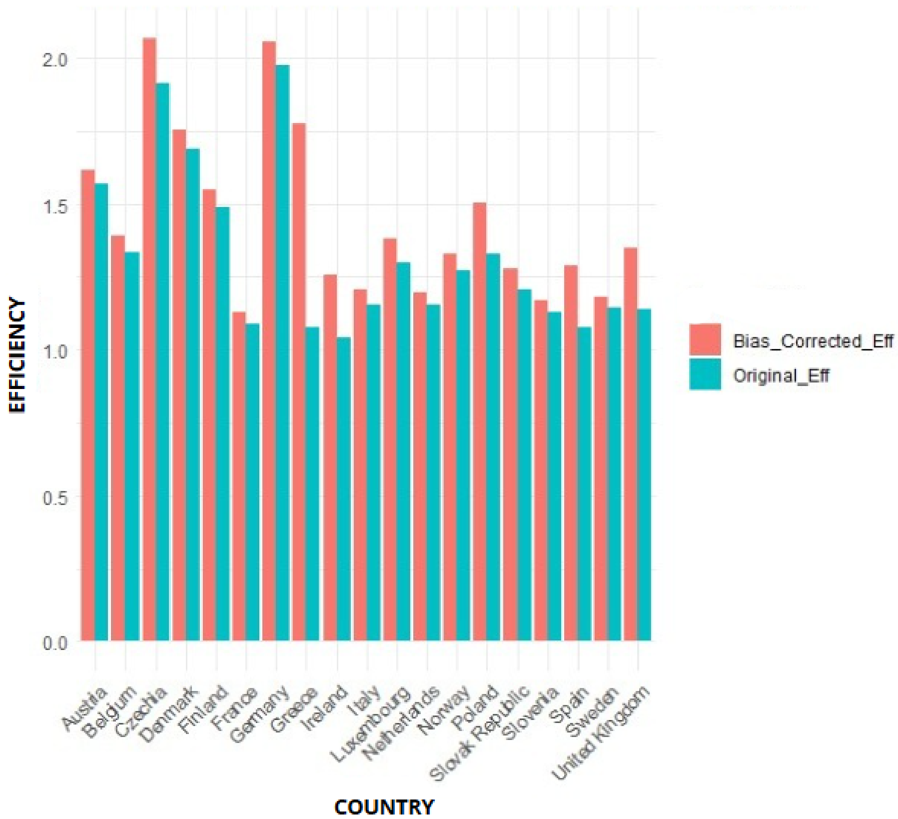

Figure 11: Comparison between biased efficiency and unbiased efficiency

Figure 12 shows the boxplot of the bias-corrected efficiency distribution in each country to compare variability and average values across DMUs. For instance, countries like Poland, Czechia, and Finland have a wide dispersion of values (as can be seen by the width of the boxplot or even the length of the whiskers), reflecting significant heterogeneity in DMUs performance. This means that while some units in those countries are highly efficient, others perform much less effectively. On the contrary, countries like Luxembourg, Ireland and Spain are in the category of the most efficient, with several DMUs reaching the efficient frontier (E=1), defining the optimal use of the resources involved. However, these countries are also among those presenting outliers, whose presence indicates some DMUs with extraordinarily high performances or some very low-efficient units as exceptional cases to be further deeply analyzed.

Additionally to this, the chart in *Figure 13* highlights a clear pattern where smaller countries, such as Luxembourg, tend to have efficiency distributions closer to the DEA frontier, while larger countries, like Germany, show significant room for improvement despite the notably stable economic and industrial systems, which would suggest a rich resourceful country. This distribution logically suggests the importance of tailoring strategies to each country's unique context to improve efficiency outcomes.

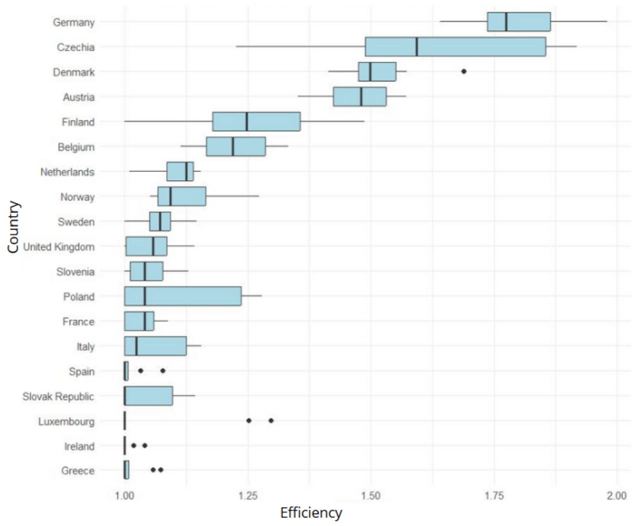

Figure 12: Boxplot fficiency distribution by country

The DEA frontier graph *(Figure 24 in the appendix)* represents the relationship between inputs and outputs for the DMUs and positions them relative to the efficient frontier. Many DMUs are located close to the frontier, indicating high unbiased efficiency, while some DMUs are positioned significantly below the frontier, highlighting inefficiencies that could be further analyzed. As a conclusion, the shape of the frontier reflects the VRS model used allowing greater flexibility in production technology.

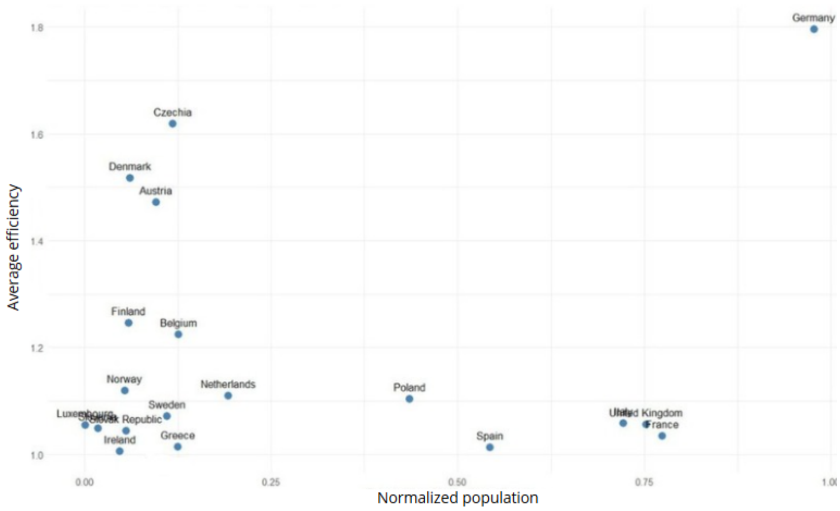

Figure 13: Relationship between average efficiency and normalized population by country

The *Figure 25* shows the efficiency trend in time by country while the bar chart at *Figure 26* puts into an histogram the average efficiencies by country computed with VRS *(both in the Appendix)*. Together, they highlight how some nations, like Czechia and Denmark, achieve higher inefficiency scores, suggesting badly optimized resource use, while others, such as Ireland and France, almost consistently perform at the efficiency highest levels, although presenting some slight negative trends in the latest years of the analysis. Moreover it also reveals stability for certain countries, like Slovenia, while others, such as Finland and Belgium, show respectively marked improvements or declines in specific periods, potentially reflecting policy changes or the presence of external factors.

