International Journal of Mathematical, Engineering and Management Sciences
Vol. 10, No. 2, 300-349, 2025
https://doi.org/10.33889/IJMEMS.2025.10.2.017

Ram Arti
Publishers

Incorporating Contextual Factors into a Comprehensive Analysis of Operational Efficiency and Service Quality in Healthcare Sector

Utsav Pandey
Decision Sciences Area,
Indian Institute of Management Lucknow, Lucknow, 226013, Uttar Pradesh, India.
Corresponding author: utsav.pandey@iiml.ac.in

Sanjeet Singh
Decision Sciences Area,
Indian Institute of Management Lucknow, Lucknow, 226013, Uttar Pradesh, India.
E-mail: sanjeet@iiml.ac.in

(Received on August 16, 2024; Revised on November 12, 2024; Accepted on January 5, 2025)

Abstract
Service quality is believed to influence the productive efficiency of firms, particularly in a service focused industry such as healthcare. However, there is mixed evidence in the literature of both positive and negative correlation (e.g., the cost drivers of care providers vis-à-vis capacity expansion for better quality of service) between quality and efficiency. To address this challenge, a two-phase data-driven analysis is undertaken. In the first stage, an output-oriented Data Envelopment Analysis is employed to model the interdependency between operational efficiency and service quality by assessing the allocation of the input resources for achieving these two objectives. While accounting for the external influences and avoiding the 'best practice trap' in the healthcare sector, a set of classification algorithms are used to quantify the impact of external factors on efficiency levels. The proposed model is empirically tested using healthcare data of 31 provinces of China for a period from 2013 to 2018. The results show that the efficiency scores in operational productivity and quality of service are 67% and 64%, respectively. The major source of inefficiency is the number of cases in observation rooms (almost 47%) followed by health examination (22%). The provinces are categorized into three classes (optimally chosen number of clusters) using K-means clustering. The second phase of the analysis starts with selecting a subset of relevant features from 33 explanatory variables using information gain and correlation analysis. The proposed two-phased integrated technique enhances the performance of healthcare services and provides a roadmap for improvement for inefficient regions.

Keywords- Healthcare efficiency, Quality, DEA, Healthcare analytics.

1. Introduction

China's healthcare expenditure as a proportion of its total GDP remains at 6%, significantly lower than the US with 17%, its projected annual growth rate is 8.4%. Increased services provided per patient and rising cost per service are two primary reasons behind the increase in the healthcare costs, with their respective contributions of 4.3% and 2% (Eggleston et al., 2008). Furthermore, China's life expectancy has surged by 8.68 years over the past two decades (from 2000 to 2020), particularly among women, which saw an even greater increase of 10.16 years (Dwyer-Lindgren et al., 2017). Without efficient running of the hospitals, these costs are only expected to increase (Wagstaff et al., 2009). Besides, concerns for the quality of healthcare in China were raised persistently over the years (Eggleston et al., 2008). In an increasingly customer-driven market, there is a shift in the definition of quality which is now defined by what patients experience. However, evidence suggests that due to the absence of a strong monitoring system, the quality standards in China have not progressed as expected (Eggleston et al., 2008). The aging population will significantly increase the demand for high-quality long-term services (see **Figure 1**). Quality is primarily dynamic and multifaceted; interdependency on factors such as accessibility of healthcare, quality of life, etc., makes it difficult to measure precisely. There is no consensus on the common set of measures to use and how to interpret them from different perspectives of the patients, physicians, hospital authorities, and policymakers. Therefore, it is crucial to identifying the desired attributes of high-quality service and then using those attributes as quantifiable measures. The primary challenge of healthcare sector at a global scale is to balance the cost of providing quality service and cost of operational efficiency. To achieve this, Data Envelopment Model (DEA) is employed to handle multiple inputs and distinct sets of multiple outputs pertaining to both the perspectives: quality and efficiency.

The significant challenge for the policymakers is to improve the efficiency of the institutions to contain the cost without compromising on the standard quality of the service provided (Yang and Zeng, 2014). However, limited resources inevitably cause a trade-off between quality and quantity (Newhouse, 1970). Though China successfully established a national-level health scheme to cover a massive size, i.e., 90% of its total 1.8 billion population, the implementation, service, and costs depend heavily on the regional monitoring body. As a result, different barriers, including overall inflation, inaccessibility to healthcare, and disparities between urban and rural areas, have emerged (Cai et al., 2017). The measurements also vary disproportionately when larger units are in question or the level of analysis is at macro-levels. Since not all micro-level measurements can be aggregated to a larger capacity, e.g., from region to national level (Ozcan, 2014), we need macro-level proxies for the same set of measurements. Further, the regional administrations are responsible for providing the required healthcare needs, yet there is variability among their services. In light of these findings, the DEA framework treats the regional health sectors as individual decision-making units (DMU).

As health care managers face challenges from government regulations, insurance policy and global competition, adaptation of new technology and information or managing data centers, allocation of resources and performance evaluation become key elements of strategic decision making. The two most important productivity indicators used in the health care system are inpatient and outpatient counts (Ersoy et al., 1997). Therefore, as a cumulative measure, the total number of patients discharged is considered as output. In a dynamic system such as health care, there are multiple by-products which are sometimes undesirable. Too much focus on improving the desirable outcome (e.g., patients discharged) might lead to unintended consequences like increased fatality rates. Previous research identifies three primary forms of inputs: financial (e.g., operating cost), human-resources (i.e., doctor, nurses, technical staff), and equipment (e.g., general bed, ICU bed) (Ferreira et al., 2018).

This paper, therefore, investigates the effects of various externalities on the outcome of the healthcare sector while explaining the in-class variation of the efficiency scores of the regions within a particular class. We make a simplistic assumption of capturing the externalities via five overarching themes and each of the themes contain measurable factors (**Figure 2**). Considering the challenges mentioned above, we have used DEA framework, that treats the regional health sectors as individual decision-making units (DMU) in the DEA framework. We include the inpatient fatality and observation-room fatality rates as bad outputs. Also, the number of practicing doctors and the number of hospital beds act as inputs. This framework allows the units under evaluation to freely choose the best set of weights according to their performance. Thus, these weights reflect the relative strengths of various measures pertaining to the aspects of quality and productivity. The integrated efficiency can be decomposed into quality and productive efficiency. Besides, unidimensional strategies for optimizing quality and operational efficiency individually also exist side-by-side with the overall approach which would help us compare the result of overall approach compared to unilateral approaches.

However, the measurable factors cannot be included in the DEA model for two reasons. First, they are not directly controlled by the healthcare policymakers. Second, these factors influence both the inputs and outputs and thus cannot be segregated into any one of the two set in the DEA framework. For example, education influences the number of doctors in a society, and it also helps improve general awareness in people resulting in higher number of health examination. Other factors can also be shown to have significant impact on the inputs and outputs in different degree. Economic resources have been shown to have a positive effect on the health status as affluent nations and societies have access to better medical care. Disposable income levels of the citizens in each of the Chinese regions are used as a proxy variable to capture economic prosperity, the first externality. In the second phase of the analysis, a classification model is developed to understand the interrelationship between contextual environmental factors and the efficiency levels. Initially, the efficiency scores are used to create the optimal number of classes for each window. In earlier studies (Tsolas et al., 2020) all the regions are categorized under pre-defined classes based on their efficiencies.

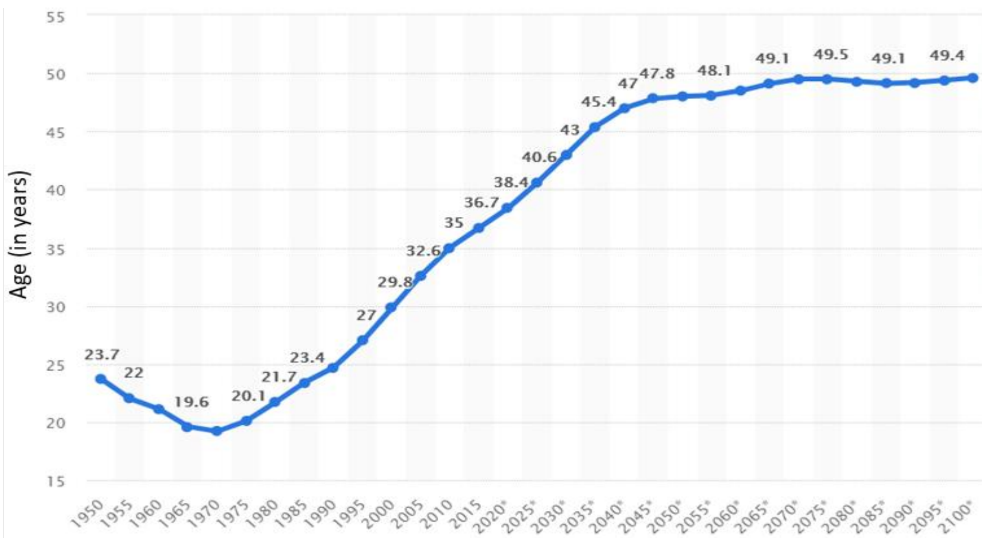


Figure 1. Median age of Chinese population (Source: Statista).

This study makes a significant contribution to the existing literature in both theoretical aspects and practical decision-making by demonstrating the interrelationship between quality and efficiency. The methodological contributions are: First, to the best of our knowledge, this is the first attempt to examine the interrelationship between quality and efficiency using an integrated-DEA framework without assuming any pre-defined substitutable relationship between the two aspects being examined. Second, there are individual measurements for the best possible quality and efficiency scores. When these best-possible values are compared with the corresponding values of quality and efficiency achieved under overall approach, one can assess the trade-off analysis using two indicators: cost-of-efficiency and cost-of-quality. Third, to account for the various socio-economic factors determining the advancement in quality and efficiency, a set of machine learning techniques are employed to evaluate importance of the external variables. Furthermore, there are some important insights for the policy-makers: First, the results show that quality is single-handedly responsible for almost three fourths of the overall inefficiency. The two main reasons for poor quality of service are cases in observation room and health exams. Second, the indicators of operational efficiencies show no sign of deterioration. A remarkable improvement can be observed in fatality in observation rooms which has reduced from 31% to 6%. Third, a key insight from the regional analysis is that in most of the factors (input or output) only a handful of regions are responsible for the bulk of the inefficient practices. In line with previous studies, this study finds a huge disparity between efficient regions and inefficient regions while evaluating productivity vis-à-vis quality. The efficient regions are largely efficient in both measures. On the other hand, inefficient regions experience disproportionate inefficiency and uneven emphasis on both the aspects. The reason for high mortality among the inefficient states can be traced back to incompetency in the quality standards. Fourth, the classification analysis reveals that medical factors (visit, medical technology personnel), socio economic factors like Guaranteed livelihood, and unemployment rate are the three major contributing factors in determining the category of the regions. Furthermore, it is noteworthy that the top-most determining factors seldom reappear across the windows. This trend indicates that there is remarkable variance in the root cause of inefficiency, suggesting a dynamic landscape of influential forces that shape regional efficiency over time.

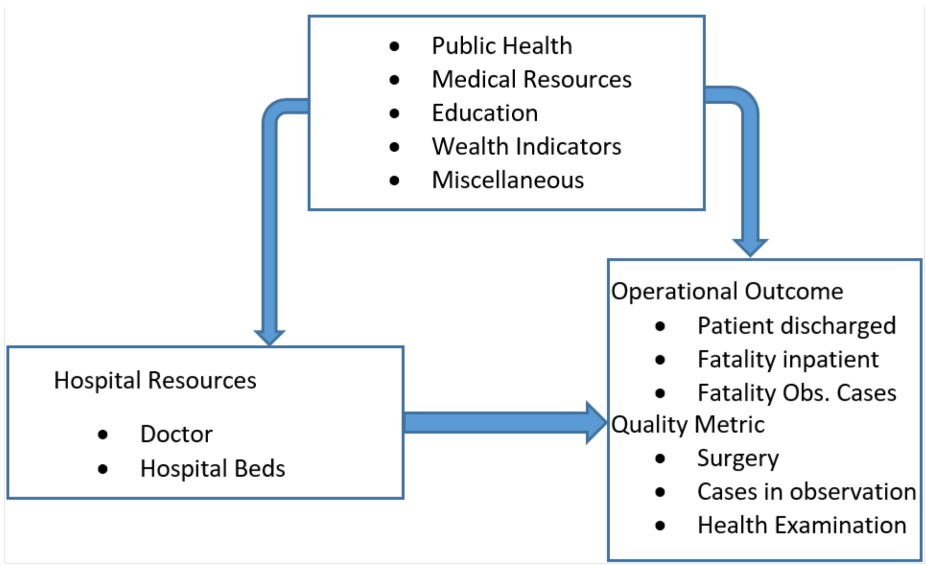


Figure 2. Interaction of the external variables with the designated inputs and outputs.

The remainder of the paper is organized as follows. Section 2 reviews the prior relevant literature. Section 3 presents the conceptual framework of the problem in question and develops an appropriate DEA model and extends it. The following section reports the empirical analysis and discusses research findings. Section 5 and 6 discuss the implications and conclusions of the study.

2. Literature Review

With the advancement of computing devices and software technologies, various quantitative techniques have been widely applied to simulate, study, and improve the medical system. Queuing theory drives this stream of research, focusing on the developing new rules for inpatient and outpatient services (Smith and Warner, 1971). Different variations particularly focus on scheduling and planning activities, such as addressing no-show patients that impact appointment scheduling (LaGanga and Lawrence, 2012), variations in ancillary tasks affecting patient waiting-time and resource utilization (Salzarulo et al., 2011), emergency blocking in a multistage treatment procedure with patient routing (Bretthauer et al., 2011), and capacity constraints (Chen et al., 2020), among others. The primary methodology adopted in these studies is simulation-based analysis (Badri and Hollingsworth, 1993) or heuristics (LaGanga and Lawrence, 2012). Recently, the majority of the studies in healthcare embraced machine learning (Priya and Kumar, 2015), data mining (Kusiak, 2006), or AI-driven decision making (Pan et al., 2019). Hybrid techniques, like iterative principal component analysis with a hybrid fast correlation-based filter (HFBF), are also employed to predict disease detection and progression over machine learning or decision forest tree (Priya and Kumar, 2015). As the industry faces challenges such as financial (Kaplan and Porter, 2011), cost management (Finkler and Wirtschafter, 1993), regulatory (Hotchkiss and Jacobalis, 1999), and high demand (Sanmartin et al., 2002), an objective assessment of healthcare sector is necessary for informed decision-making. In the management literature, regular operational activities and the potential sources of deficiencies are explored at length (Geng et al., 2019) also in light of lean performance (Henrique et al., 2021).

With the development of the concepts such as productivity and technical efficiency frontier (Farrell, 1957), research in DEA has witnessed a wide-spread applications in various areas. In the earlier phases, DEA was mainly used on public institutions like bank (Sherman and Gold, 1985), educational institutions (Singh and Ranjan, 2018), library (Pandey and Singh, 2022), agriculture (Pandey and Singh, 2021), and R&D activities (Feng et al., 2021) etc. Besides, it has been used very effectively on some unconventional fields of research as well. Further, regional studies comparing the innovation system in Russia and key factors contributing to its success has been reviewed and considered (Zemtsov and Kotsemir, 2019). Chen et al. (2022) studied the trend of research works in social science area in China using Malmquist index. Gender equality in the field of science and technology in Spain is evaluated and corresponding ranking is completed using meta-frontier analysis (Puertas et al., 2023). Role of social and scientific resources and the critical barriers in cross-border knowledge diffusion have also been examined (Ho and Liu, 2013).

Health economists have employed the frontier analysis using both parametric and non-parametric methods (Hollingsworth, 2003). Comprehensive reviews of existing literature on the productivity of healthcare institutions using non-parametric methods have been conducted (Hollingsworth, 2008; Kohl et al., 2019). Stochastic frontier analysis (SFA) has been extensively utilized in the healthcare sector (Hollingsworth et al., 1999; Jacobs, 2001). The literature on healthcare productivity and efficiency has seen a rapid increase, with 55% of studies published after 2000 (up to 2006). DEA is the framework used in 67% of these studies in some form as a primary or secondary analysis tool (Hollingsworth, 2008). Among these studies, 52% have treated hospitals as DMUs, and only a few (4%) have analyzed country-wise evaluations with thirty or more countries constituting the dataset. Traditional DEA models have been extended, for example, through window analysis (Co and Chew, 1997), and supplemented with further statistical analysis, such as Tobit regression (Nayar et al., 2013).

Trade-offs and compromises among organizational capabilities are pervasive in every echelon of industry and public life (Porter, 2011). Operations managers in the manufacturing sector grapple with this challenge at various levels, incorporating them into their strategic planning to retain a competitive advantage (Schmenner and Swink, 1998). However, several research studies (Talluri et al., 2013) suggest that these contrasting capabilities could sometimes converge to form a compatible competency up to certain levels.