Furthermore, looking at the map in *Figure 28 (in the appendix)*, a clear heterogeneity emerges in efficiency levels among the various European countries. Central regions appear to exhibit higher inefficiency (closer to yellow), whereas Southern regions and parts of Western Europe tend to display higher efficiency levels (mainly purple or darker shades). This distribution might or might not reflect differences in economic contexts, infrastructure, or regional policies influencing the operational efficiency; for instance Germany's situation regarding resources does not entirely mirror the results obtained, which can be explained by misallocation of said resources. Another interesting point is the concentration of intermediate levels (colors leaning towards red or orange) in certain transitional regions between countries with high and low efficiency. This could indicate a scenario where socio-economic or structural factors have a mixed impact on productive and managerial capacity.

5.3 SFA

Now, let's move on to a **parametric analysis** to try to compare with the efficiency levels identified earlier. Specifically, we employed a **Stochastic Frontier Analysis** (SFA) which is a parametric approach used to evaluate the efficiency of DMUs by estimating a production or cost frontier, that could be either the **Translog** or the **Cobb-Douglas** function. Unlike Data Envelopment Analysis (DEA), SFA accounts for statistical noise in the data, distinguishing between inefficiency and random error.

This method is widely used in contexts where the presence of stochastic noise is significant and could distort efficiency estimates. For example, SFA is valuable in analyzing data that may be influenced by random shocks, measurement errors, or other exogenous factors. As already introduced, the model assumes a specific functional form for the production frontier and separates deviations from the frontier into two components:

- **Inefficiency**: Systematic deviations reflecting suboptimal performance.
- **Random Noise**: Non-systematic deviations capturing stochastic variations.

In this study, we applied SFA to analyze the efficiency in the scope of our study and to identify possible noise in our data. Using normalized and logged data, we modeled separately our two outputs (health status and inverse mortality) as a function of the same five input variables introduced for the DEA analysis. The SFA model has been estimated using the frontier package in R, with the goal of evaluating how efficiently healthcare systems across countries transform their technological and financial resources into positive health outcomes.

For our model, we tested both the Cobb-Douglas and the Translog functions as our frontiers, to identify the most suitable for the analysis.

In our specific case, the result given by this SFA on R using the Translog function, highlighted that there's no significant interrelationship between the variables of our model. As a consequence, despite the low log likelihood (which describes the ability of the model to adapt to the data), we had to select the Cobb-Douglas as the most suitable for our model.

Function	Log likelihood valuea	Mean efficiency
Cobb-Douglas	-102,79	0,6891
Translog	38,92	0.8602

Figure 14: Comparison between SFA function with mort_inv as output

Figure 14 showcases the results of our analysis. The outcome of the SFA with the Cobb-Douglas confirmed that all the inputs had significant positive effects, indicating that increased resources generally improve healthcare outcomes. The only relevant difference with the DEA analysis lies in the input variable describing the "tech health expenditure", whose influence on the outputs turned out to be almost insignificant for our SFA. It's important to point out how it was still one of the least influencing variable in the DEA, but still held some significance, as assessed by *Figure 6*.

That being said, we noticed that for both of our models, Gamma presented a high value, exceeding 0.8 and suggesting that statistical noise is relatively low. Therefore, the non-parametric DEA approach is sufficient to effectively explain the resulting efficiency, proving to be overall more robust. This leads us to exclude the SFA, which, despite its advantage of distinguishing noise from inefficiency, does not add any significant value. To sum up, we can say that the results revealed minimal noise in the dataset, aligning with the findings from the DEA analysis, where efficiency scores were comparable across countries. As DEA is non-parametric and does not rely on restrictive assumptions about the production frontier, it offers a more flexible framework for analyzing multi-output systems like the one we are dealing with. As stated, SFA serves as a powerful tool in contexts with high stochastic variability, allowing researchers to disentangle inefficiency from noise. However, in this study, the low noise level in the data suggests that the DEA approach already adequately captures the efficiency of healthcare systems. The SFA results reinforce the robustness of the DEA findings, confirming that inefficiency is the primary source of deviation from optimal performance. Furthermore, the selected SFA function presented low log likelihood, strengthening the theory of the inadequacy that this analysis brings to our study. So, while SFA offers valuable validation, we concluded that the added value in this specific application is limited by both the small level of noise and the nature of our model, highlighting the better suitability of DEA.

5.4 Considerations on inputs efficiency

So back to the Bootstrap DEA results, we performed additional analysis to showcase relationships between the inputs selected and efficiency score, while always keeping in mind the **output-oriented** approach, in order to identify interesting trends about how any input alone can influence the final results. In *Figure 15*, where the linear combination of inputs was considered, the observed trend reveals a U-shaped relationship, where inefficiency tends to increase at both very low and very high input linear combinations, while mid-range input levels are associated with relatively higher efficiency. At the extremes of the input combination spectrum, inefficiency increases, suggesting two possible scenarios: at low input levels, insufficient resources may limit the ability of decision-making units (DMUs) to achieve optimal outputs; at high input levels, the inefficiency could result from diminishing returns or mismanagement of surplus resources. These findings are significant for resource allocation and policy-making, as they underscore the importance of identifying an optimal range of input combinations that minimizes inefficiency, also revealing the risks of operating at extremes where inefficiency is likely to rise either due to resource shortages or poor management of excess inputs. The relationship between all the input variables and the bias-corrected efficiency can be found in the *Appendix*, from *Figure 29* to *Figure 32*: they all resemble more or less accurately this U-shape trend, except for the health expenditure variable, whose mid to high values are usually connected to a higher level of efficiency.

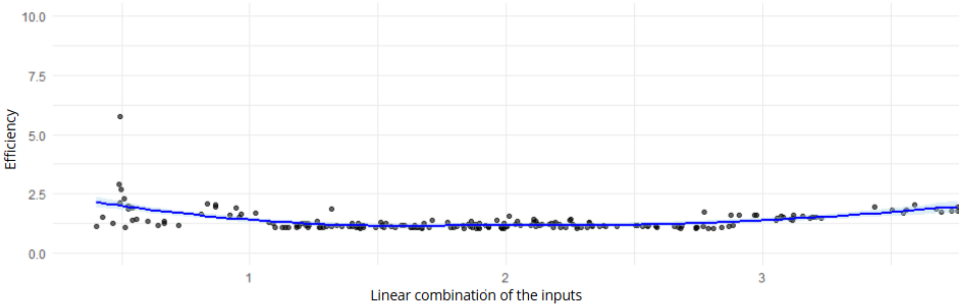

Figure 15: Inputs relationship with efficiency

As an addendum, *Figure 33* illustrates the relationship between technological expenditure per capita (which we consider to be one of the most significant input for the nature of our research) and efficiency: a general negative correlation emerges between technological spending and inefficiency since as per capita spending increases, inefficiency tends to decrease. However, a clear correlation between the two can't be drawn: countries like Germany, Austria, and France, representing the highest spending countries in such field, exhibit high levels of technological spending but differ in their inefficiency levels suggesting that higher investments can lead to varying outcomes depending on the context or the effectiveness of resource utilization. The same happens at the other end, with countries like Poland, Ireland and Czechia, all exhibit very low levels of technological spending but again a very different efficiency rate. Most countries with an intermediate expenditure such as Sweden, Spain, and Slovenia show low inefficiency representing an interesting benchmark for identifying efficient technological management strategies, strengthening the previously attained result that highlighted how the highest efficiency is given for intermediate values of inputs in general. In summary, the graph suggests that greater technological spending generally tends to reduce inefficiency although the results are not linear and strongly depend on each country's ability to manage technological resources effectively.