Nevertheless, the prevailing belief is that service quality and operational efficiency, as two aspects in the service industry, often behave as "zero-sum" trade-off, where gains in one aspect necessitates losses in the other (Sarmiento and Shukla, 2011). Recent studies have shown that service quality to be essential for enriching customer satisfaction (Sampson and Froehle, 2006) and attracting new customers (Campbell and Frei, 2010). These studies also highlight the need for additional resources, capital, and labor to provide high-quality service.

In a general view, high-quality services also increase costs, limiting the accessibility of medical care and consequently restricting healthcare accessibility. A focus on economic efficiency leads to a reduction in the cost of operations by minimizing wastes in production, sometimes entailing the adoption of new production principles (Brohman et al., 2009) such as automation or lean management. However, technology-driven services can result in lower customer satisfaction and perceived quality levels (Campbell and Frei, 2010).

Thus, complementarity and compromises are identified as characterizing the relationship between efficiency and quality.

Decision-makers must fulfill efficiency and effectiveness requirements while navigating resource allocation (Hill et al., 2007), location, and capacity constraints (Smith-Daniels et al., 1988), among other factors. Additionally, decision control mechanisms in hospitals are responsible for the evidence-based and experiential quality of healthcare (Senot et al., 2016). The quality of medical care delivery and efficiency of the medical units also vary across states or regions (Jha et al., 2005). Therefore, achieving maximum efficiency to contain the costs while providing high-quality service is the most practical dichotomy in both academia and industry (Delellis and Ozcan, 2013).

The performing resources in healthcare facilities are medicines, equipment, bed staffing, room staffing, sizing, and technical and non-technical staff (Fetter and Thompson, 1965). A large body of literature has been developed, addressing each of these areas in minute details. An overview and summary of the literature reviews are also well-documented (Fone et al., 2003; Gholami-Zanjani et al., 2018). Although there is no comprehensive index for measuring quality in healthcare, existing measures can be broadly categorized into two categories: technical and interpersonal (Navarro-Espigares and Torres, 2011).

Recently, there has been an abundance of healthcare information generated from different wearable health devices and sensors in monitoring equipment (Kamble et al., 2019). Big data analytics can provide the much-needed competence to consolidate the information and support an evidence-based decision support system with predictive capabilities (Wang et al., 2018; Attaran and Gunasekaran, 2019). However, the success of such information management systems and the Internet of Things (IoT) depends primarily on the adaptation of it and the user's ability to effectively use it (Martínez-Caro et al., 2018).

Online communities also generate a volume of information regarding patients' concerns, feedback, and sentiments (Shirazi et al., 2021). In general, healthcare 4.0 deals with the technological and social-cultural aspects to attain a more customer-centric and resilient healthcare system (Tortorella et al., 2021; Ramadevi et al., 2016). The need for such a resilient health sector with capabilities for anticipation, adaptation, and recovery is even more evident in the challenges posed by the COVID-19 pandemic (Belhadi et al., 2021).

Only a few selected indicators are chosen in this study in consultation with US national (Rockville, 2015). The most important in the list of attributes is accessibility to health services or the infrastructural support which is captured using the number of health examinations performed. Since specialty or super-specialty hospitals provide treatment for only specific diseases, disease-specific weights and the weighted counts are not considered here. Besides, limiting the region-wise analysis to specific diseases would fail to capture the proficiency of the medical units in countering all types of diseases and illnesses. Relevance, as another crucial factor is concerned with whether the provided medical service was actually required or not. This measure also captures the safety precaution taken by the medical staff. Inpatients are kept under observation after any given medical treatment. Therefore, the number of cases in observation rooms indicates the alacrity of the medical practices. Lastly, the effectiveness of the treatment of illness is reflected in the number of surgeries performed in a hospital.

3. Conceptual Framework

Data Envelopment Analysis (DEA) is a linear programming technique that derives the relative productivity of each of the Decision-Making Units (DMUs) in a set of DMUs (Charnes et al., 1978). Suppose there are N homogeneous DMUs that consume m inputs to produce s desirable outputs and t bad outputs. The quality of the service produced is measured using K metrics. The ith input, rth output, and pth bad output of

DMU n are denoted by x_i^n, y_r^n, and b_p^n , respectively. The classical definition of efficiency, proposed by Farrell (1957), is the maximum proportional reduction (or augmentation) in the inputs (or outputs) while keeping the amount of outputs (or inputs) constant at the current level. Since all the inputs (or outputs) are contracted (or expanded) in the same proportion, the resulting efficiency is called radial efficiency. For inputs, the radial efficiency is denoted by θ in: $\sum_{n=1}^{N} \lambda_n x_i^n \leq \theta x_i^o$ where λ_n signifies the weighted contribution of n^{th} DMU. This radial efficiency can be categorized into two types: weak and strong efficiency. However, the radial measure of efficiency overlooks the non-radial portion of inefficiency. For example, some inputs of a DMU may be contracted more than the common proportion in which every other inputs are contracted. In other words, if for f^{th} input the constraint holds at inequality: $\sum_{n=1}^{N} \lambda_n x_f^n < \theta x_f^o$. That means this input can be contracted even more, given by the value: $\theta x_f^o - \sum_{n=1}^{N} \lambda_n x_f^n$. Such DMUs are called weakly efficient if the radial proportion is one. But there are still inefficiencies in those DMUs. Strongly efficient DMUs are free from such inefficiencies. In this case the input constraints hold at equality: $\theta x_f^o = \sum_{n=1}^{N} \lambda_n x_f^n$. In other words, there is no opportunity of any marginal improvement in any inputs since $\theta x_f^o - \sum_{n=1}^{N} \lambda_n x_f^n = 0$. To identify both strong and weak inefficiencies, additive models (Ali and Seiford, 1990) or slack based models (Tone, 2011) are proposed. The convex hull created by this framework enclosing the input and outputs values of the DMUs marks the efficient production frontier. Thus, inefficient DMUs can reduce input consumption or augment output productions accordingly to reach the nearest point on the efficiency frontier. The primary advantage of this framework is that it derives the efficiency frontier empirically from the data sets without any presumption about the shape of the production function (Banker et al., 1984). There is an alternative definition of efficiency, in the presence of multiple inputs and outputs, based on relative productivity as the ratio of the weighted sum of the outputs to the weighted sum of inputs (Charnes and Cooper, 1984).

A popular extension of DEA model is to consider multiple stages functioning in series to convert inputs into outputs. In a two-stage DEA model, the output of the first stage functions as inputs to the second stage. In this use-case of healthcare sectors in China, the inputs are Hospital beds and Doctors. If one considers operational efficiency to be the 1st stage, then the stage-1 outputs e.g., Patient Discharge, Death in Observation-room, and Death in Inpatient should be the input to stage-2 for measuring service quality which is captured by three indicators: Cases in Observation, Surgical Operation, Health Exam. However, this grossly mis-represent the entire process of healthcare management. Secondly, one must consider that the three outputs of operational efficiency as well as the three indicators of service quality all compete for the same set of input resources e.g., Hospital beds and Doctors. Thus, a stage-wise DEA analysis does not fit in the actual process of healthcare delivery system.

The proposed models can be adopted individually to measure performance in individual aspects of productivity-oriented and quality-oriented approaches. The overall approach developed in this study integrates these two aspects and produces a single synthetic measure focused on the overall performance assessment. In DEA, time-varying data typically calls for window analysis to identify performance changes and frontier shifts (Asmild et al., 2004). Similar to a moving average, this analysis recognizes trends in performance over a fixed time period and examines the consistency of these performance trends.

3.1 Proposed Methodology

Quality metrics are denoted by q_k^n. Under the assumption of constant return-to-scale assumption (CRS), the pure-operational efficiency of DMU 'O' can be expressed as follows:

$$\begin{aligned}
&\text{Max } \pi_o^1 = \quad \textstyle\sum_{r=1}^{s} \frac{c_r}{y_r^o} + \sum_{p=1}^{t} \frac{c_p}{b_p^o} \\
&\text{subject to} \\
&\textstyle\sum_{n=1}^{N} \lambda_n x_i^n \leq x_i^o, \forall i \in m, \\
&\textstyle\sum_{n=1}^{N} \lambda_n y_r^n = y_r^o + c_r, \forall r = 1 \dots s, \\
&\textstyle\sum_{n=1}^{N} \lambda_n b_p^n = b_p^o - c_p, \forall p = 1 \dots t, \\
&\lambda_n \geq 0, \forall n = 1 \dots N.
\end{aligned} \tag{1}$$

To obtain the BCC or VRS equivalent, the constraint $\sum_{n=1}^{N} \lambda_n = 1$ for convexity principle needs to be added to model (1). The intensity variable $\lambda_n, \forall n = 1 \dots N$ denotes the contribution of nth DMU in the construction of the frontier for the DMU under evaluation. The fraction of excess bad outputs c_p and shortfalls of the desirable outputs c_r are maximized parallelly. Both the output types are present in the model (1) and it reflects that the limited amount of input resources should be utilized to improve both the types of the outputs. Let the DEA model (1) be termed as DEA for operational efficiency, and the operational efficiency score be denoted by θ_o^1 for O^{th} DMU.

On the other hand, the productivity for quality metrics, alternatively, could be considered in reference to input consumptions. The corresponding pure-service-quality efficiency score can be calculated as:

$$\begin{aligned}
&\text{Max } \pi_o^2 = \quad \textstyle\sum_{k=1}^{K} \frac{c_k}{q_k^o} \\
&\text{subject to} \\
&\textstyle\sum_{n=1}^{N} \mu_n x_i^n \leq x_i^o, \forall i \in m, \\
&\textstyle\sum_{n=1}^{N} \mu_n q_k^n = q_k^o + c_k, \forall k = 1 \dots K, \\
&\mu_n \geq 0, \forall n = 1 \dots N.
\end{aligned} \tag{2}$$

The quality measures are included in the objective expression but it does not take into account the operational metrics. Thus, model (2) can be termed as DEA for service quality, and the score θ_o^2 for O^{th} DMU be called quality efficiency.

In the previous two models, the framework explicitly tries to optimize the performance based on a single construct i.e., either operational efficiency or service quality. However, in real-life problems, decision makers are expected to optimize both these constructs simultaneously. The model for measuring the performance of the regions based on both operational and quality parameters can be obtained by combining the above two models into Equation (3).

$$\begin{aligned}
&\text{Max } \pi_o^3 = \quad \textstyle\sum_{k=1}^{K} \frac{c_k}{q_k^o} + \sum_{r=1}^{s} \frac{c_r}{y_r^o} + \sum_{p=1}^{t} \frac{c_p}{b_p^o} \\
&\text{subject to} \\
&\textstyle\sum_{n=1}^{N} \gamma_n x_i^n \leq x_i^o, \forall i \in m, \\
&\textstyle\sum_{n=1}^{N} \gamma_n y_r^n = y_r^o + c_r, \forall r = 1 \dots s, \\
&\textstyle\sum_{n=1}^{N} \gamma_n b_p^n = b_p^o - c_p, \forall p = 1 \dots t, \\
&\textstyle\sum_{n=1}^{N} \gamma_n q_k^n = q_k^o + c_k, \forall k = 1 \dots K, \\
&\gamma_n \geq 0, \forall n = 1 \dots N.
\end{aligned} \tag{3}$$

The optimal solutions in the DEA models determine the cumulative percentage of shortfalls in the desired outputs and the excesses in the undesired outputs. The objective values can thus be interpreted as measures of improvement possibilities. It is important to note that the models can identify weak efficiency, and only

DMUs identified as strongly efficient are marked as efficient.

In this particular use case, quality and efficiency are handled jointly in a single DEA model. Note, the same DEA model can be extended to incorporate more than two organizational goals e.g., sustainability, humanitarian operations, societal responsibilities and community engagement etc. Even though the objective of this paper is not to delve deeper into the multi-objective setting, a virtual use case is considered with l number of constructs. The l^{th} construct has K_l number of output indicators and the respective k^{th} output is denoted by y_{lk}^n for n^{th} DMU. The following DEA model presents a generalizable performance evaluation framework:

$$\text{Max } \pi_o^{3\prime} = \sum_{k=1}^{K_1} \frac{c_{1k}}{q_k^o} + \cdots + \sum_{k=1}^{K_l} \frac{c_{lk}}{q_k^o}$$

subject to

$$\sum_{n=1}^{N} \gamma_n x_i^n \leq x_i^o, \forall i \in m,$$
$$\sum_{n=1}^{N} \gamma_n y_{1k}^n = y_{1k}^o + c_{1k}, \forall k = 1 \dots K_1,$$
$$\dots$$
$$\sum_{n=1}^{N} \gamma_n y_{lk}^n = y_{lk}^o + c_{lk}, \forall k = 1 \dots K_l,$$
$$\gamma_n \geq 0, \forall n = 1 \dots N.$$

Lemma 1. *All the three models* (1), (2), *and* (3) *are feasible and bounded.*
Since model (3) is in a more generalizable form than the other two, we provide evidence for both feasibility and boundedness based on Equation (3). The proofs for Equations (1) and (2) follow the same approach. At first, the feasibilities are checked. One obvious solution to Equation (3) is to assign a weight of unity for the DMU 'O' and the rest are assigned zero weights.

$$\gamma_o = 1;\ \gamma_n = 0\ \forall n \neq 0.$$

Not only are all the constraints satisfied, but the slack and surplus variables are set to zero: c_r, c_p, and c_k. Thus, model (3) solves and produces efficiency of DMU 'O' to be 1 in a suboptimal solution. To prove the boundedness of Equation (3), the theory of dual linear programming (LP) is used. According to the duality property, if the dual problem of an LP is feasible, then the original LP is bounded. Therefore, the dual formulation of Equation (3) is written as follows:

$$\text{Min } \sum_{i=1}^{m} v_i x_i^o + \sum_{r=1}^{s} u_r y_r^o + \sum_{p=1}^{t} e_p b_p^o + \sum_{k=1}^{K} w_k q_k^o$$

subject to

$$\begin{aligned} & \sum_{i=1}^{m} v_i x_i^n + \sum_{r=1}^{s} u_r y_r^n + \sum_{p=1}^{t} e_p b_p^n + \sum_{k=1}^{K} w_k q_k^n \geq 0, \forall n = 1 \dots N, \\ & -u_r \geq \frac{1}{y_r^o}, \forall r = 1 \dots S, \\ & e_p \geq \frac{1}{b_p^o}, \forall p = 1 \dots t, \\ & -w_k \geq \frac{1}{q_k^o}, \forall k = 1 \dots K, \\ & v_i \geq 0, \forall i; u_r, w_k, e_p\ unrestricted. \end{aligned} \tag{4}$$

The weights of desirable outputs and quality measures are negative; whereas the weights of inputs and bad outputs are positive. The signs of these notations can be interpreted from the first set of constraints: $\sum_{i=1}^{m} v_i x_i^n + \sum_{r=1}^{s} u_r y_r^n + \sum_{p=1}^{t} e_p b_p^n + \sum_{k=1}^{K} w_k q_k^n \geq 0$. It can be written as:

$$\sum_{i=1}^{m} v_i x_i^n + \sum_{p=1}^{t} e_p b_p^n \geq -\sum_{r=1}^{s} u_r y_r^n - \sum_{k=1}^{K} w_k q_k^n, \forall n.$$

Let's denote a new set of the weights for the desirable outputs, $\hat{u}_r$ and quality measures $\widehat{w}_k$ as $\hat{u}_r = -u_r$ and $\widehat{w}_k = -w_k$ where both the new set of weights are positive.

$$\sum_{i=1}^{m} v_i\, x_i^n + \sum_{p=1}^{t} e_p\, b_p^n \geq \sum_{r=1}^{s} \hat{u}_r\, y_r^n + \sum_{k=1}^{K} \widehat{w}_k\, q_k^n, \forall n.$$

Finally, dividing both sides by the LHS of the above expression gives the following ratio.

$$\frac{\sum_{r=1}^{s} \hat{u}_r y_r^n + \sum_{k=1}^{K} \widehat{w}_k q_k^n}{\sum_{i=1}^{m} v_i x_i^n + \sum_{p=1}^{t} e_p b_p^n} \leq 1,\ \forall n.$$

This ratio is an extended form of the ratio of the weighted outputs to the weighted inputs, following the classical definition of efficiency. The numerator is adjusted with the weighted quality measures since the inputs are now responsible for maintaining these quality measures along with the outputs. Similarly, the denominator is adjusted with the weighted bad outputs denoting the cost at which the desirable outputs and quality measures are produced. To prove the feasibility of Equation (4), assume that the second, third, and fourth sets of constraints are satisfied at equality. To achieve this, the weights of good outputs, bad outputs, and quality measures are set accordingly as follows:

$$u_r = -\frac{1}{y_r^o}, \forall r = 1 \dots S, e_p = -\frac{1}{b_p^o}, \forall p = 1 \dots t, w_k = -\frac{1}{q_k^o}, \forall k = 1 \dots K.$$

The lower limit of the weighted inputs is thus set as:

$$\sum_{i=1}^{m} v_i\, x_i^n \geq \sum_{r=1}^{s} \frac{y_r^n}{y_r^o} + \sum_{k=1}^{K} \frac{q_k^n}{q_k^o} - \sum_{k=1}^{K} \frac{b_p^n}{b_p^o}.$$

Now, it is known that not all x_i^ns are zero (i.e., output production is not possible without any inputs). Therefore, the values of v_i can be set arbitrarily high to satisfy the above condition. This proves that Equation (4) is feasibly and this implies that Equation (3) is bounded.

None of the DEA models, (1), (2) or (3) reports efficiency score directly. The optimal objective value denotes the amount of total inefficiency as a fraction of the respective parameters. Yet, one may conclude that the efficiency varies inversely with the total inefficiency i.e., π_o^1, π_o^2, and π_o^3. The resultant efficiency score can be thus formulated as $\theta_o^1 = \frac{1}{1+\pi_o^1}$, $\theta_o^2 = \frac{1}{1+\pi_o^2}$, and $\theta_o^3 = \frac{1}{1+\pi_o^3}$. A DMU is efficient when the cumulative fractional slacks and surpluses sum up to zero. It is important to note that the slacks are expressed as the fraction of the input and output to make the measure of inefficiency unit invariance. The separability of the inefficiency as denoted by the slack and surplus variables in the objective function based on their association to the productivity or quality aspect enables us to examine the extent of compatibility between the two.

***Definition* 1.** A DMU is said to be strongly efficient if no inputs or outputs can be improved without worsening other inputs or outputs.

Relative efficiencies in DEA model (3) is always calculated over a set of feasible production-mixes i.e., by $T = \{(x, y, b, q) \mid x \geq \sum_{n=1}^{N} \gamma_n\, x_i^n, y \leq \sum_{n=1}^{N} \gamma_n\, y_r^n,\ b \geq \sum_{n=1}^{N} \gamma_n\, b_p^n, q \leq \sum_{n=1}^{N} \gamma_n\, q_k^n\}$. In other words, input x produces good and bad outputs y and b while maintaining quality metrics at level q. The weak efficiency frontier for production possibility set T is given by $\mathbb{F}^w(T) = \{\{(x, y, b, q) \mid x > \hat{x}, y < \hat{y}, b > \hat{b}, q < \hat{q} \Rightarrow (\hat{x}, \hat{y}, \hat{b}, \hat{q}) \notin \mathrm{T}\}$. When the inequalities conditions are satisfied strictly and $(\hat{x}, \hat{y}, \hat{b}, \hat{q}) \in \mathrm{T}$, then we can say that $(\hat{x}, \hat{y}, \hat{b}, \hat{q})$ strongly dominates (x, y, b, q). The reason it is called weakly efficient because it ignores the possibility of unilateral improvement in one or a few of production parameters (x, y, b, q). For example, for a point on the weakly efficient frontier: $(x, y, b, q) \in \mathbb{F}^w(T)$ if $\exists\ (\hat{x}, \hat{y}, \hat{b}, \hat{q}) \in$

T such that $x > \hat{x}, y = \hat{y}, b = \hat{b}, q = \hat{q}$, one cannot say $(\hat{x}, \hat{y}, \hat{b}, \hat{q})$ strongly dominates (x, y, b, q) and hence both of them have the same weak efficiency. But as per definition 1, since the input vector can be improved $\hat{x} < x$, without worsening good output (y), bad output (b), or quality (q), the current product mix (x, y, b, q) is not strongly efficient.

***Lemma* 2.** *All the three models* (1), (2), *and* (3) *report "strong" efficiencies.*

Suppose DMU-O is not "strongly" efficient with product mixes as (x_o, y_o, b_o, q_o). This means at least one of the inequalities is satisfied strictly: $y \leq \sum_{n=1}^{N} \gamma_n y_r^n,\ b \geq \sum_{n=1}^{N} \gamma_n b_p^n, q \leq \sum_{n=1}^{N} \gamma_n q_k^n$. Suppose, it is the input constraint that is satisfied at strict inequality. Now, if $\sum_{n=1}^{N} \gamma_n y_r^n > y_r^o$, then $c_r > 0$. Thus, $\pi_o^3 > 0$ and $\theta_o^3 < 1$, i.e., the DMU-o is not efficient in (3). Clearly DMU-o is efficient in model (3) if $\theta_o^3 = 1$, or $\pi_o^3 = 0$. In other words, $c_r = 0$, $c_p = 0$, and $c_k = 0$. Therefore, the output and quality constraints are held at equality: $y = \sum_{n=1}^{N} \gamma_n y_r^n,\ b = \sum_{n=1}^{N} \gamma_n b_p^n, q = \sum_{n=1}^{N} \gamma_n q_k^n$. It also means that $\nexists (y, b, q) \in$ T such that $y > y_o$ or $b < b_o$ or $q > q_o$. Therefore, DMU-o is "strongly" efficient. A similar approach shows the same for model (1) and (2).

The optimal solution of Equation (3), can be expressed as the summation of operational slack $\pi_o^{31} = \sum_{r=1}^{s} \frac{c_r}{y_r^o} + \sum_{p=1}^{t} \frac{c_p}{b_p^o}$ and quality slack $\pi_o^{32} = \sum_{k=1}^{K} \frac{c_k}{q_k^o}$. Such decomposition leads to the separation of operational and quality efficiency, denoted by θ_o^{31}, and θ_o^{32} respectively, from the overall efficiency. They are calculated as $\theta_o^{31} = \frac{1}{1+\pi_o^{31}}$, and $\theta_o^{32} = \frac{1}{1+\pi_o^{32}}$. To distinguish it from θ_o^1 and θ_o^2, the newly developed scores θ_o^{31}, and θ_o^{32} are termed as "overall-operational" and "overall-quality" efficiency score. It is clear from Equation (3), that the same amount of inputs are consumed to enhance the capabilities along two dimensions efficiency and quality. If more inputs are employed to improve quality, the operational measures suffer and vice-versa. Therefore, they share a substitutable relation between them. Such types of compromises are known as zero-sum trade-offs in manufacturing literature (Sarmiento and Shukla, 2011).

***Theorem* 1.** *The decomposed operational (or service quality) efficiency score in Equation* (3) *sets the maximum value of the operational efficiency in Equation* (1) (*or service quality in Equation* (2)).
Suppose the optimal solution of Equation (3) is γ_n^*, c_r^*, c_p^*, and c_k^*. It is important to note that this solution is feasible in both pure-operational and pure-service-quality DEA models i.e., Equations (1) and (2), respectively. This implies that at optimality $\pi_o^{31} \leq \pi_o^1$ and $\pi_o^{32} \leq \pi_o^2$ conditions hold. As a result, the efficiency scores in Equations (1) and (2) are bounded from above by the decomposed operational and service-quality based efficiency scores in Equation (3): $\theta_o^{31} \geq \theta_o^1$ and $\theta_o^{32} \geq \theta_o^2$.

In addition, model (3) expresses the inefficiency measure as $\pi_o^3 = \sum_{k=1}^{K} \frac{c_k}{q_k^o} + \sum_{r=1}^{s} \frac{c_r}{y_r^o} + \sum_{p=1}^{t} \frac{c_p}{b_p^o}$. Since, it reports strong-efficiency, i.e., if any single value of c_k, c_r or c_p is non-zero then $\pi_o^3 > 0$. Therefore, any efficient DMU in model (3) will have $\pi_o^3 = 0$. As a result, the individual measures of inefficiency in quality and operational efficiency are both zero: $\pi_o^{31} = \sum_{r=1}^{s} \frac{c_r}{y_r^o} + \sum_{p=1}^{t} \frac{c_p}{b_p^o} = 0$ and $\pi_o^{32} = \sum_{k=1}^{K} \frac{c_k}{q_k^o} = 0$. Thus, strong-efficiency is preserved in the case of both quality and operational efficiency measure.

Though the theoretical construct of Equation (3) ensues substitutability between them, the framework analyzing actual data might lead to a different observation. Quality measures and operational efficiency may have a circular relationship where one influences the other. For example, thorough healthcare examinations and observations in the room might result in fewer fatalities among inpatients. Contrary to the popular wisdom of substitutability and differentiality, the overall approach in Equation (3) demonstrates the need for a minimum level of compatibility, as the two aspects are adjusted to optimize the cumulative result. There are seminal works theorizing these concepts (Schmenner and Swink, 1998) that advanced the concept of frontier trade-offs, which is consistent with the idea of potential cooperation between two competitive constructs. The overall DEA framework represents a balanced approach in contrast to operation and quality-centric approaches. Nonetheless, decision-makers need to quantify the loss of one metric due to the incorporation of another. This measure has been captured by the cost of efficiency value as derived below:

$$CoE = \frac{\pi_o^2 - \pi_o^{32}}{\pi_o^2}.$$

The amount of quality loss equals the difference between the scope of the percentage improvement in quality metrics in quality-centric approach and in the overall approach. The quality loss when scaled relatively to the maximum achievable quality improvement, it is termed as Cost of Efficiency (CoE). Note, the cost of efficiency attains its highest value of 1 or 100% when the scope of improvement in quality metric in the overall approach becomes zero. This presents a situation where the quality aspect of healthcare service is completely ignored in favor of operational efficiencies. Then CoE attains its lowest value of 0.

Similarly, we define the cost of quality as the efficiency loss due to the incorporation of the quality aspect of healthcare service and can be denoted by,

$$CoQ = \frac{\pi_o^1 - \pi_o^{31}}{\pi_o^1}.$$

The value of the metric varies between zero and one. When the cost of quality is very high (almost 1), it can be concluded that the aspect of operational efficiency is completely discounted in favor of the quality of service.

In healthcare sector, introduction of high-technology treatment, process innovation, or new practices can improve the resulting outcomes and it happens frequently. As classical DEA models focus on static performance evaluation, it fails to measure the gradually evolving state of production systems in a particular sector. The total planning horizon of T years is divided into $(T - W + 1)$ number of windows of length W $(= 3$ or 4 years). To incorporate the inter-temporal factors, classical DEA framework had been extended (Cooper et al., 2007) where each of the production units in each year is evaluated against its peers in the same year and all production units (including itself) in rest of the years in the same window. The length is set in such a way that there is minimal amount of variability in terms of technical changes that makes the comparison of peers within a window as fair as possible. The other advantage of using all the DMUs within a window of length W is that it helps to the train the classification algorithm with more data than what would have been possible with a single year analysis.

4. Classification Algorithms

Since the DMUs operate in a larger economic and social setup, often there is significant impact of external environment factors on the performance metrics of those DMUs. One popular way to assess such impact is use of OLS regression. However, the underlying assumptions of regression are required to be fulfilled. Often such assumptions are violated in real world problems and two primary reasons for that are: one, the independent variables should have no or very low correlation. The explanatory variables broadly belong to five categories e.g., Medical, Education, Public Health, Wealth, and miscellaneous. Thus, assumption of uncorrelated features is difficult to hold in our case-study. Two, the error should follow normal distribution with zero mean and a constant variance. Since in DEA the efficiency scores are mostly skewed and use of a linear function is difficult for approximation. However, such difficulties can be circumvented by the use of a technique that does not require such the restrictive assumptions. Machine Learning techniques such as, Support Vector Machine (SVM), Decision Tree (DT) and Naïve Bias (NB) classifiers can handle correlated features. Furthermore, these techniques can accommodate non-linear relationships among predictors.

Once the first level of analysis is completed using DEA, the efficiency scores of the DMUs need to be clustered into well-defined classes. In most efficiency clustering studies so far, researchers have segregated DMUs into a pre-determined set of classes based on their efficiency scores (Tsolas et al., 2020) However, this study takes a dynamic approach where the number of clusters is determined based on the similarity-dissimilarity measures of the underlying efficiency scores.

One of the primary objectives of this study is to predict the future classification of a futuristic DMU based on the given set of inputs and outputs. The proposed framework function independently of any pre-defined set of clusters but rather depend on natural clustering of data. In unsupervised learning, there is no labelling of cluster number, even though there might be hidden cluster patterns in the data. To achieve this, we have used K-means clustering method. For a given value of k, the number of clusters, the data points are classified into k partitions to optimize an objective function: the similarity among points in the same cluster. This method starts with k arbitrarily chosen points acting as centroids of k clusters, and the remaining objects are assigned to one of these k clusters, which has the minimum distance, or highest similarity. It then calculates new centroid for each cluster. Finally, it converges after a number of iterations when the centroids of the cluster stop getting changed in successive iterations. If there is no natural clustering rule available, k is chosen arbitrarily.

The value of k is determined through two techniques: Silhouette index and Elbow method. In the Elbow method, the sum of squared distances between each data point in a cluster and the centroid of the cluster is considered. As the number of clusters increases, the distance falls. However, there exists a point up to which the fall is significant, but after that point, the fall becomes marginal. To avoid overfitting the model, the optimal k value is considered, where the sum of square distance stops decreasing at a significantly faster rate.

However, because it unilaterally focusses on only intra-cluster distance i.e., the similarity among cluster points, the measures of dissimilarities among points in different clusters mostly get ignored. Thus, it is always a good idea to combine the results of Elbow method with the Silhouette index, which places equal importance on similarity within clusters and dissimilarity among different clusters. In Silhouette index, for each data point i in a particular cluster, say M, the inter-cluster distance between the point under consideration i and all other points j in the closest cluster, say N, is calculated and the mean distance is denoted as $x(i)$. Similarly, the average distance between the specific data point i and other data points, i', in the same cluster M, is denoted by $y(i)$. The Silhouette index value is calculated as a ratio: $\frac{x(i)-y(i)}{Max\{x(i),y(i)\}}$. The closer the value is to one, the better the classification. To enhance the robustness of the model, we have consulted the outputs of both methods.

Since the objective of the second phase of the analysis is to explore the interdependency of healthcare efficiency on various environmental and contextual variables. This study has not reused the input and output variables in the training of the supervised machine learning algorithms. Furthermore, the accuracy indicators only reinforce our conclusions regarding the interdependency of efficiency level on external factors. Although accuracy measures are important indicators of the efficiency of the classification models, these measures are not used from this perspective in this study. The purpose of these classification algorithms is to predict the efficiency class of a region based on environmental or contextual factors.

It is important to note that the phase two is multiclass classifier model. A range of classifying techniques are used to suit the nature of the data, such as Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM), and Naïve Bias (NB) classifier. One advantage of these algorithms, except SVM, is that it can handle qualitative (categorical or nominal/ordinal) variables. Categorical variables need to be pre-processed using dummy variables for SVM. Except for logistic regression, there are no pressing assumptions to be fulfilled for the application of the classification techniques.

4.1 Feature Selection

We initialize the analysis by employing ad-hoc features that were selected based on our domain knowledge and a literature survey. The five overarching factors; Public health, medical resources, Education, Wealth indicators, and Miscellaneous factors together comprise a total of 33 sub-factors. Due to the limited data points (regions) relative to the number of features, it is necessary to control the degrees of freedom of the DMUs by restricting the number of features within a specified limit. The literature refers to the phenomenon of numerous features overwhelming the few training examples as the curse of dimensionality (Charles et al., 2019). Another potential benefit derived from this approach is the reduction of redundancy, particularly when dealing with other variables or set of variables.

To select features with maximum discriminatory power, we employ rank techniques to order them before selecting top features. This method is popular due to its simplicity and effectiveness in practical applications (Chandrashekar and Sahin, 2014). For this purpose, we utilize variable ranking approach where each feature i is assigned a scoring function S_i and the features are arranged in decreasing order of S_i. A substantial body of literature is dedicated to the information-theoretic approach for deriving scores for ranking (Bekkerman et al., 2003; Forman, 2003). Mutual information calculations between explanatory and target variables have also been explored (Dhillon et al., 2003; Torkkola, 2003).

However, relying solely on the ranking score may introduce redundancy in the model. Achieving the current level of accuracy might be possible with a smaller subset of variables in the absence of redundancy. To mitigate such situations, we also incorporate correlation among the variables as a filtering criterion before including them in the classification model.

Although high model accuracy is always a desirable trait in any data-driven analysis, it often leads to the inclusion of an increasing number of explanatory variables in the analytic technique. However, in managerial decision-making, a more crucial aspect is to identify the most important set of variables that truly drive the interrelationship with the dependent variable. This step also eliminates variables with high correlation to other explanatory variables. The inclusion of interdependent or colinear variables does not significantly enhance explanatory power and can destabilize current analytical techniques. To address the importance of variables while avoiding correlated variables, we have implemented a two-step validation technique.

The variable importance is measured using an information theory metric known as information gain, which is proportional to how much information a feature holds about a particular class (Gray, 2011). It is based on entropy calculation, which measures the impurity or uncertainty in a dataset. Information gain represents the reduction in entropy in the class if we split the dataset according to the feature. Once we calculate the information gains of each variable, we arrange them in decreasing order. Initially, the top most variable on the list is selected by default. We choose the second-highest variable only if it does not have a high correlation with the first one. This process continues we exhaust all the variables.