6 Conclusions

This report aimed at estimating efficiency in the healthcare sector, with a precise focus on the new practices based on emerging technologies such as Digital Twins. Unfortunately, the collection process of data relative to DT was not sufficiently conclusive due to the absence of enough data, given its sensitivity, to build a stable dataset. However, we overcame this problem by shifting our interest on efficiency in healthcare in a broader sense, collecting observations regarding numerous inputs that can be considered proxy to DT, such as technological equipment availability and more, able to represent the impact of technology in healthcare. Before drawing conclusions, we mentioned how the dataset ultimately included about ten years of observations for each variable, although we believe that repeating this same exact study in a few years in the future would notably benefit the results, giving better insights on the efficiency estimate due to the presence of more precise data. Given the results of DEA and Boostrap DEA in particular, we managed to observe how the relation between population and efficiency for smaller countries such as Luxembourg often presents excellent results due to their streamlined management and fewer complexities, while instead for larger countries like Germany the scores are worse, which is indicative of how scale can pose challenges in achieving optimal resource utilization. Other examples emerged in the analysis regarding the other variables, such as technological expanditure which points out how greater spending in this context generally tends to reduce inefficiency, although the results are not linear and strongly depend on each country's ability to manage technological resources effectively. The same thing can be said for GDP per capita and others, which keep on showing how having more resources does not always lead to be more effective. Keeping in mind this affirmation, we can recommend several general strategies to yield better results: first of all resource allocation should address regional disparities, which often challenge larger nations due to uneven resources distribution and complexity overall. Another important point can be made regarding AI-driven technologies such as DT, where data collection and usage, interdisciplinary collaboration between institutions and knowledge sharing among countries are essential in order to standardize and optimize relative medical applications, making it possible for inefficient countries to implement what efficient ones do and enhancing efficiency as a consequence. In order to do this, we strongly suggest strengthening policy and regulation to ensure data privacy and security, given the sensitivity of healthcare data, and we feel like policymakers should create incentives for innovation as grants and tax benefits to accelerate integration of such technologies. Finally, the introduction of continuous performance monitoring systems, supported by data-driven analyses (like DEA), is critical for identifying inefficiencies and enabling timely interventions ensuring that strategies are constantly updated based on observed results, fostering a dynamic and efficient management framework.

7 References

References

Ahmed, I., Ahmad, M., & Jeon, G. (2022). Integrating digital twins and deep learning for medical image analysis in the era of covid-19. *Virtual Reality & Intelligent Hardware*, *4*(4), 292–305.

Awan, K. A., Din, I. U., Almogren, A., & Rodrigues, J. J. (2024). Meditwin: A web 3.0-integrated digital twin for secure patient-centric healthcare in the metaverse. *IEEE Transactions on Consumer Electronics*.

Bjelland, Ø., Rasheed, B., Schaathun, H. G., Pedersen, M. D., Steinert, M., Hellevik, A. I., & Bye, R. T. (2022). Toward a digital twin for arthroscopic knee surgery: A systematic review. *IEEE Access*, *10*, 45029–45052.

Cellina, M., Cè, M., Alı, M., Irmici, G., Ibba, S., Caloro, E., Fazzini, D., Oliva, G., & Papa, S. (2023). Digital twins: The new frontier for personalized medicine? *Applied Sciences*, *13*(13), 7940.

Chen, J., Yi, C., Du, H., Niyato, D., Kang, J., Cai, J., & Shen, X. (2024). A revolution of personalized healthcare: Enabling human digital twin with mobile aigc. *IEEE Network*.

Chen, J., Wang, W., Fang, B., Liu, Y., Yu, K., Leung, V. C., & Hu, X. (2023). Digital twin empowered wireless healthcare monitoring for smart home. *IEEE Journal on Selected Areas in Communications*.

De Benedictis, A., Mazzocca, N., Somma, A., & Strigaro, C. (2022). Digital twins in healthcare: An architectural proposal and its application in a social distancing case study. *IEEE Journal of Biomedical and Health Informatics*, *27*(10), 5143–5154.

Geissler, F., Heiß, R., Kopp, M., Wiesmüller, M., Saake, B., Wuest, W., & May, M. S. (2021). Personalized computed tomography–automated estimation of height and weight of a simulated digital twin using a 3d camera and artificial intelligence. *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, *193*(04), 437–445.

Hu, H., & Zheng, X. (2024). Augmented and virtual reality-based cyber twin model for observing infants in intensive care: 6g for smart healthcare 4.0 by machine learning techniques. *Wireless Personal Communications*, 1–17.

Ilan, Y. (2023). Department of medicine 2040: Implementing a constrained disorder principle-based second-generation artificial intelligence system for improved patient outcomes in the department of internal medicine. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, *60*.

Khan, S., Ullah, S., Ullah, K., Almutairi, S., & Aftan, S. (2024). Implementing autonomous control in the digital-twins-based internet of robotic things for remote patient monitoring. *Sensors*, *24*(17), 5840.

Konopik, J., Wolf, L., & Schöffski, O. (2023). Digital twins for breast cancer treatment–an empirical study on stakeholders' perspectives on potentials and challenges. *Health and Technology*, *13*(6), 1003–1010.

Kulkarni, C., Quraishi, A., Raparthi, M., Shabaz, M., Khan, M. A., Varma, R. A., Keshta, I., Soni, M., & Byeon, H. (2024). Hybrid disease prediction approach leveraging digital twin and metaverse technologies for health consumer. *BMC Medical Informatics and Decision Making*, *24*(1), 92.

Li, L., Camps, J., Rodriguez, B., & Grau, V. (2024). Solving the inverse problem of electrocardiography for cardiac digital twins: A survey. *arXiv preprint arXiv:2406.11445*.

Manocha, A., Afaq, Y., & Bhatia, M. (2023). Digital twin-assisted blockchain - inspired irregular event analysis for eldercare. *Knowledge-Based Systems*, *260*, 110138.

Mohapatra, S., & Bose, S. (2020). An appraisal of literature for design and implementation of developing a framework for digital twin and validation through case studies. *Health and Technology*, *10*(5), 1229–1237.

Moztarzadeh, O., Jamshidi, M., Sargolzaei, S., Keikhaee, F., Jamshidi, A., Shadroo, S., & Hauer, L. (2023). Metaverse and medical diagnosis: A blockchain-based digital twinning approach based on mobilenetv2 algorithm for cervical vertebral maturation. *Diagnostics*, *13*(8), 1485.

Rowan, N. J. (2024). Digital technologies to unlock safe and sustainable opportunities for medical device and healthcare sectors with a focus on the combined use of digital twin and extended reality applications: A review. *Science of the Total Environment*, 171672.

Schwartz, S. M., Wildenhaus, K., Bucher, A., & Byrd, B. (2020). Digital twins and the emerging science of self: Implications for digital health experience design and "small" data. *Frontiers in Computer Science*, *2*, 31.

Shen, M.-d., Chen, S.-b., & Ding, X.-d. (2024). The effectiveness of digital twins in promoting precision health across the entire population: A systematic review. *NPJ Digital Medicine*, *7*(1), 145.

Sun, T., He, X., Song, X., Shu, L., & Li, Z. (2022). The digital twin in medicine: A key to the future of healthcare? *Frontiers in Medicine*, *9*, 907066.

Tao, K., Lei, J., & Huang, J. (2024). Physical integrated digital twin-based interaction mechanism of artificial intelligence rehabilitation robots combining visual cognition and motion control. *Wireless Personal Communications*, 1–16.

Vallée, A. (2023). Digital twin for healthcare systems. *Frontiers in Digital Health*, *5*, 1253050.

Veluvolu, K. C., Raman, R., et al. (2024). An insight in the future of healthcare: Integrating digital twin for personalized medicine. *Health and Technology*, 1–13.

Wu, Y., Wu, Y., Yang, R., Feng, M., & Pu, G. (2024). Cyber-physical wireless networks for smart health monitoring for elderly persons. *Wireless Personal Communications*, 1–22.

Xing, X., Del Ser, J., Wu, Y., Li, Y., Xia, J., Xu, L., Firmin, D., Gatehouse, P., & Yang, G. (2022). Hdl: Hybrid deep learning for the synthesis of myocardial velocity maps in digital twins for cardiac analysis. *IEEE Journal of Biomedical and Health Informatics*, *27*(10), 5134–5142.

8 Appendix

CIMO ELEMENT	DESCRIPTION
C (CONTEXT)	**Healthcare Application:** Digital twin are used in diagnostic, predictive analytics, operational optimization, personalized medicine and public health.
I (INTERVENTION)	**Digital model creation for patients and healthcare systems:** DTs simulate patient health, optimize results allocation, and integrate with IoT and AI technologies.
M (MECHANISM)	**Advanced Technologies and Predictive Analytics:** DTs use IoT for real-time monitoring, deep learning for diagnostics and predictive models for early intervention.
O (OUTCOME)	**Improved Patient Outcomes and Operational Efficiency:** Enhanced diagnostics, faster treatments, reduced costs for the healthcare providers and optimized resource use.

Figure 16: CIMO Framework for Digital Twin in healthcare system

(”digital twin” OR ”digital twinning” OR ”digital model”)

AND

(”healthcare” OR ”health system” OR ”hospital management” OR ”patient care” OR ”surgical”)

AND

(”efficiency” OR “effectiveness” OR ”optimization” OR ”cost reduction” OR ”resource allocation”)

Figure 17: Search Query (timestamp: 15/11/24 9:32 AM)