4.2 Decision Tree
It is a tree-based classification model that splits the data in a top-down manner, following a greedy approach (Quinlan, 1986). Some of the most widely used versions are ID3, C4.5, and CART. This technique is popular due to the simple interpretation of its result. It graphically illustrates how a decision is made and the criteria that guide the decision-making process. The model selects the feature with the highest information gain value to split the root node and, in the next step, chooses the best variable to split each of the leaf nodes. It recursively uses this divide-and-conquer philosophy until it exhausts all the variables or reaches a state of pure classification node. Since information gain has already been used to select variables, a different attribute selection criterion is employed, known as the Gini index. To address the overfitting problem in our study, we have set a limit to the depth of the tree, determining how many times the non-leaf nodes can be split; this technique is also called tree-pruning (Quinlan, 1987).

4.3 Logistic Regression
It is one of the earliest and most widely used classification methods. It can be conceptualized as a linear regression; however, the dependent variable is then transformed using sigmoid function, ensuring it ranges between 0 and 1. This transformed value serves as the probability of the event occurring. Logistic regression is highly effective and easy to apply in binary classification problems, as it maps the variable space into probability space where the method anticipates either of the two discrete values, 1 or 0. However, for the purposes of this study, we have used one-vs-rest (OVR) technique for all the binary classifiers to ensure that they perform equally well in multiclass problems. In the OVR technique, the multiclass problem is decomposed into multiple binary classification problems. In each of these problems, one class is examined, while the rest of the classes are combined into a single class referred to as "the rest" class.

4.4 SVM Classifier
This technique is based on finding a supporting hyperplane that separates the two classes while maximizing the margin between them. The objective is to determine an equation in higher dimensions. Data points can exist on either side of this plane, which is derived by inputting the values of the points into the hyperplane equation. In a binary classification problem, the two classes can be associated with the positive and negative signs of hyperplane equation when solved for a particular data point. The absolute value of the equation represents the distance of the point from the plane and serves as a reliability measure.

4.5 Naïve Bayes
The working of this classifier is based on the statistical theory of Bayes' theorem. It calculates the posterior probability of a hypothesis, which is "the observed data point belongs to a particular class". It can be shown that the posterior probability is directly proportional to the probability of the observed point given the hypothesis, under the assumption that the probability of the observed point and probability of the hypothesis are equal across the classes. One fundamental assumption behind this is that the features are independent of each other for any given class.

5. Applications to Chinese Healthcare System
5.1 Input-Output Selection
China's healthcare sector is primarily structured into three levels: primary, secondary, and tertiary (Lu et al., 2008). Primary healthcare encompasses prevention services, medical care, rehabilitation, and family planning. At the secondary level, the routine services and major diseases are managed by community health centers and hospitals. Third-level healthcare institutions lack support, both technically and financially, from higher-tier institutions. Therefore, we exclude these health centers from the scope of our study. Previous research (Song et al., 2019) has provided significant evidence of considerable spillover effects during

patients transition between hospitals. To address these externalities, we evaluate healthcare performance in each of China's 31 provinces or regions, rather than focusing on individual medical units. Such aggregated data annuls the need to address micro-level data e.g., the complexities of the diseases (as often indicated by case mix index) or the severity of the patients.

Previous studies have shown that health factor of any region depends on various factors such as demographics, socio-economic conditions, the financial wellbeing of the citizens, and ecological parameters, among others (Hood et al., 2016). In the second phase of our study, we investigated the underlying association between external factor or a set of factors and health-service efficiency. Instead of choosing a-priori a set of external factors, our study identifies a set of overarching constructs, including the constructs that are remotely associated with the health-sector), which can severely impact health services. This study has excluded irrelevant factors to avoid any spurious correlation. Afterward, we consider all possible variables within each of these constructs. A set of feature selection criteria has been imposed on the entire dataset to remove highly correlated data and data that have little impact on health outcomes.

Table 1. Descriptive statistics of input and output parameters.

Summary	Inputs		Operational output			Quality		
	Bed	Doctor	Patient discharge	Death in observation-room	Death in inpatient	Cases in observation	Surgical operation	Health exam
Mean	305.7	85561.2	729.15	15.07	27.95	163.77	164.94	1460.97
Variation	347.88	2782496101	234886.18	266.86	444.92	18158.4	14266.1	2525655.7
Std. Dev.	18.65	52749	484.65	16.33	21.1	134.75	119.44	1589.23
Max.	349.6	210806	1684.2	71.70	77.34	561.4	556.46	8490.95
Min.	271.5	4791	28.71	1.828	6.9	91.4	4.98	142.2
Range	78.1	206015	1655.5	71.51	76.66	552.26	551.48	8348.75

5.2 External Variable Selection

Outcomes of healthcare service are impacted by a wide range of factors beyond just the healthcare infrastructure. This includes socio-economic factors, standard of living, promotion and awareness of healthy behaviors among the citizens, physical environment, to name a few. We chose to categorize our external variables into five broad categories: Public health measures, medical factors and infrastructure, educational factors, Wealth indicators, and other Miscellaneous factors Within each of these overarching categories, we further identified a set of more quantifiable variables. **Figure 3** represents the schematic overview of quantitative techniques and the nature of the underlying data sets. **Tables 10, 11, 12, 13** and **14** in appendix (B) exhibit the summary statistics of all such quantifiable variables under the five broad categories.

5.3 Empirical Results

This section explains how to derive useful insights from the two-phase analytical framework. The data is obtained from the Statistical Yearbook, China. An overview of the data for the 31 provinces of China is presented in **Table 1**. The average overall efficiency for window-1 (i.e., from 2013 to 2015), for the Chinese regional health administrations stands at 58%. Regions such as Fujian, Hunan, Guandong, and Qinghai have attained the optimal level of outputs. Some regions such as Jiangsu, Sichuan, and Yunnan fell just short of achieving full overall efficiency. Previous research works considered only a sub-set of the input-output variables used in this paper (Li et al., 2019; Jiang et al., 2017). Understandably, the results have changed-but not for the highly inefficient regions. For example, Jilin, Hebei, Inner-Mongolia, and Liaoning have consistently performed poorly (Jiang et al., 2017).

Interestingly, Fujian, Hunan, and Guandong are the three regions that exhibit full efficiency throughout the periods studied here. Jiangsu, Yunnan, Qinghai, and Jianxi maintain a very high-efficiency standard with minimal variance across the years (see **Figure 4** for window-1, and **Figures 5** to **7** in Appendix (A) for windows -2, -3, -4, respectively). The rest of the regions perform poorly in either of the two aspects of medical parameters. Beijing, Inner Mongolia, Jilin, Heilongjiang, and Shaanxi require the deployment of an all-around development plan and policy restructuring to bridge the wide gap (the cumulative increment of output needs to be at least around 80%) in performance compared to the efficient regions. A similar type of inequality trend based on geographical area has been quantified by Zhang et al. (2018). A few regions, such as Chongqing, Henan, and Anhui have enough room for improvement (projected output increment should cumulatively be approximately 30%).

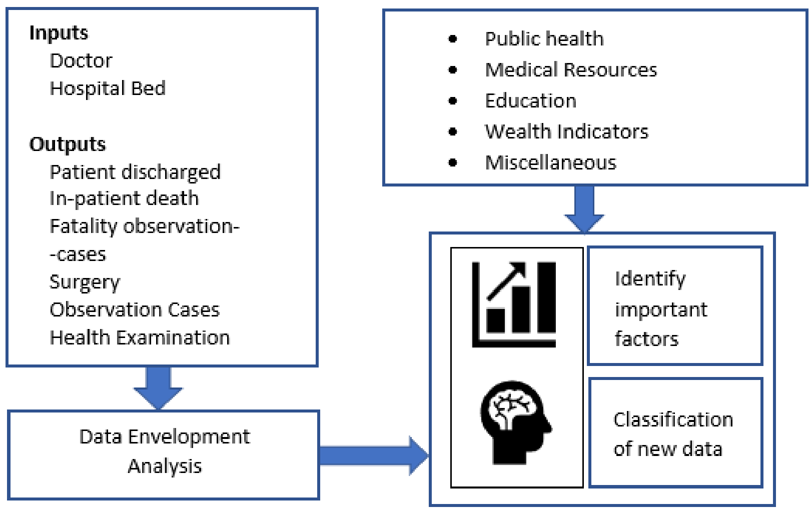


Figure 3. Schematic view of the analysis framework.

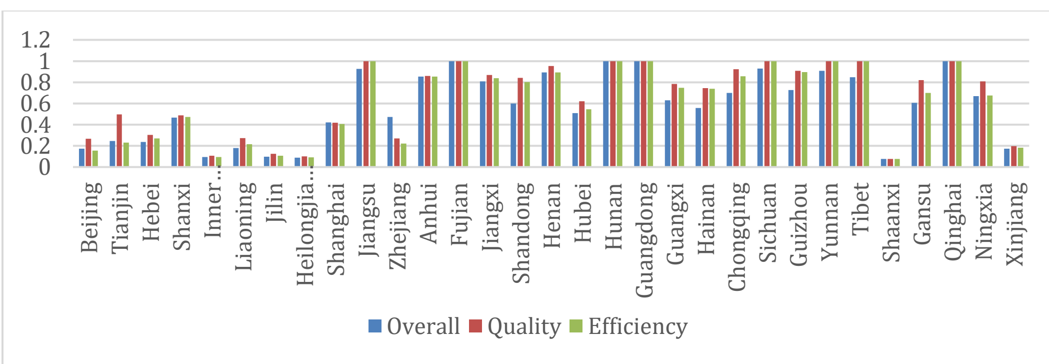


Figure 4. Window-1 average efficiency of regional health sectors.

Improvement is quantified by analyzing the 3-year moving average of the efficiencies of the regions. This kind of trend is more reliable than a year-on-year analysis, as the average efficiency of three years in every window cancels out any external fluctuations and accidental outliers. The comparative analysis for window -2 to -4 can be found in Appendix (A). Based on the scores, the Chinese regions can be divided into three groups. Within the first group, some regions made continual improvements, such as Yunnan (90% to

100%), Ningxia (67% to 90%), Chongqing (70% to 100%), Shandong (60% to 78%), Hubei (51% to 62%), Xinjiang (17% to 25%), Inner Mongolia, and Jilin (9% to 12%). Whereas several other regions exhibit only a recent increment trend, such as Beijing, Fujian, Hubei, and Yunnan, Shaanxi. The second group includes regions that managed to maintain their efficiencies across the windows, such as Tianjin, Zhejiang, Liaoning, Jiangsu, and Hebei. There is serious concern over the third group where the efficiency trend of a handful of regions has exhibited a continual drop, e.g., Henan (90% to 73%), Hainan (55% to 28%), Guangxi (62% to 43%) Gansu (60% to 50%), Anhui (85% to 74%).

The primary finding of this study is the breakup of overall efficiency into quality-centric and efficiency-centric scores. If one considers the overall efficiency scores, they hover around the value of 58% across all the windows with very small variations. However, this observation is grossly misleading. In **Table 2**, the year-wise efficiency scores for operational, quality, and overall approaches are reported. The scores show a continual improvement of operational efficiency. Whereas similar upward trends in the other two aspects remain sporadic indicating insufficient resources for performing adequate surgeries and conducting proper health examination. In order to smoothen out yearly fluctuation due to uncontrollable and unforeseen determinants, in **Figure 5** exhibits moving-window based efficiency assessment and it clearly shows that quality score is dropping (except in window-4). On the other hand, efficiency scores remain at 68% after a slow start in window-1. It is safe to conclude that the Chinese health sector is currently more efficiency driven.

Table 2. Year-wise efficiency score for quality, operational, and overall measure.

	2013	2014	2015	2016	2017	2018
Quality	0.64	0.66	0.60	0.61	0.61	0.64
Operational	0.61	0.67	0.64	0.68	0.69	0.73
Overall	0.61	0.58	0.56	0.58	0.58	0.62

The efficiency trends of the quality and operations -specific approaches vary to a great extent (**Figure 5**). There are sixteen regions that gradually improved their operational efficiency over the years, e.g., Anhui, Beijing, Guangxi, Hainan, Jianxi, Jiangsu, Tianjin, etc. This number reduces to only eight in service quality efficiency. Regions like Gansu, Hunan, Jiangxi, Shaanxi, Shanghai, and Yunnan exhibited high quality trends.

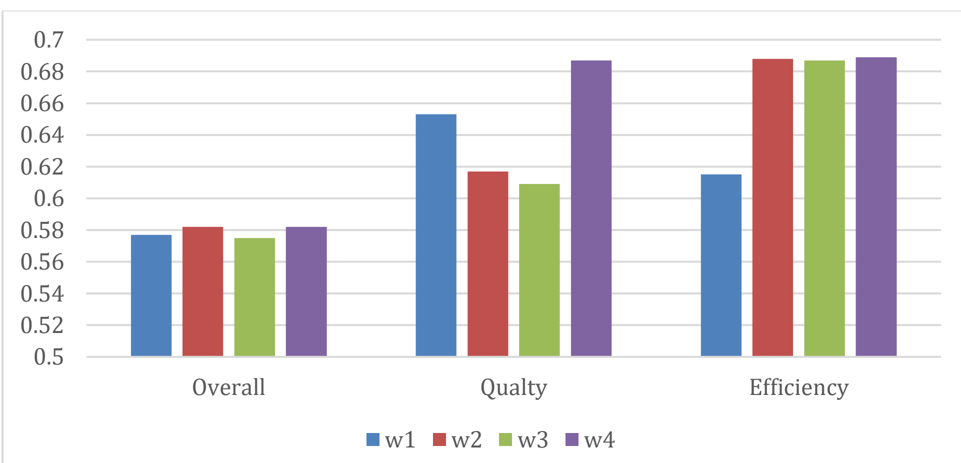


Figure 5. Average window efficiency in the three approaches.

In the overall performance metric, only six regions observed considerable decline in operation efficiency: Shanxi, Jiangsu, Guangxi, Hainan, Guizhou, and Gansu. This number grows to eleven in service quality e.g., Tianjin, Hebei, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Guangxi, Hainan, Sichuan, and Gansu. An important insight for the policy-makers is that the regions that suffered poor service quality have predominantly exhibit poor operational efficiency. Even though researchers exploring the impact of quality on the operational efficiency have reached conflicting conclusions, it is very evident from our study that safety- measured by health examine- and effective medical treatment-measured by observation cases and surgery largely determine the outcome of this sector namely patient discharge and low mortality rate. Previous studies (Zhang et al., 2018) also indicate to the fact that institutional infrastructure is a major source of inequity which further justifies the findings of this paper. Zhang et al. (2018) also showed that inequity among the provinces with respect to institutional infrastructure is gradually improving. In our study, we have gone one step further in our analysis which reveals that as higher number of surgeries are undertaken, it improves both the fatality rates among inpatients and observation room.

The upkeep and maintenance of quality standards are often disregarded in hospitals. This is evident from the share of inefficiency emanating from the three quality measures, which approximately account for 75% of the total overall inefficiency, on average (**Figure 6**). **Figure 6** also shows the individual contributions of all the outputs to the overall inefficiencies in the healthcare system. There are a few interesting trends, e.g., the share of observation cases has been increasing considerably. A similar trend is observed for health exam, where its share increases from 9% in window-1 to 25% in window-4.

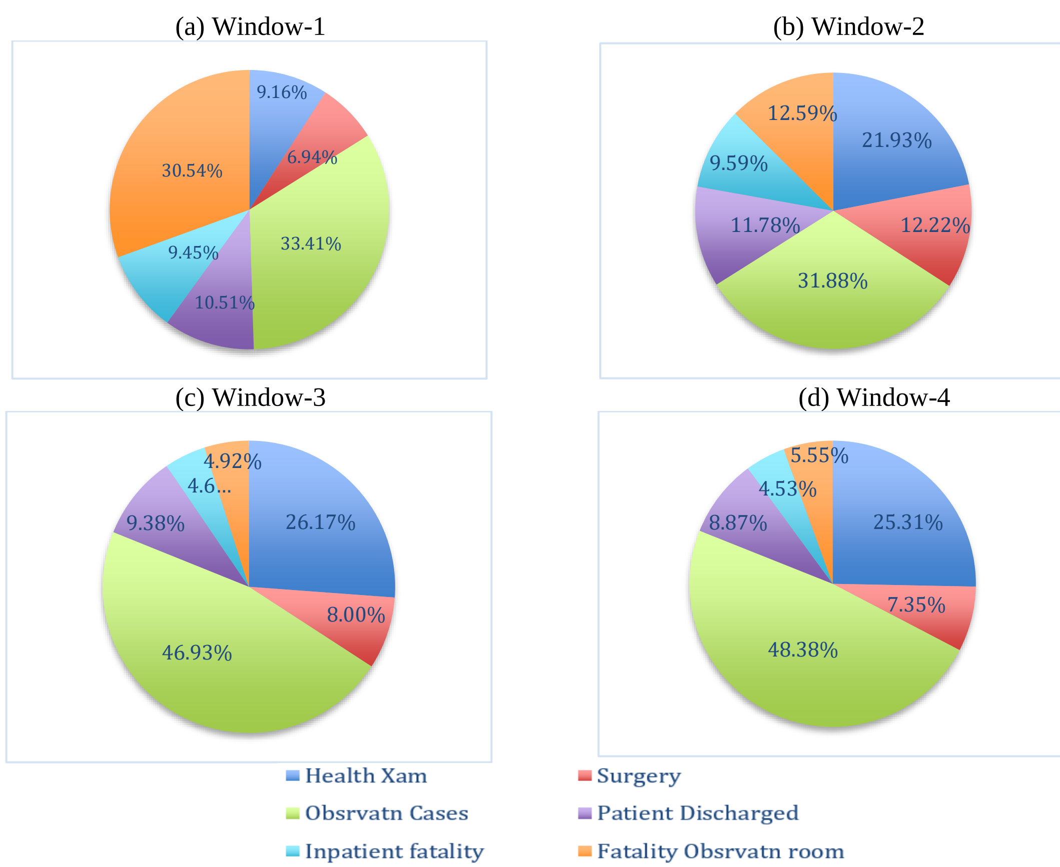


Figure 6. Sources of inefficiency in overall approach (window wise).

However, if we turn our focus towards quality-centric approach, where we concentrate on the indicators primarily responsible for high-quality healthcare service, an opposite trend is found as observed in the overall approach (**Figure 7**). For example, the individual contribution of the observation room has significantly decreased over time. However, the number of health exams has become more severe in the health service sector. The primary sources of inefficiency in medical health services have not changed much over the time horizon. The pattern here is the same as in the overall approach, wherein the observation cases contribute to almost 60% of total health issues, followed by health examinations (30%), and Surgury (10%) (see **Figure 7**).

When all the resources are targeted towards improving operational efficiency in terms of the number of patients and fatality rates (bad outputs), the window-wise trends for operational efficiencies are not consistent, as in the previous two cases (**Figure 8**). However, in most cases, fatality in observation cases has been the most predominant contributor to overall inefficiency. There has been a significant improvement in patient discharge from 40% to 7% of total inefficiency. The other two factors, number of patients treated and inpatient fatality, vary greatly in terms of their share of inefficiency.

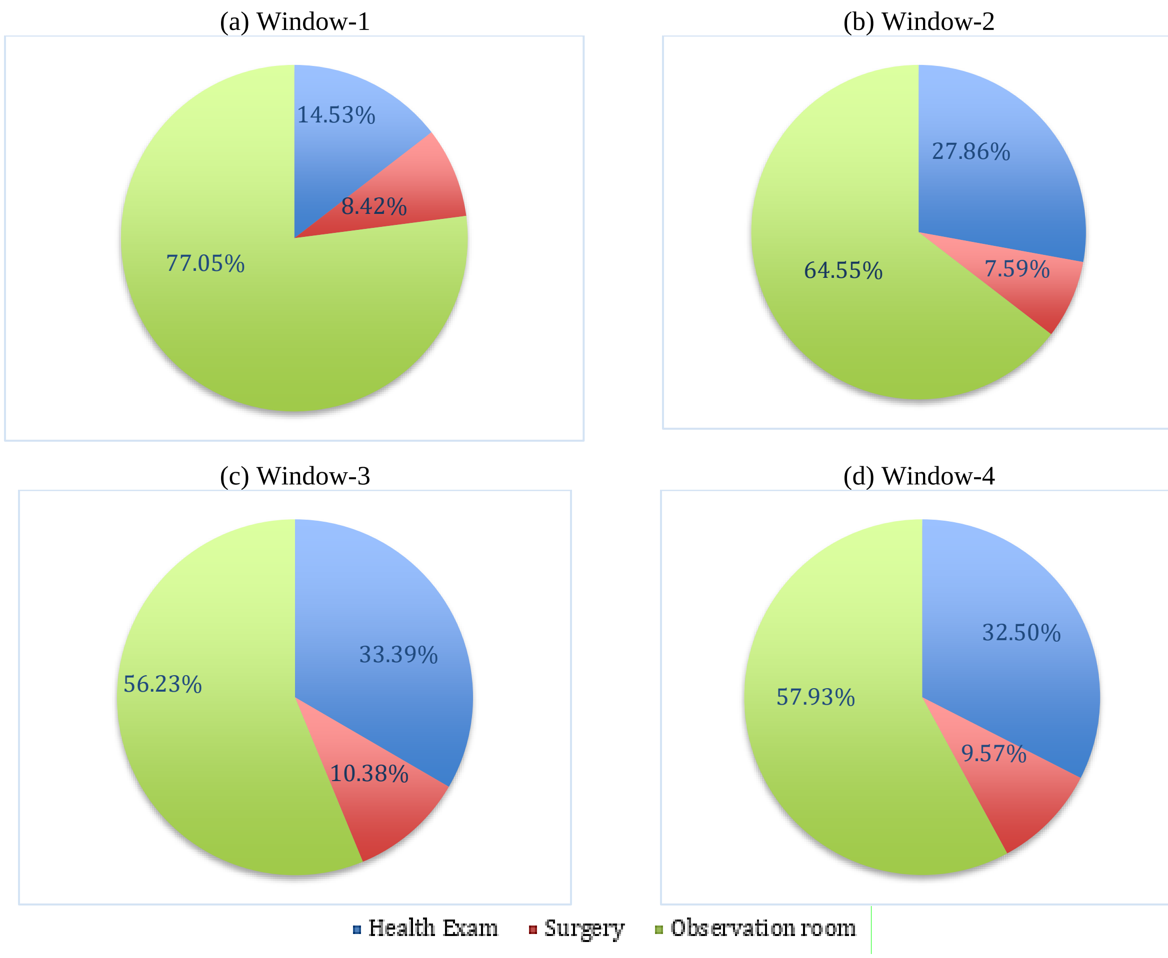


Figure 7. Sources of inefficiency in quality-centric approach (window wise).

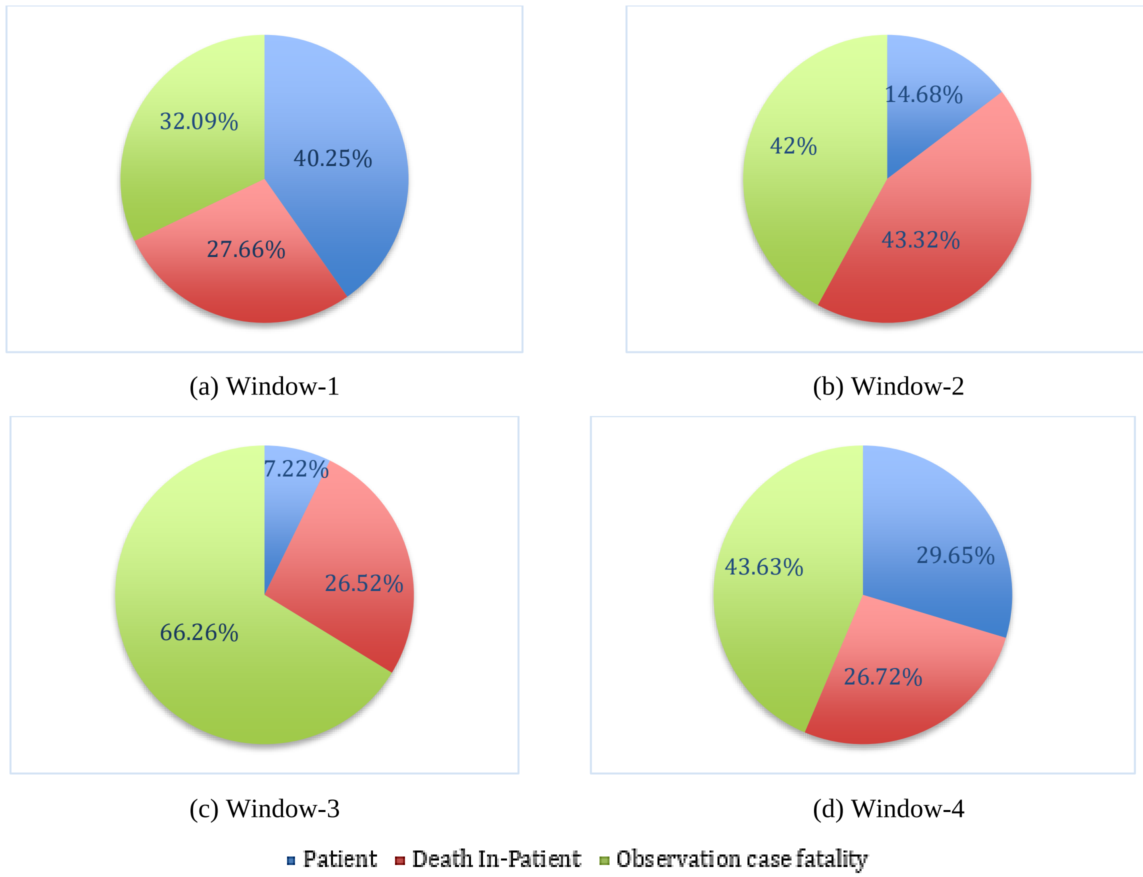


Figure 8. Sources of inefficiency in operational approach (window wise).

The classification of the efficiency scores in window-1 for overall efficiency is performed based on the optimal k-value, as shown in **Figures 9(a)** and **9(b).** A similar approach is taken to calculate the same for window -2, -3, and -4.

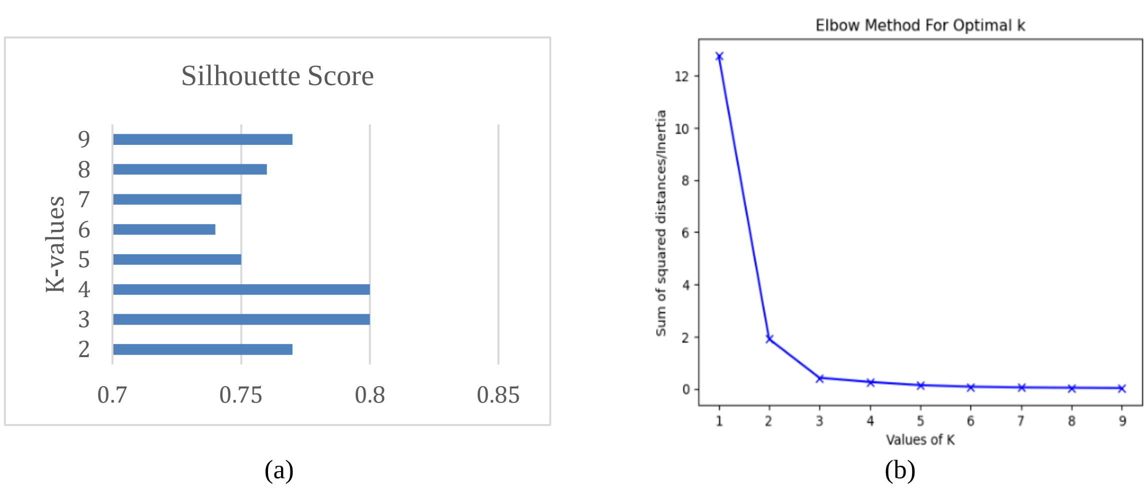
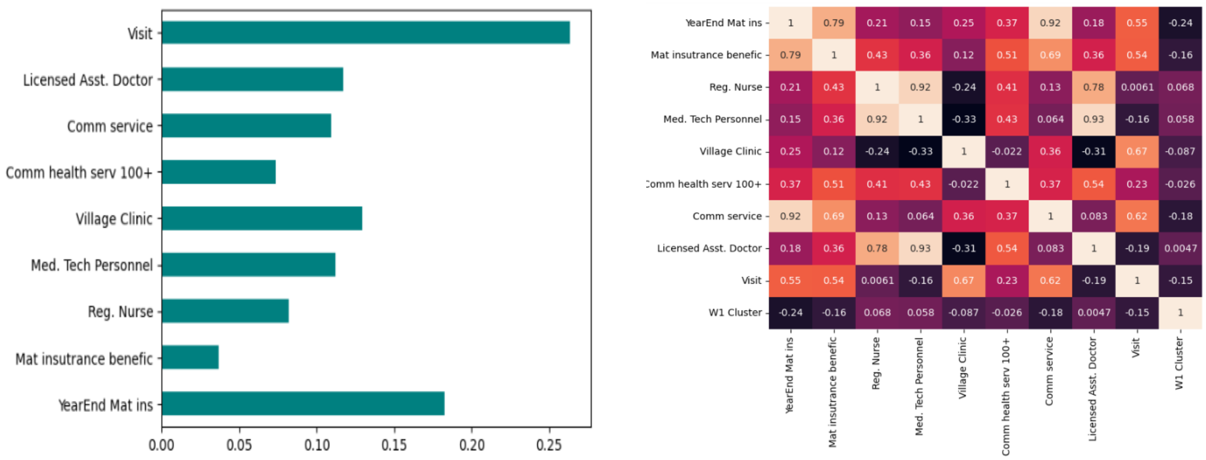


Figure 9. Selection of optimal K value.

The next step in the prediction model is to list all the variables in each of the five constructs. In order to keep only the relevant health factors, yet not redundant, the information gain value and the correlation coefficients for window-1 are shown in **Figures 10(a)** and **10(b).**

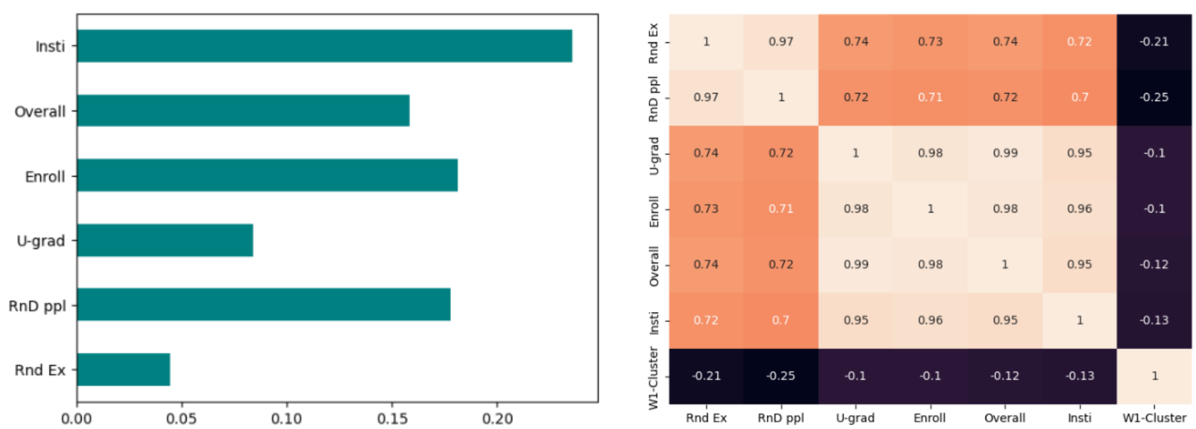


(a) (b)

Figure 10. Information gain and correlation coefficients of medical factors.

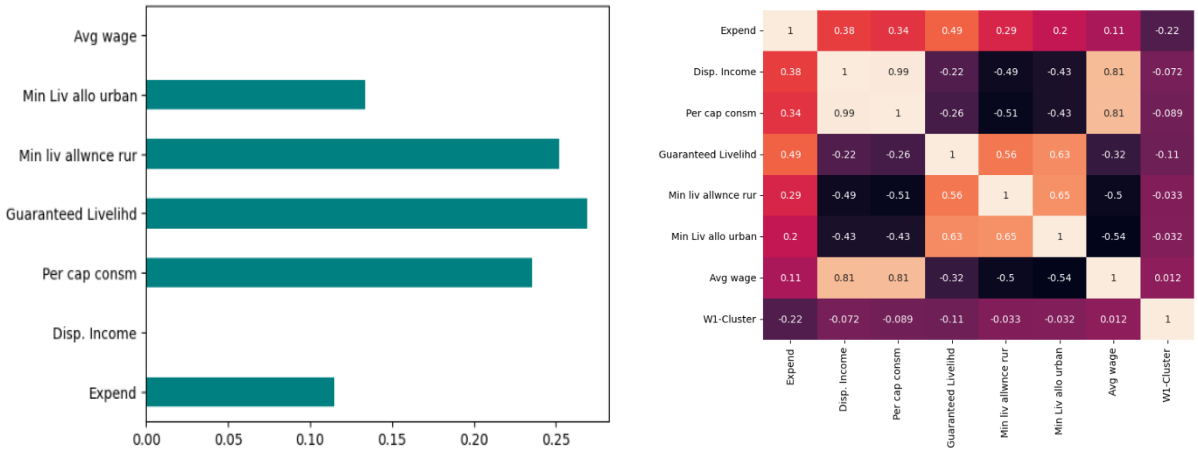
If we follow the two-step validation method, as discussed earlier for feature selection, the first choice should be "Visit". As per the IG value, Year-end-Maternity insurance factor should get selected. However, it is highly correlated with "Visit" and thus, is dropped. In a similar way, if we progress, licensed assistant doctor and technical personnel are selected. The similar analysis for window -2, -3, and -4 is given in **Figures 8, 9**, and **10** in Appendix C.