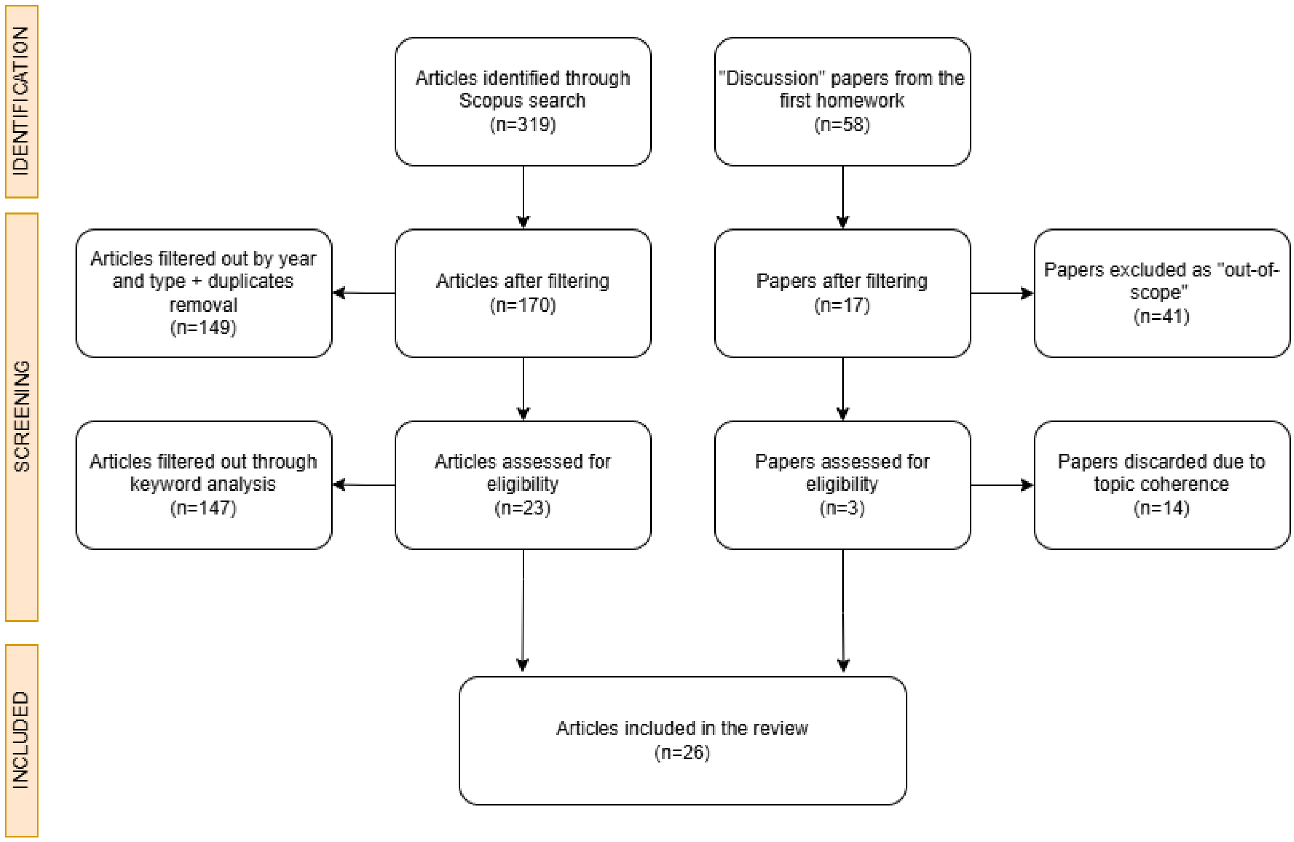

Figure 18: PRISMA

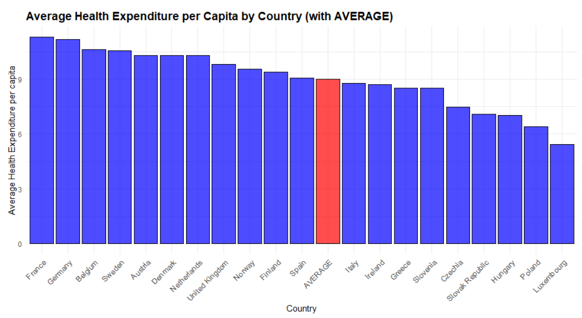

Figure 19: Average health expenditure per capita by country

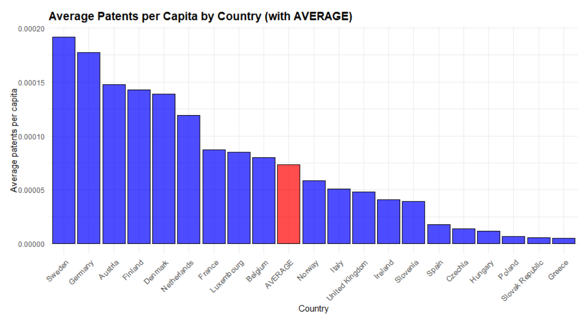

Figure 20: Average patents per capita by country

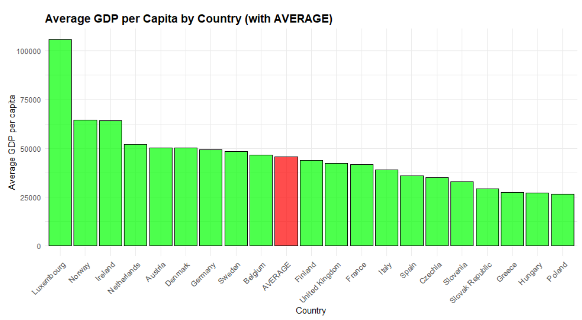

Figure 21: Average GDP per capita by country

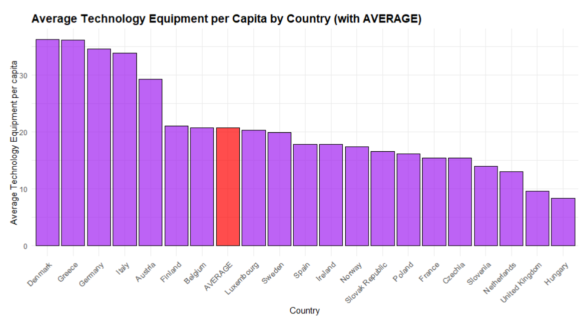

Figure 22: Average tech equipment available per capita by country

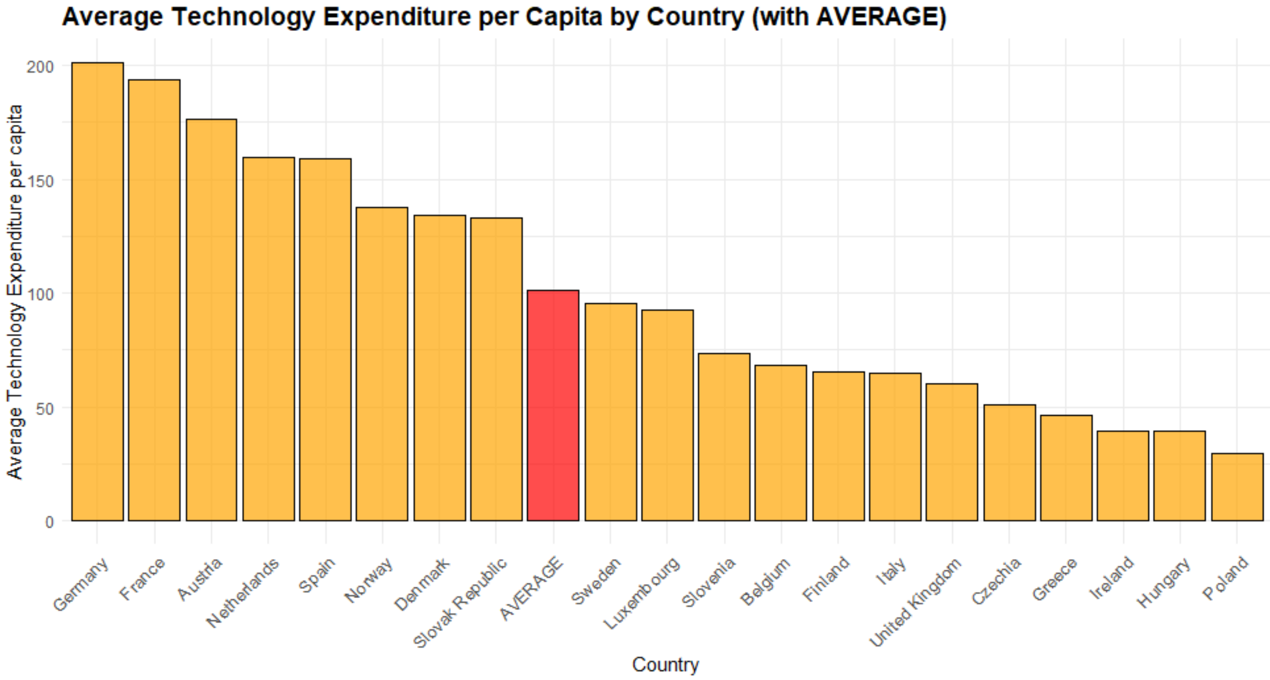

Figure 23: Average part of healthcare expenditure spent in technology by country

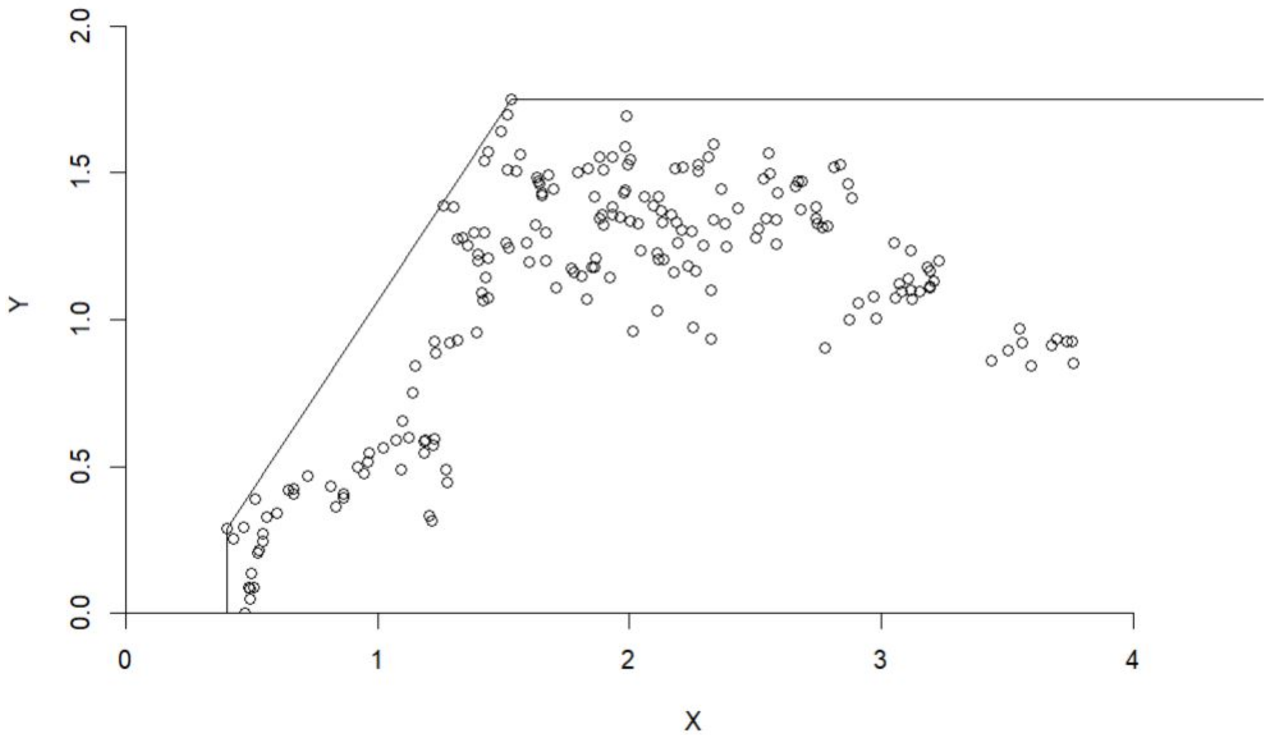

Figure 24: DEA frontier

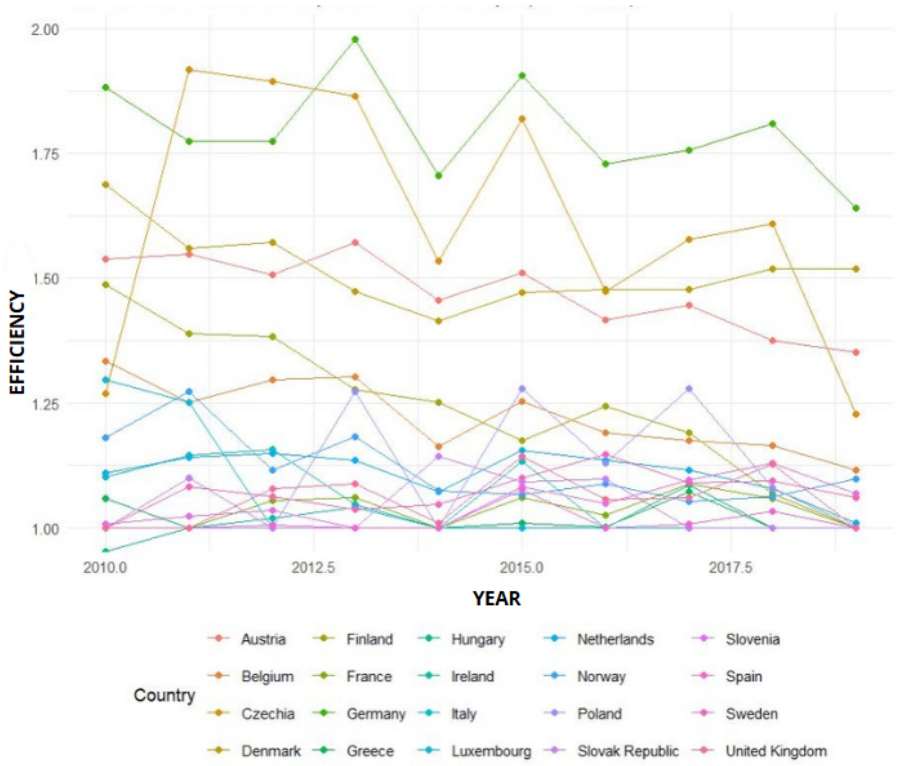

Figure 25: Average efficiency by country through the years

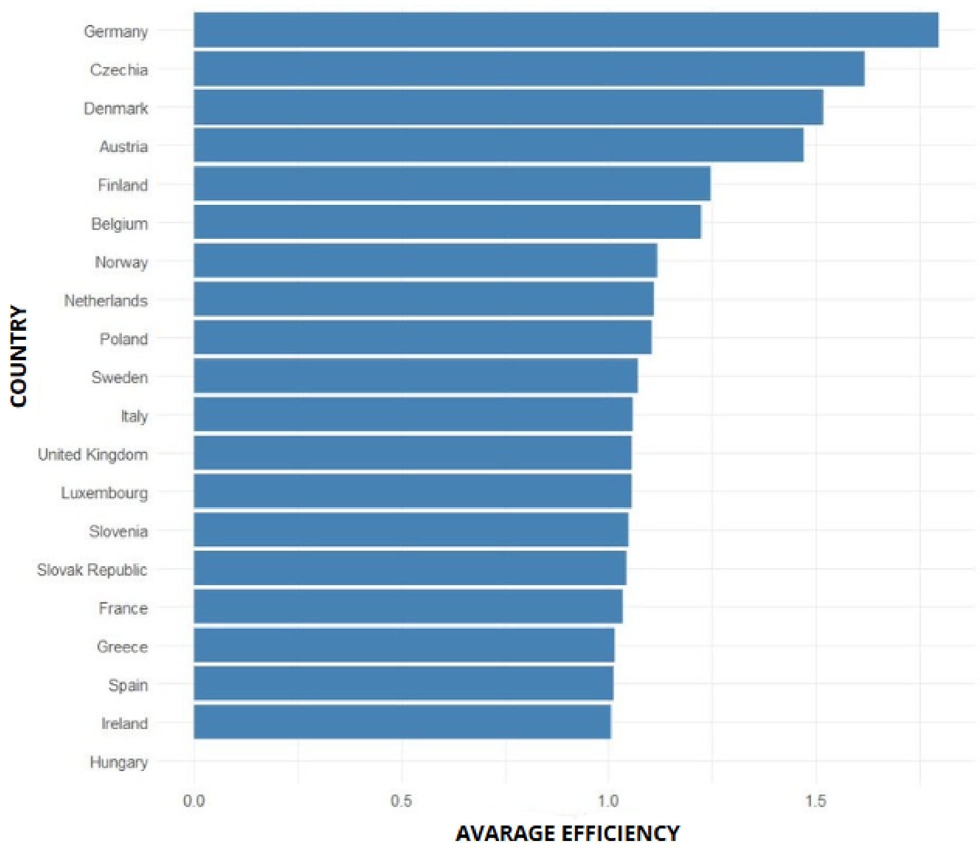

Figure 26: Average efficiency by country

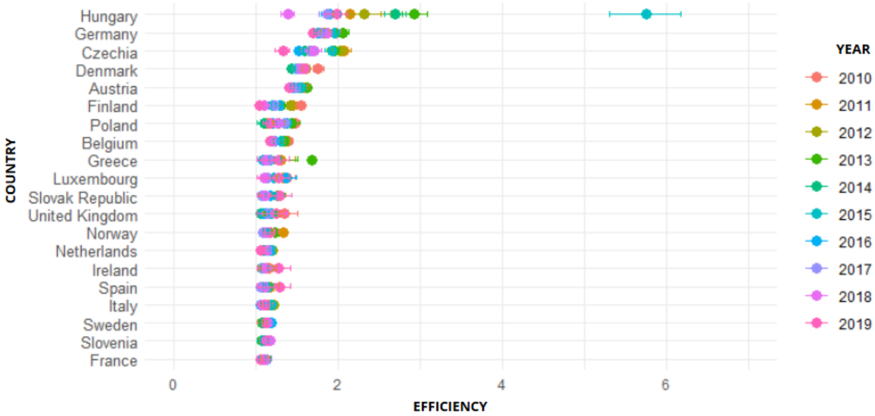

Figure 27: Confidence intervals of efficiency by country and years

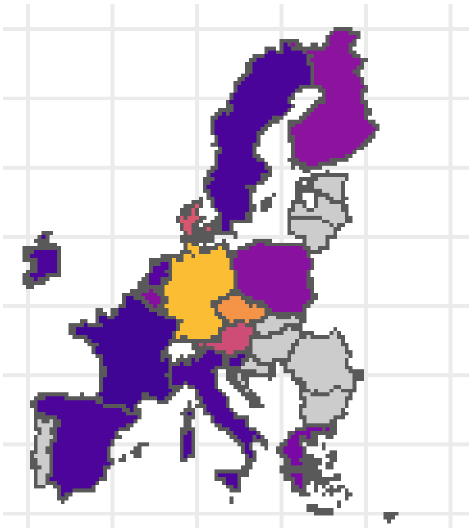

Figure 28: Average efficiency geographic distribution

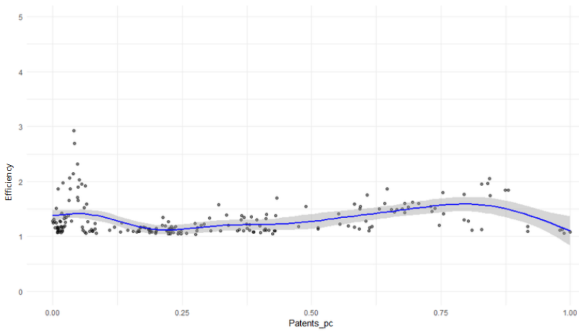

Figure 29: Patents relationship with efficiency