Interestingly, all the variables in the Education construct for window-1 are highly correlated (**Figure 11(b)**). Therefore, despite the high predictive power of the individual variables (**Figure 11(a)**), only the one with the largest IG value "Insti" gets selected. The same analysis for windows -2, -3, and -4 can be found in **Figures 11, 12,** and **13** in Appendix C.



(a) (b)

Figure 11. Information gain and correlation coefficients for educational factors.



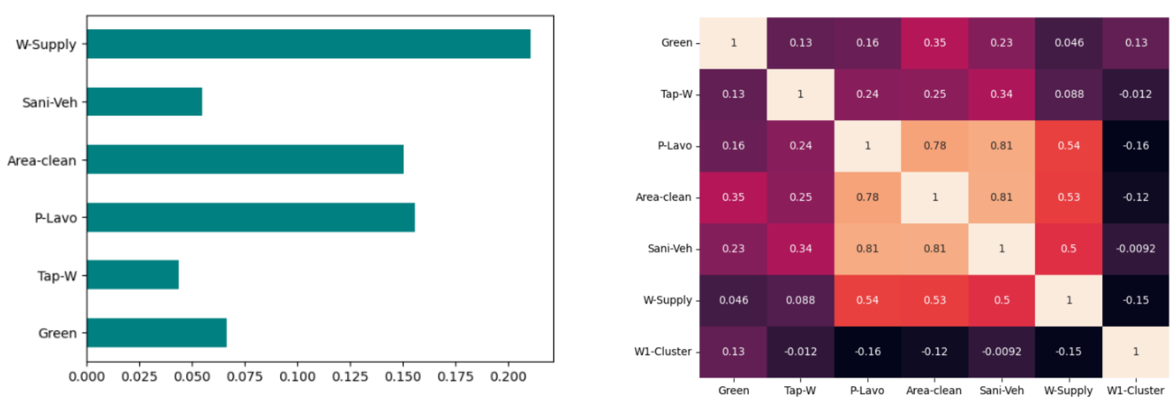
(a) (b)

Figure 12. Information gain and correlation coefficients for wealth indicators.

The construct "Wealth" has Guaranteed Livelihood with the largest IG value **Figure 12(b).** However, it is highly correlated with two variables: minimum living allowances in rural and urban areas. Thus, the last choice is per capita consumption and expenditure. The same analysis for windows -2, -3, and -4 are available at **Figures 17, 18**, and **19** in Appendix C.

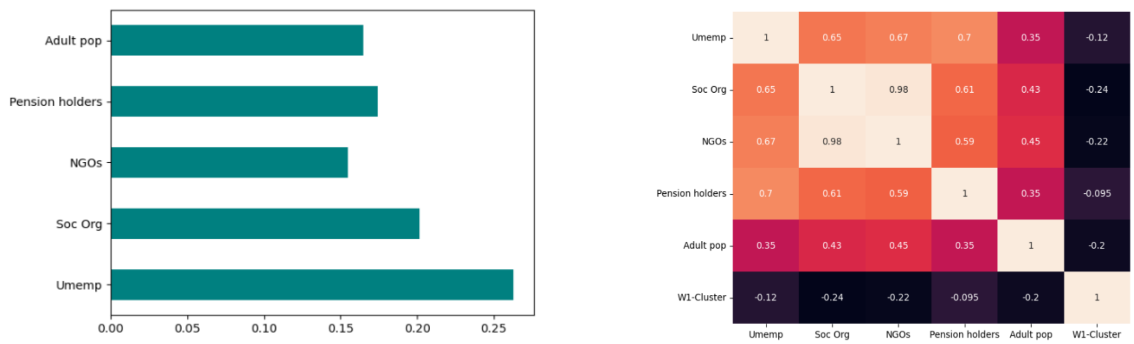
The construct "Public Health" has water supply as the variable with the highest IG value (**Figure 13(a)**). Both variables, public laboratories and area under cleaning program have very high correlation (**Figure 13(b)**). Thus, the next two best variables, green areas and number of sanitization vehicles, are included. The same analysis for windows -2, -3, and -4 is shown in **Figures 14, 15** and **16** at Appendix C.

For miscellaneous factors, the number of unemployed adults tops the information gain list (**Figure 14(a)**). However, the next two variables on the list-Social Organizations and Pension holders-are highly correlated. Thus, the only variable left with low correlation is adult population **Figure 14(b).** For the remaining three windows, refer to **Figures 20, 21** and **22** in Appendix C. The variables that are finally eligible for being considered in the classification method across the windows are shown in the **Table 3**.



(a) (b)

Figure 13. Information gain and correlation coefficients for public health.



(a) (b)

Figure 14. Information gain and correlation coefficients for miscellaneous factors.

Table 3. Window-wise explanatory variables corresponding to five external factors.

	Medical indicators	**Education**	**Public health**	**Wealth**	**Misc.**
W-1	No. of Visits	Insti	Water Supply	Guaranteed Livelihood	Unemployed
	Med. Tech. Personnel		Tap Water	Per Cap Consumption	Adult pop
	Lisc. Assts. Doctors		Green Area	Expenditure	
W-2	No. of Visits	Insti	Public Lavatory	Min. Liv allow. Rur	Unemployed
	Community serv. 100+		Tap Water	Disposable income	Adult pop
	Lisc. Assts. Doctors		Green Area	Expenditure	
W-3	No. of Visits	Insti	Public Lavatory	Min. Liv allow. Rur	Unemployed
	Med. Tech. Personnel		Tap Water	Expenditure	Adult pop
	Lisc. Assts. Doctors		Green Area		
W-4	No. of Visits	Insti	Water Supply	Min. Liv allow. Rur	Unemployed
	Med. Tech. Personnel		Tap Water	Average wage	
	Lisc. Assts. Doctors		Green Area	Expenditure	

6. Classification Analysis

This study employs machine learning algorithms to identify the appropriate classification of the regional medical services based on the external variables, not on input and output values. The aim is to highlight the actual predicting power of the model depending solely on external variables, without "feeding" the actual efficiency score to the model. The correlation values between the selected set of explanatory variables from the five types are kept to the minimum level. In case of high correlation, we have considered the one with larger information gain and dropped the other.

The classification techniques reveal interesting interdependencies between the contextual factors and the overall efficiency of the regions. As the number of datasets is limited, we have used k-fold cross validation technique to boost up the training dataset for the classifiers, especially logistic regression (LR). As we set out to explore the interlinkages of the contextual variables, **Figure 15** shows the variables that have real influence on the efficiency levels of the regions and in which direction it unfolds. It is important to note that the sign of the explanatory variables do not imply positive or negative impact on the health outcome, but rather on the productivity of the health outcome. For example, medical technical personnel and Green area are the two factors which will reduce the per unit productivity in the health sector. The Adj-R^2 values of the window-based logistic regression are 0.54, 0.59, 0.55, 0.58, respectively.

There are a few eminent patterns in the series of efficiency models which are listed below.

(i) The sets of significant variables in each window are distinctly different, with very little repetition. No. of visits has a significant influence only in window-2. Similarly, variables such as green area, lisc. Assistant doctors, tap water, and guaranteed livelihood significantly influence the efficiency for a single window and never appear henceforth. This points to the fact that improving performance in the areas of concerns led to a different set of problems in other areas. Thus, a holistic approach of understanding the primary determinants is key to long-term efficiency.

(ii) The number of significant variables is gradually reducing with years. However, there is no significant drop in the R^2 values across the windows. If we ignore the effect of relatively small sample size, then this only refers to the fact that most of the regions have improved some of those sources of inefficiency and now there are only a few selected parameters that separate the efficient regions from the rest. As a result, the effect-size of these parameters have increased considerably.

(iii) Out of the twelve independent contextual variables, only five variables, e.g., Medical technical personnel, Green area, Guaranteed Livelihood, Per Capita Consumption, Unemployment, Licensed Doctor and Tap water have statistically significant level (at or less than 8% $p - value$) of influence over the efficiency level. The results reveal a great deal of resources related to overutilization or underutilization of resources in the sector depending on whether the effect size is negative or positive, respectively. Guaranteed Livelihood and Number of visits exhibit increasing return-to- scale. So, increasing these two variables results into improvement of the outputs at a much higher rate. However, most of the other variables e.g., institutions, medical technical personnel, tap water, and green areas exhibit decreasing return-to-scale. If the values of these variables are increased, then the output levels would improve but at a lower rate, thus overall productivity will deteriorate.

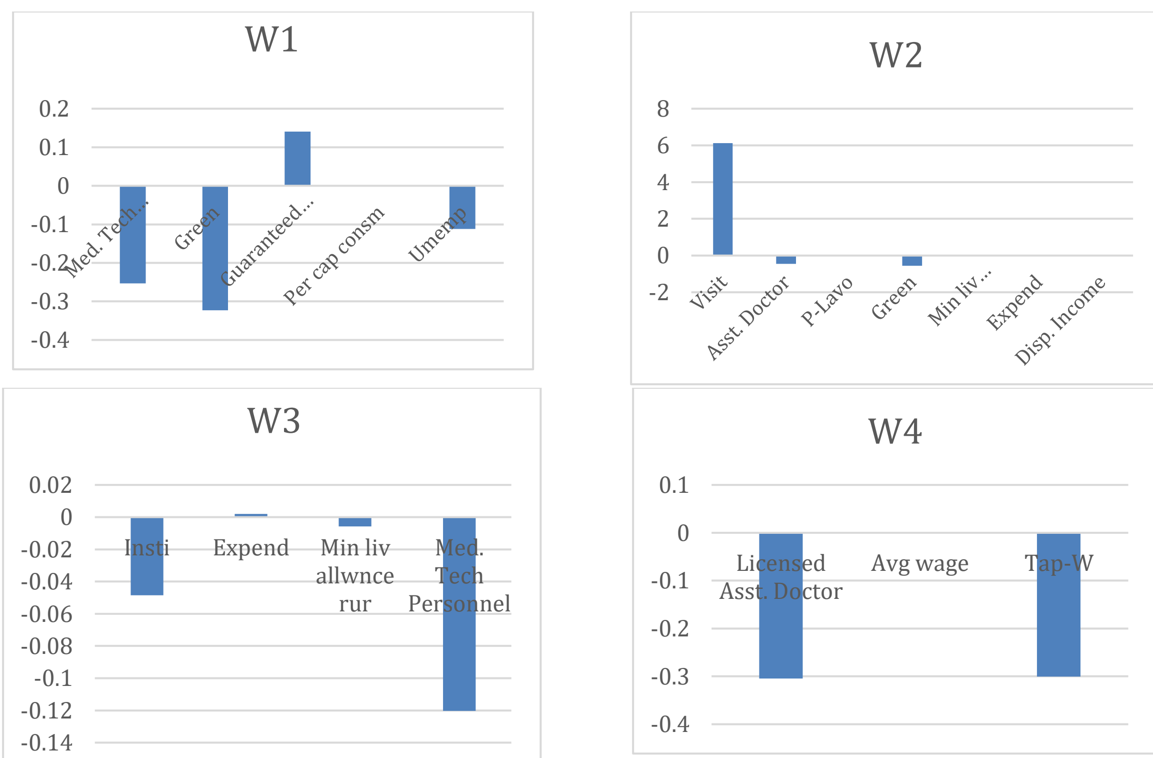


Figure 15. Impact of external factors on health outcomes (Logistic regression analysis).

Guaranteed livelihood has a positive contribution towards productivity. If the number of people under centralized livelihood guaranteed increases by 10,000, it will make the regions 15% more likely to become a higher efficiency region. The regions which are trying to get out of the low- productivity trap need to focus on number of visits, technical personnel, Green Area, Guaranteed livelihood (all significant at 5% significance level). In this analysis, per capita consumption is ignored for its low effect size. The influence of number of visits has overshowed the other variables as it exponentially increases the odds of being efficient. Also, the number of citizens covered under the guaranteed livelihood scheme can more than double the likelihood of being efficient for every ten thousand people brought under the scheme. Thus, guaranteed livelihood scheme has proven to be a booster for improving the health outcome of any region irrespective of its efficiency.

The dominance of literacy on the public health indicators has been established in previous literature (Wu et al., 2015). In this paper, a wide range of variables related to literacy has been considered and number of institutions turns out to be the most influential one. Secondly the number of technical personnel is shown to have played a significant role in determining the overall performance (Du, 2017). In our study, we have identified specific variables driving efficiency in each of the four windows e.g., Medical technical personnel, Asst. doctors, Medical technical personnel, and Licensed asst. doctors in the four windows respectively. In decision tree analysis, one can see number of visits is an influential factor (Zhang et al., 2015) showed a similar trend where number of visits and bed utilization strongly determined the efficiency scores. Since number of beds is used as an input in our study, in output-oriented model we can’t associate inefficiency with bed utilization.

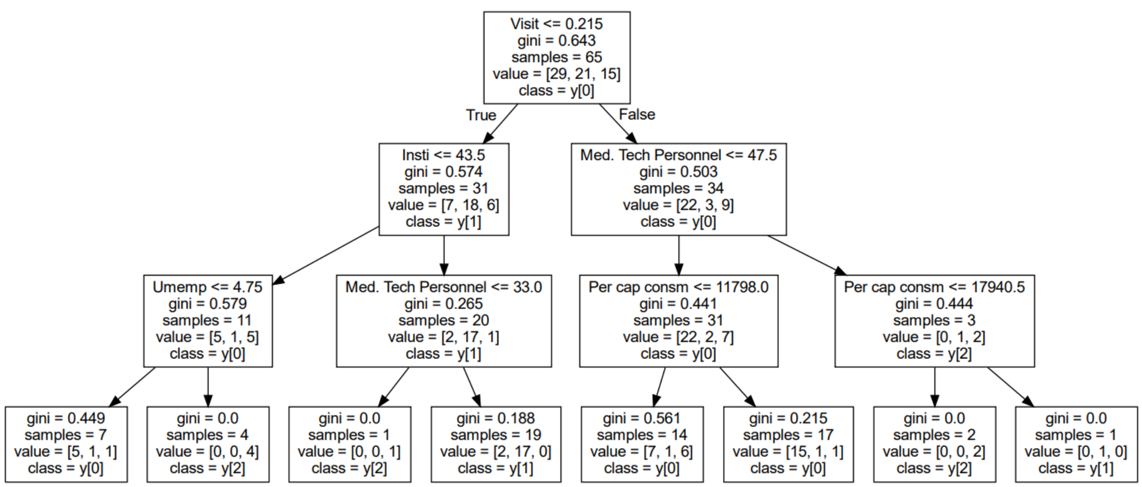


Figure 16. Decision tree for Window-1.

Some interesting trends arise from the series of decision trees for window -1, -2, -3, and -4.

(i) The root nodes in each of the four windows, i.e., visit, assistant licensed doctors, and medical technical personnel in medical sector, are continuously changing to new values. The variables at the root node or closer to it represent a more impactful classification factor, compared to the variables appearing in the leaf nodes or closer to leaf nodes. Thus, **Figures 16, 17, 18** and **19** reveal that the roles of variables in determining the efficiency levels of the regions are also shifting.

(ii) The factors at the leaf nodes or at a deeper level in the tree are associated with specific scenarios conditioned on the status of the variable appearing at a higher level in the tree. It can be seen that there is more agreement among the variables as the tree grows in depth. It implies that the region-specific characteristics of healthcare efficiency are more or less stable across the windows.

(iii) In most of the cases, the gini-impurity values attain considerably smaller values in the leaf-level nodes. In majority of the times, it shows that three variables are sufficient to explain the higher efficiency values.