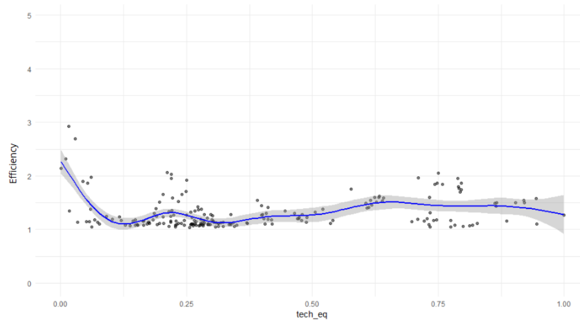

Figure 30: Tech equipment available in relationship with efficiency

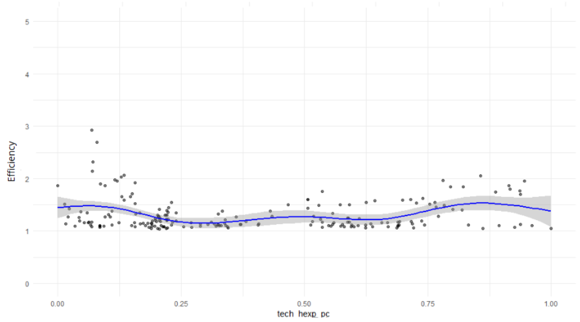

Figure 31: Expenditure in tech health sector in relationship with efficiency

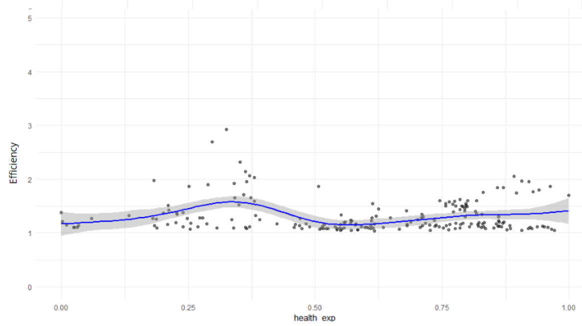

Figure 32: Health expenditure in relationship with efficiency

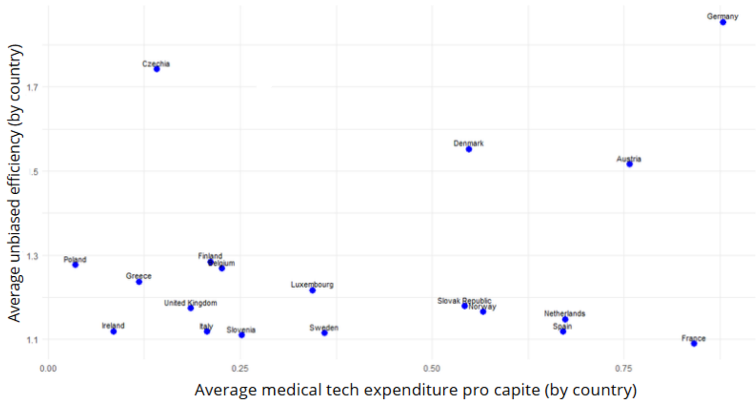

Figure 33: Relationship between tech health expenditure and average unbiased efficiency by country

Variables	Definition
Healthcare Technological Patents	The number of patents registered in the healthcare sector, reflecting innovation and technological advancement.
Total Health Expenditure	The total amount of financial resources spent on healthcare services and infrastructure within a given population.
Technological Health Appliances Expenditure	The specific expenditure dedicated to acquiring and maintaining advanced technological healthcare equipment.
Medical Technology Availability	The availability of a specific medical technology (used as a proxy) to gauge the level of access to modern medical tools.
Average Length of Stay (ALOS)	The average number of days patients spend in hospitals, indicative of the efficiency and resource utilization of healthcare systems.
Hospital Discharges	The total number of patients discharged from hospitals, representing the output of healthcare services.
Population	The total number of people in the region or country under analysis.
Population Over 65	The segment of the population aged 65 and older, representing a group with higher healthcare needs.
Healthcare Personnel (Active Physicians)	The number of practicing physicians per person, indicating the availability of medical professionals to meet population needs.
GDP (Gross Domestic Product) Per Capita	A measure of economic output per person, used to understand the financial capacity of a population to support healthcare.
Health Status	Data on perceived health status, measuring the percentage of the population aged 15 and over reporting their health as 'good/very good' (or excellent).
Mortality	The number of deaths per 100,000 people, used as an indicator of overall healthcare effectiveness and population health.