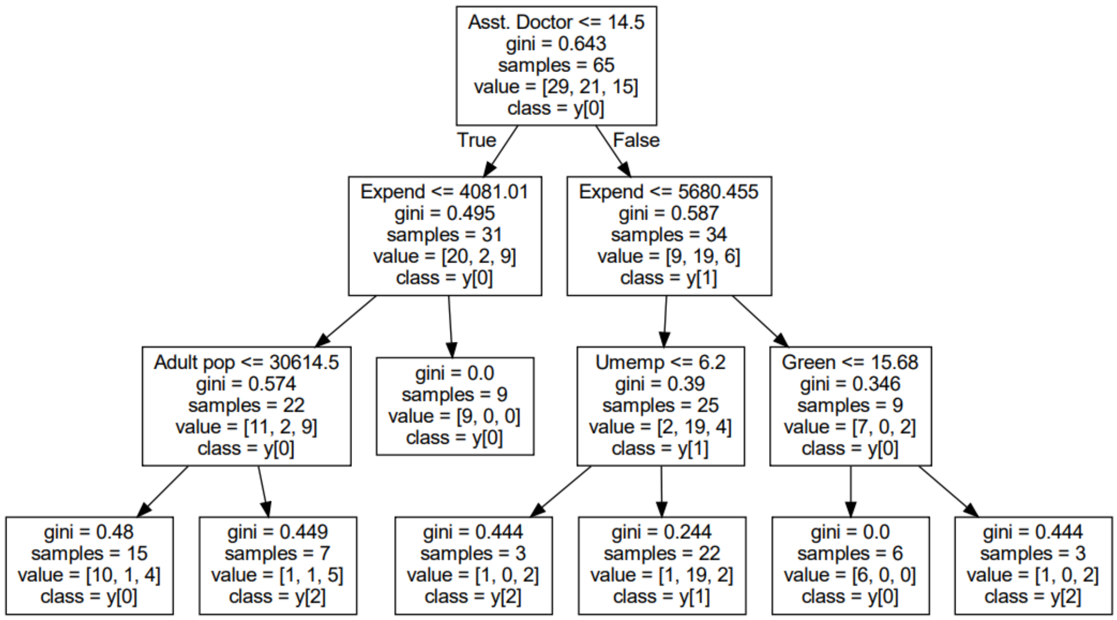


Figure 17. Decision tree for Window-2.

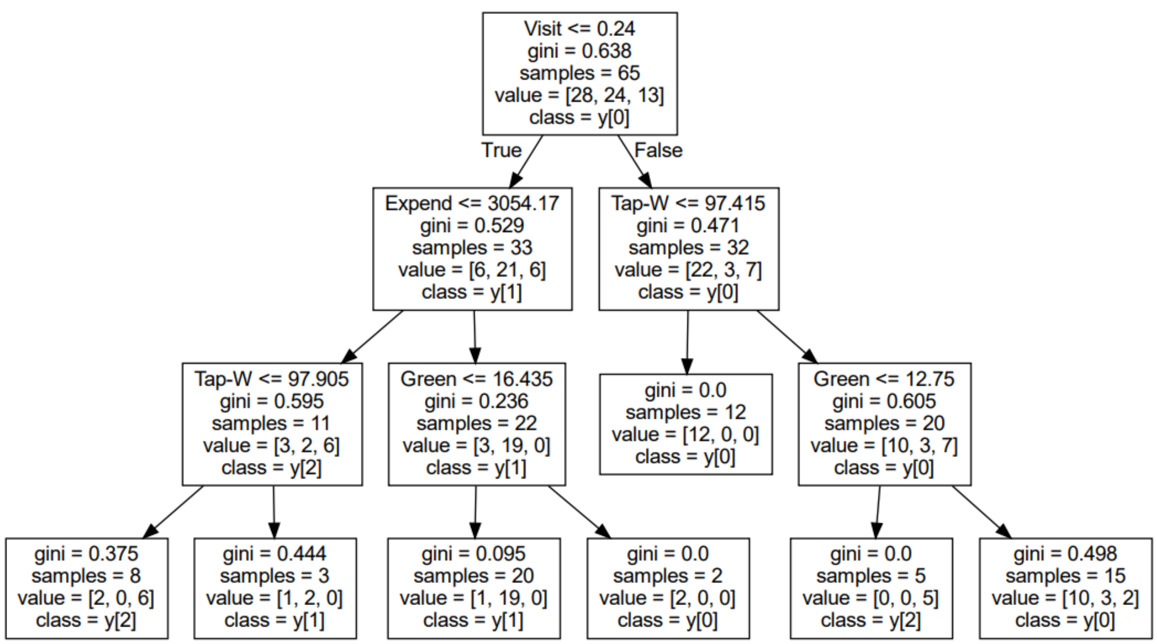


Figure 18. Decision tree for Window-3.

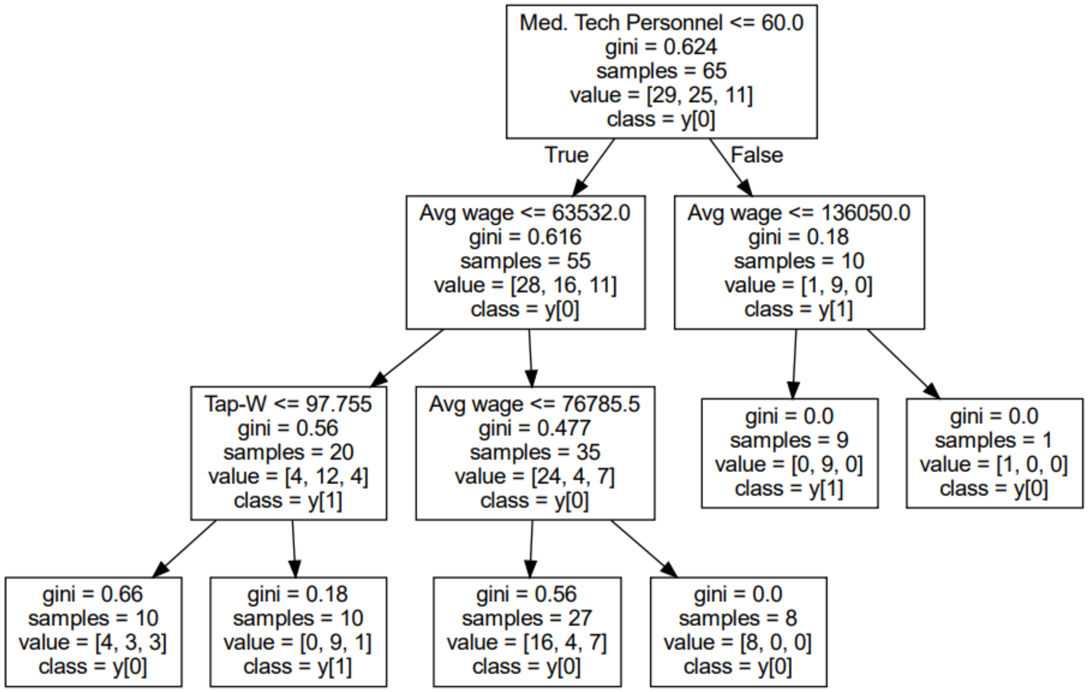


Figure 19. Decision tree for Window-4.

The key advantage of using linear Support Vector Machine (SVM) method, despite its lower accuracy, lies in its ability to explicitly extract information about the features. In a linear SVM, the weight vector signifies the individual contributions of the respective explanatory variables towards the classification. The highest absolute value of the weight, $|w_i|$, is assigned to feature i, representing its highest contribution (W-map method). Similarly, we have selected the top-5 feature set in each window of analysis based on the absolute weight values in **Table 4**.

Table 4. Relative importance of external variables (SVM approach).

Rank	**1**	**2**	**3**	**4**	**5**
w1	Institution	Min Liv.allowance	Green Area	Unemployed	Med. Technical personnel
w2	Institution	Med. Technical personnel	Guaranteed Livelihood	Green Area	Licensed Asst. Doctor
w3	Institution	Med. Technical personnel	Unemployed	Min Liv.allowance	Licensed Asst. Doctor
w4	Unemployed	Med. Technical personnel	Green Area	Institution	Tap-W

There are some interesting patterns emanating from **Table 4** as listed below:

(i) There are only a handful features which appears to be important e.g., institution, medical technical personnel, minimum living allowance, licensed doctor etc. This result is very similar to the one obtained using LR.

(ii) There is more amount of convergence in the top-ranked features than the bottom-ranked features.

(iii) Note that the list is based on individual contributions, and it may not represent the best set of variables for predicting the classification. For selecting the optimal feature set, various wrapper methods, such as RFE W-map, can be employed.

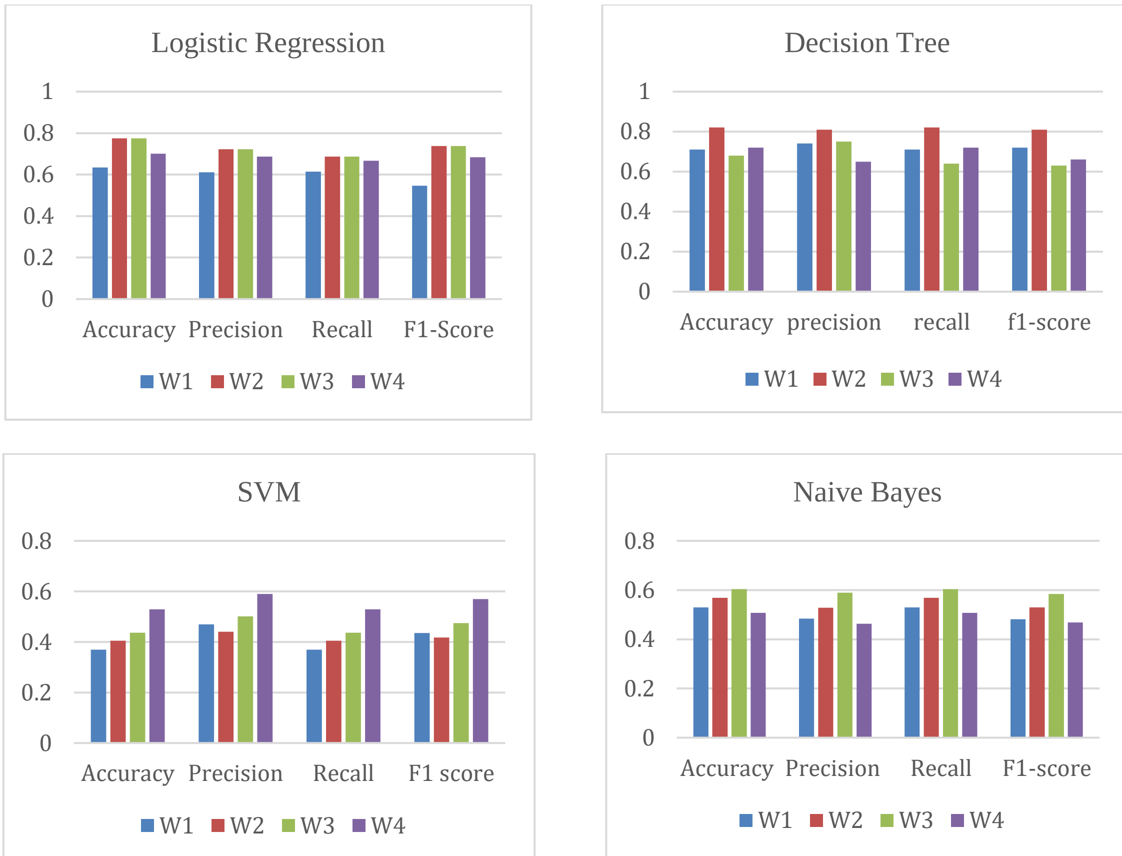


Figure 20. Window-wise accuracy levels of four classifiers.

Figure 20 shows the performance of different algorithms across different evaluation measures e.g., accuracy, precision, recall, F1-score. It is clearly seen that there all the four classification algorithms perform similarly in all the four evaluation metrics across the four windows. SVM shows a continuous improvement every year in all four metrics.

LR and DT models have equivalent accuracy levels across windows and both outperform NB and SVM (see **Figure 21**). NB model in turn outperforms SVM in terms of accuracy. In real life scenarios, contextual variables are rarely independent of each other, and proof of non-collinearity cannot be translated into independence. Usually, under such circumstances, SVM generally outperforms NB. Given the nature of the dataset, there could be several reasons behind this phenomenon. The usage of "linear" SVM restricts its full potential. The reason for choosing linear SVM model is to extract the relative importance of the individual explanatory variables, which is the primary objective of the second-phase of this study. In cases where a non-linear and rough decision boundary is more likely, as in the considered case study, SVM's performance might be compromised. Another possible reason could be that the non-collinearity assumption is weakly held in some of the prediction models, and Linear-SVM is more prone to such perturbances than NB. Thus, despite a discriminative approach for classification, SVM lags behind NB, which bases its prediction on existing examples without facing specific restriction apart from a limited set of training examples, partly handled with k-fold cross validation techniques. Consequently, a generative technique like NB comfortably outperforms SVM.

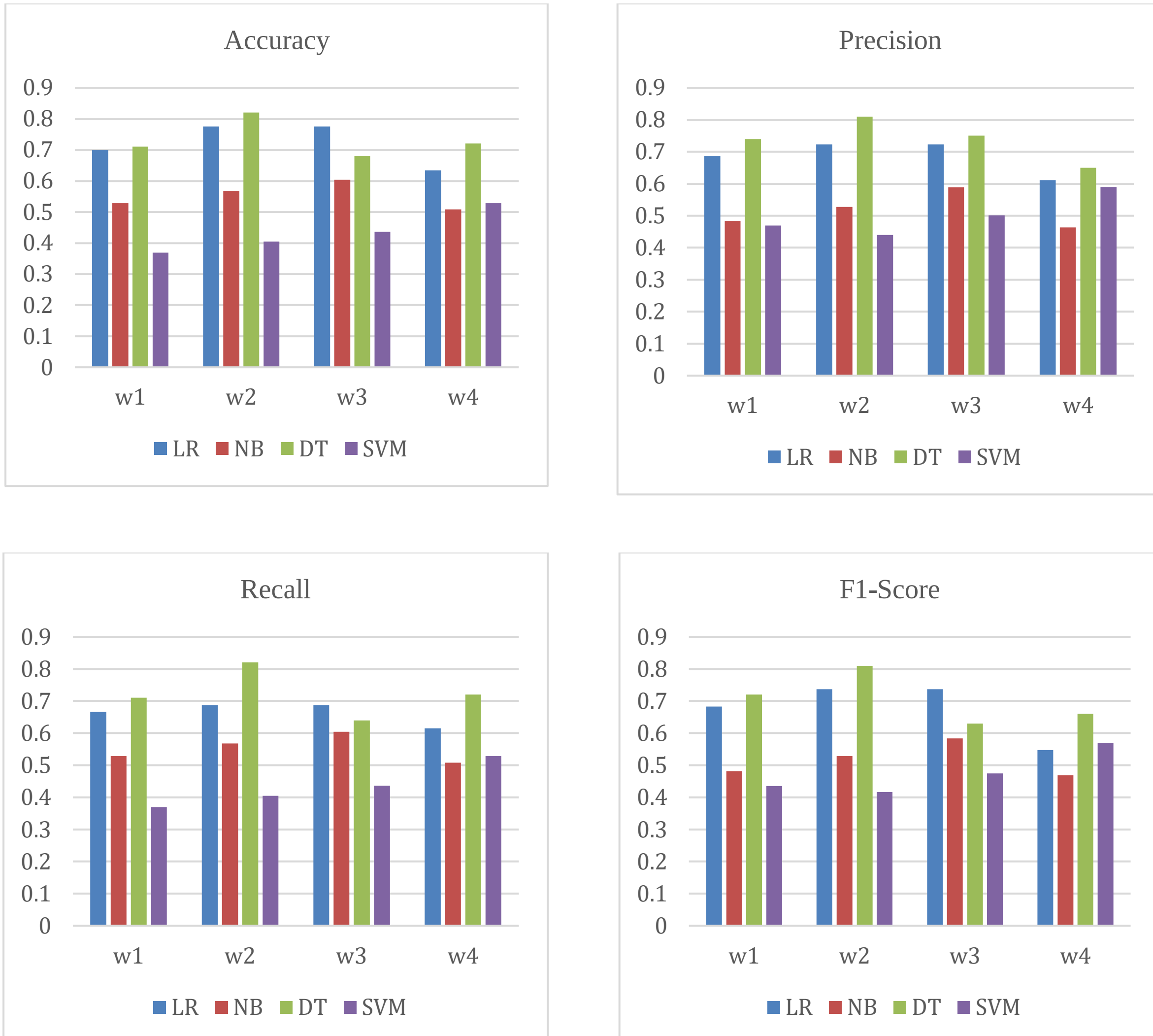


Figure 21. Comparison of accuracy, precision, recall, and F1-score of the classifiers.

In this analysis, DT is slightly better than LR, precisely due to two reasons. First, a decision tree is free from collinearity or linear assumptions, without which the accuracy of LR suffers, and the results become unreliable. Second, outliers in the data weaken the LR model. A decision tree thus has a natural advantage over LR model. However, the primary advantage of LR is that it reports significance levels of the explanatory variables, which can't be accessed in any other methods undertaken in this study. On the other hand, DT manages to provide a slightly better result as compared to LR. This is due to the fact that while handling continuous input data, it is prone to lose information about classification trends. Both LR and DT models convincingly outperform the NB model. The key advantages of using NB model lie in its linear learning time in the training sample size, and in its prediction time, which is independent of the training sample size (Pazzani and Billsus, 1997). Thus, the model is fast in learning as well as in prediction. The accuracy can be significantly improved with a larger dataset (Han and Kamber, 2001).