Table 3: Variables' Description

	Authors	*Title*	*Year*	*Method*	*Paper Goal*
1	**Ahmed et Al.**	**Integrating Digital Twins and Deep Learning for Medical Image Analysis in the Era of COVID-19**	**2022**	**Cascade recurrent convolutional neural network (RCNN)**	**To develop a digital twin-based healthcare system integrated with deep learning for medical image analysis, particularly for COVID-19 diagnosis using X-ray images**
2	**Awan et Al.**	**MediTwin: A Web 3.0-Integrated Digital Twin for Secure Patient-Centric Healthcare in the Metaverse**	**2024**	**Blockchain technology for secure data synchronization and cryptographic resilience**	**To design a robust integration of digital twins into the Metaverse, addressing challenges like data synchronization and security, and enabling patient-centric healthcare**
3	**Bjelland et Al.**	**Toward a Digital Twin for Arthroscopic Knee Surgery: A Systematic Review**	**2022**	**PRISMA protocol and DEA**	**To explore the potential of digital twins for improving surgical training and planning in arthroscopic knee surgery, using patient-specific models and intraoperative data**
4	**Cellina et Al.**	**Digital Twins: The New Frontier for Personalized Medicine**	**2023**	**Narrative review**	**To discuss the potential applications of digital twins in personalized medicine, including disease progression monitoring, treatment optimization, and medical education**
5	**Chen et Al.**	**Digital Twin Empowered Wireless Healthcare Monitoring for Smart Home**	**2023**	**Wireless healthcare monitoring using intelligent algorithms for fall detection and atrial fibrillation screening**	**To enhance healthcare monitoring in smart homes through real-time digital twin systems integrated with wearable devices and artificial intelligence**
6	**Chen et Al.**	**A Revolution of Personalized Healthcare: Enabling Human Digital Twin with Mobile AIGC**	**2024**	**Mobile artificial intelligence-generated content (AIGC) technologies**	**To enable real-time personalized healthcare through the development of human digital twins using mobile AIGC for applications like surgery planning and personalized medication**
7	**De Benedictis et Al.**	**Digital Twins in Healthcare: An Architectural Proposal and Its Application in a Social Distancing Case Study**	**2023**	**Architectural framework design and case study**	**To propose a generalized digital twin architecture and demonstrate its application for social distancing in workplace settings**
8	**Geissler et Al.**	**Personalized Computed Tomography – Automated Estimation of Height and Weight of a Simulated Digital Twin Using a 3D Camera and Artificial Intelligence**	**2021**	**Machine Learning with 3D cameras**	**To automate height and weight estimation using digital twin models for precision in computed tomography protocol design**
9	**Hu et Al.**	**Augmented and Virtual Reality-Based Cyber Twin Model for Observing Infants in Intensive Care**	**2024**	**Machine learning with convolutional probabilistic multilayer perceptron neural networks**	**To develop a cyber twin system for real-time monitoring of infants in ICUs using augmented and virtual reality**
10	**Ilan Y.**	**Department of Medicine 2040: Implementing a Constrained Disorder Principle-Based Second-Generation Artificial Intelligence System for Improved Patient Outcomes**	**2023**	**Constrained Disorder Principle (CDP) for enhanced digital twins**	**To use CDP-based digital twins to improve personalized treatments, diagnoses, and chronic therapy outcomes**

	Authors	Title	Year	Method	Paper Goal
11	Khan et Al.	Implementing Autonomous Control in the Digital-Twins-Based Internet of Robotic Things for Remote Patient Monitoring	2024	IoRT, VE and DTs	To facilitate remote patient monitoring using digital twins and autonomous robots for real-time health data collection
12	Konopil et Al.	Digital Twins for Breast Cancer Treatment – An Empirical Study on Stakeholders' Perspectives on Potentials and Challenges	2023	Qualitative content analysis based on semi-structured interviews	To assess the potentials and challenges of implementing digital twins in breast cancer care from stakeholders' perspectives
13	Kulkarni et Al.	Hybrid Disease Prediction Approach Leveraging Digital Twin and Metaverse Technologies for Health Consumer	2024	Denoising AutoEncoder with Broad Learning System (DAE-BLS)	To enhance robust disease prediction by combining DAE for feature extraction with BLS for incremental learning
14	Li et Al.	Solving the Inverse Problem of Electrocardiography for Cardiac Digital Twins: A Survey	2024	Deterministic and probabilistic computational modeling, ECG inverse problem-solving techniques	To improve the accuracy of cardiac digital twins by solving the inverse problem in electrocardiography, focusing on reconstructing cardiac sources and estimating electrophysiological parameters
15	Manocha et Al.	Digital Twin-Assisted Blockchain-Inspired Irregular Event Analysis for Eldercare	2023	Integration of IoT, Blockchain for data security, and deep learning for sequential data processing	To develop a framework combining digital twins and blockchain for secure, real-time monitoring of irregular events in eldercare
16	Mohapatra et Al.	An Appraisal of Literature for Design and Implementation of Developing a Framework for Digital Twin and Validation Through Case Studies	2020	Literature review and case study validation	To develop a framework for the design and implementation of digital twins, analyzing their potential applications across different domains
17	Moztarzadeh et Al.	Metaverse and Medical Diagnosis: A Blockchain-Based Digital Twinning Approach Based on MobileNetV2 Algorithm for Cervical Vertebral Maturation	2023	MobileNetV2 architecture within blockchain ecosystems	To develop a low-cost, blockchain-secured digital twin for dental diagnosis and treatment using MobileNetV2
18	Rowan N. J.	Digital Technologies to Unlock Safe and Sustainable Opportunities for Medical Device and Healthcare Sectors	2024	PRISMA review of literature; integration of digital twin and extended reality for medical device lifecycle optimization	To explore the use of digital twins in combination with extended reality to improve medical device safety, sustainability, and training practices
19	Schwartz et Al.	Digital Twins and the Emerging Science of Self: Implications for Digital Health Experience Design and "Small" Data	2020	N-of-1 evaluation	To leverage digital twins for personalized healthcare interventions and enable self-scientist approaches for end-users
20	Shen et Al.	The Effectiveness of Digital Twins in Promoting Precision Health Across the Entire Population: A Systematic Review	2024	Systematic review with Joanna Briggs Institute scales for quality evaluation	To assess the impact of digital twins on precision health outcomes at a population level

	Authors	Title	Year	Method	Paper Goal
21	Sun et Al.	The Digital Twin in Medicine: A Key to the Future of Healthcare	2022	IoT, Big Data, and AI	To evaluate the potential of digital twins in enabling precise diagnosis and personalized treatment in healthcare
22	Tao et Al.	Physical Integrated Digital Twin-Based Interaction Mechanism of AI Rehabilitation Robots	2024	Cyber-physical systems, visual cognition, and motion control	To design a virtual twin-based interaction mechanism for rehabilitation robots that adapts to patient needs in real-time
23	Vallèe A.	Digital Twin for Healthcare Systems	2023	Predictive analytics and machine learning integrated with digital twin technology	To enhance patient outcomes and operational efficiency in healthcare systems through digital twins
24	Veluvolu et Al.	An Insight in the Future of Healthcare: Integrating Digital Twin for Personalized Medicine	2024	Systematic literature review, analysis of real-time healthcare applications	To explore the integration of digital twins in personalized medicine, focusing on operational control, safety, and health management
25	Wu et Al.	Cyber-Physical Wireless Networks for Smart Health Monitoring for Elderly Persons	2024	Bi-Neural Network (Bi-NN), Attribute-Based Encryption (ABE)	To create a digital twin-based system for secure, real-time health monitoring of elderly individuals
26	Xing et Al.	HDL: Hybrid Deep Learning for the Synthesis of Myocardial Velocity Maps in Digital Twins for Cardiac Analysis	2021	Hybrid Deep Learning (HDL) using UNet and Generative Adversarial Networks (GANs)	To create synthetic 3Dir MVM data for digital twins to improve cardiac motion analysis and efficiency in clinical studies

Figure 34: Articles for the systematic review