8. Conclusion

In this study, the interdependency between efficiency and effectiveness is revisited using the DEA framework. In principle, this conflicting relationship is pervasive in most of the managerial fields, including supply chain management, retail management, and production planning. Thus, this study is valuable to both academicians and practitioners alike. Addressing this conflict and striking a fair balance is even more crucial, particularly for healthcare managers given that human lives are at stake. Broadly, we make the following four policy recommendations:

- The healthcare sector should emphasize more on the quality aspect and integrate it with the overall efficiency rather than focusing on minimizing cost.
- A few provinces are suffering very low efficiency scores in service quality and operational efficiency should perform more surgeries and healthcare exams to check the high mortality rates in those provinces.
- There has been a continual and gradual improvement in the external factors directly related to public health. However, the solution mechanism of the concerned authorities is not holistic enough as one can see newer issue factors appearing subsequently and deteriorating the overall performance.
- Concerned decision-makers can choose a set of selected externalities in order to improve the overall performance faster. To achieve this, the external factors are categorized in two classes: one that improve the healthcare outputs at a much faster rate. And another that improve the healthcare outputs but at much lower rate than the first set.

Usually, statistical analysis to investigate the relationship between two constructs is popular in relevant literature. However, non-parametric techniques offer the flexibility of not fixing the production function a priori. This indicates that reaching the Pareto-efficient frontier requires different strategies for different units. Therefore, the results are more relevant to the practicality of observed datasets. In addition, DEA offers a multitude of insights about the overall situation. Along with assigning efficiency scores, it also sets benchmark for the poor performers to follow and achieve full efficiency. The underlying LP formulation of DEA can be exploited to determine the criticality levels of resources and their utilization rate.

The scope of this study is limited to six years, spanning from 2013 to 2018, covering two five-year-planning periods. The developed method enables each province to introspect and decide the best possible strategies to improve its medical care. While the literature discusses the conflict between quality and efficiency in different contextual setup, this paper follows the unique ideation of evaluating the trade-off by encapsulating both aspects in the same DEA framework. Trade-offs among multiple decision-choices have been handled using explicit functional form or by game theory. However, the proposed model explores this trade-off in a more delicate way where there is no presumption on the relation among the decision-choices. As a result of this flexibility, the proposed model can accommodate trade-off analysis in different industry or organizational sectors. For example, delivery time vis-à-vis safety concerns in food-delivery sector can also be studied using our model. Similarly, efficiency and equity are one of the major trade-off faced by the policy-makers. We have also shown in the paper that multiple factors can be analyzed simultaneously in case more than two constructs are involved. The developed model can be extended to carry out window analysis for measuring stability and consistency or by using more sophisticated techniques like Malmquist Index to identify annual changes in productivity in healthcare.

The second phase of the analysis includes the usage of Random Forest in developing a predictive model. Cautions need to be taken while selecting more sophisticated training models and determining the required level of complexity. While these advanced techniques work successfully with a small dataset, they might require different sets of resources. The use of simpler models like decision trees also leads to rule-based classification, which is easy to interpret and follow for decision-makers. The analysis can be refined and validated by testing it against the results obtained with various machine learning algorithms, such as artificial neuron network, support vector machine, etc.

This study has some limitations in the way it has defined the quality measure. Although we have followed medical protocols, providing the exact breakdown of the surgeries into specific categories, revealing the complexity of each procedure, would ensure a more comprehensive assessment. Several other proxy measures, such as patient fall, non-clinical treatment, and cleanliness might have provided more insights into the evaluation. However, the paper focuses on analyzing the available quantifiable data on a large scale, without any prejudiced view about the service productivity function.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

AI Disclosure

The author(s) declare that no assistance is taken from generative AI to write this article.

Acknowledgments

We would like to express our deepest gratitude to all the reviewers for their constructive review comments which significantly helped us improve our manuscript.

Taxonomy

- **Productivity:** It is the ratio of output to input.

- **Efficiency:** It is the ratio of the productivity of a unit to the highest productivity (also known as relative productivity).

- **Quality:** It is usually measured in terms of customer satisfaction. In this paper, one key component of quality is measured using three distinct factors that to a large extent determines customer satisfaction and effectiveness of healthcare sector.

- **Decision-making units (DMU):** In a homogeneous set of production units where each unit controls its inputs and outputs independently of the others, the individual units as often referred to as decision-making units or DMU.

- **Impurity/Entropy:** Entropy (e) is the expected value of surprises ($\mathrm{e} = -\sum \mathrm{p} \log \frac{1}{\mathrm{p}}$) where surprise is proportional to the unlikeliness of an event ($\mathrm{surprise} = \log \frac{1}{\mathrm{p}}$, where p is the probability of an event).

- **Information gain:** When a dataset is split based on some criteria on a particular relevant feature, then there is a drop in the entropy, which is known as information gain. The larger the drop, the more useful is the feature.

Decision Tree Versions

- **ID3:** The most basic version of DT that greedily chooses the features with the largest information gain at each node.

- **C4.5:** An enhanced ID3 algorithm that uses gain ratio to avoid biases present in ID3 algorithm.

- **CART:** It considers complexity along with the information gain.

- **Accuracy:** The proportion of all correct measure within the total dataset.

$$\text{accuracy} = \frac{\text{True Positive+True Negative}}{\text{Total dataset}}.$$

But many a time, the cost of a mis-classification depends whether the observed sample is positive or negative.

- **Precision:** Among all the predicted positives, what fraction is actually positive.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False Positive}}.$$

- **Recall:** It measures out of all the actual positive samples, what fraction is actually positive.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive+False negative}}.$$

- **F1-Score:** To maintain a balance between precision and recall (individually they can sometimes be misleading), the harmonic mean of precision and recall is treated a F1-score.

$$\frac{1}{\text{F1}-\text{score}} = \frac{1}{2}\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right).$$

Appendix (A)

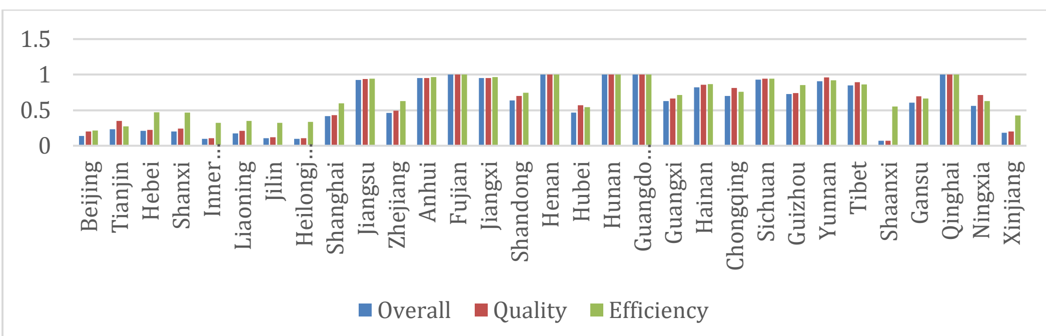


Figure 5. Window-2 average efficiency of regional health sectors.

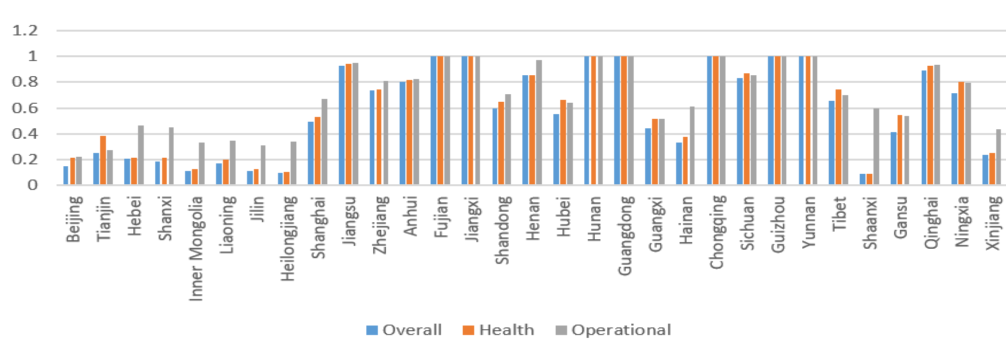


Figure 6. Window-3 average efficiency of regional health sectors.

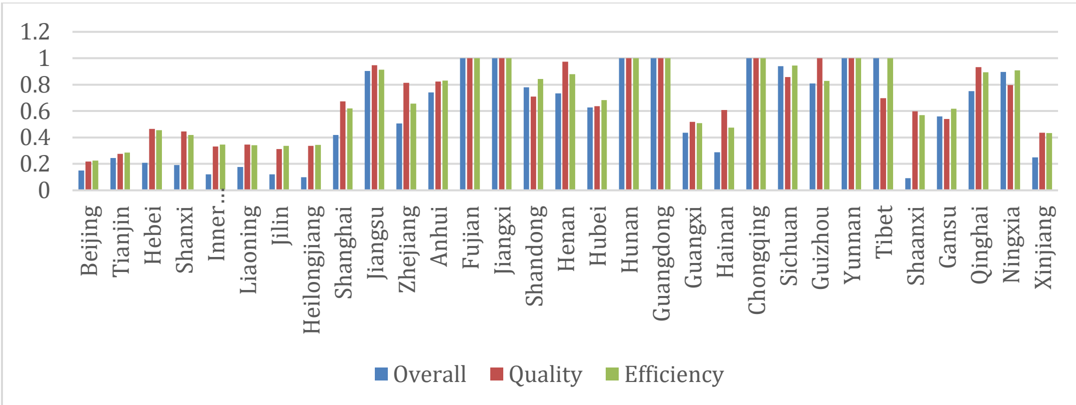


Figure 7. Window-4 average efficiency of regional health sectors.

Appendix (B)

Table 10. Summary statistics of public health measures.

	Green Area				Tap water access Urban pop			
	w1	w2	w3	w4	w1	w2	w3	w4
Average	12.68151	12.9843	13.21667	13.4228	97.36989	97.22828	97.47828	97.56419
Min	7.1	7.33	5.85	5.85	88.06	67.57	67.57	67.57
Max	19.28	19.77	19.77	20.38	100	100	100	100
Std Dev	2.62986	2.706728	2.833135	2.885726	2.56333	3.957429	3.799293	3.826089
Q1	11.005	11.145	11.5	11.63	96.355	96.72	96.76	96.67
Median	12.11	12.48	12.57	13.08	98.08	98.1	98.46	98.4
Q3	14.085	14.36	14.67	14.83	99.455	99.505	99.475	99.52

	No of public lavoratories				area under Cleaning prog.			
	w1	w2	w3	w4	w1	w2	w3	w4
Average	4013.925	4092.172	4217.699	4444.817	22069.25	23670.42	25454.88	26949.45
Min	285	303	317	317	1739	2413	976	976
Max	11739	12136	12934	13965	92161	101629	107035	112966
Std Dev	2701.9	2768.259	2842.162	2970.491	18485.05	19252.71	20527.75	21694.81
Q1	1896	1775.5	1790.5	2093	12036.5	13022.5	14583.5	14738.5
Median	3370	3550	3586	4138	16083	17380	18253	19488
Q3	5908.5	5815	5806	6088.5	26364	29353	30642.5	33635

	Env. Sanitation vehicle				Water su[pply			
	w1	w2	w3	w4	w1	w2	w3	w4
Average	4663.527	5388.151	6319.204	7252.097	197.6548	196.1151	195.5624	194.6151
Min	156	251	332	332	23.8	24.1	25.7	25.8
Max	14274	19314	20472	23213	591.3	591.3	591.3	592
Std Dev	3259.506	3746.01	4337.515	4769.095	146.2034	145.1726	144.0647	143.2975
Q1	2465	2659	3351.5	4169.5	82.2	79.75	78.25	77.45
Median	3937	4622	5462	5944	183.2	182	181.1	179.5
Q3	6132	6966.5	7721	8581	280.2	277.05	283.45	283.45

Table 11. Summary statistics of medical factors.

	No. of maternity insunce beneficiaries				registered Nurses				Medical Technical Personnel			
	w1	w2	w3	w4	w1	w2	w3	w4	w1	w2	w3	w4
Average	19.11011	23.32226	28.69226	33.49527	13.82796	14.82796	15.88172	17.15054	41.37634	43.11828	45.02151	47.63441
Min	0.47	0.54	0.65	0.91	5	5	5	5	26	28	29	29
Max	95.28	170.3	170.3	189.32	30	32	32	32	86	90	93	95
Std Dev	19.51145	26.59955	33.67131	39.25166	4.666568	4.891563	5.088147	5.043779	12.75401	13.61676	14.29112	14.66529
Q1	6.39	7.625	8.83	10.94	11	11	13	14	33	34	36	38.5
Median	12.76	14.05	18.45	23.51	12	14	15	17	38	39	40	43
Q3	22.385	27.735	31.755	34.055	16	16.5	18	19	44.5	47	48	50.5

	Village clinic				Community health services 100+ beds				Community health services			
	w1	w2	w3	w4	w1	w2	w3	w4	w1	w2	w3	w4
Average	20802.42	20696.44	20552.22	20352.91	8.677419	8.752688	9.365591	10.23656	279.172	283.7957	288.9355	294.8065
Min	1271	1218	1187	1162	1	1	1	1	7	7	7	8
Max	62311	61450	60492	60371	61	59	58	54	1078	1082	1098	1111
Std Dev	16984.27	16930.48	16847.35	16715.69	12.31501	12.09964	12.57732	13.1594	200.8232	202.4029	204.2139	206.1671
Q1	10503	10330.5	10200.5	10127	1.5	1	1	2	165	168	175.5	185
Median	15523	15475	15391	15331	4	4	4	4	245	245	248	249
Q3	27638	27031.5	26672.5	26227.5	10	10.5	12.5	13	381	397	397.5	398

	Licensed Asst. Doctor				Year-End Maternity Insurance Contributio				Visits in Township centers 100 mil			
	w1	w2	w3	w4	w1	w2	w3	w4	w1	w2	w3	w4
Average	17.17204	17.66667	18.3871	19.45161	550.556	572.6955	597.0132	625.6487	0.337527	0.345806	0.355269	0.362366
Min	10	10	10	11	20.69	22.82	23.79	24.87	0.03	0.03	0.03	0.03
Max	43	42	42	47	3081.8	3161.89	3300.89	3495.34	1.05	1.12	1.12	1.12
Std Dev	7.00019	7.392721	7.718198	8.181432	537.8804	564.2553	593.536	620.0901	0.283191	0.291608	0.298669	0.30642
Q1	13	13	14	14	256.01	267.51	285.615	300.49	0.1	0.1	0.11	0.11
Median	15	15	16	17	366.95	367.53	370.13	411.33	0.22	0.22	0.23	0.23
Q3	18	19	19	21	701.005	713.015	724.24	775.27	0.5	0.51	0.515	0.54

Table 12. Summary statistics of educational factors.

	RnD Expenditure				Full time RnD personnel person-year				No of grads with Deg or Dip			
	w1	w2	w3	w4	w1	w2	w3	w4	w1	w2	w3	w4
Average	2966300	3248694	3545328	3861553	83589.59	85831.83	86849.71	90537.27	21.27946	21.98312	22.80527	23.58398
Min	2602	2602	2602	3186	43	43	43	202	0.91	0.91	0.9	0.9
Max	15205497	16762749	18650313	21072031	441304	451885	457342	621950	48.41	50.91	57.12	58.59
Std Dev	3711471	4098000	4531702	4954964	108626.3	111838.9	115132.8	125609.5	13.46851	13.81741	14.29989	14.85287
Q1	568080	680717.5	745902.5	745902.5	16215	16773.5	17976	16664.5	11.985	12.17	12.675	13.255
Median	1688902	1960112	2238051	2420637	45809	47392	47392	49254	18.27	18.99	19.64	19.97
Q3	3286803	3578086	4001187	4552825	91796	95134	95660.5	99061	29.98	31.225	32.205	33.14

	Enroll in undergrad courses				Overall student enrollment				Regular insti of higher education			
	w1	w2	w3	w4	w1	w2	w3	w4	w1	w2	w3	w4
Average	49.59355	50.86817	52.02484	53.32204	82.16194	84.6114	86.8257	89.03753	81.50538	82.63441	83.73118	84.83871
Min	2.14	2.22	2.32	2.39	3.35	3.35	3.42	3.5	6	6	6	7
Max	104.08	107.68	110.58	114.08	190.06	199.59	201.53	214.08	162	166	167	167
Std Dev	28.11549	28.82746	29.55296	30.38434	49.95699	51.54781	53.08506	54.65327	40.08019	40.36255	40.64087	40.91883
Q1	28.96	29.665	30.935	32.72	47.515	50.335	51.225	51.47	55	55	56	56.5
Median	47.78	49.18	49.74	50.29	71.79	73.59	75.1	76.28	80	81	81	82
Q3	68.5	70.18	70.295	70.61	113.35	115.465	117.99	122.055	117.5	118	118.5	119

Table 13. Summary statistics of wealth indicators.

	Local Govt. Expenditure				Disposable income				Per cap consumption exp.			
	w1	w2	w3	w4	w1	w2	w3	w4	w1	w2	w3	w4
Average	4293.457	4730.135	5203.391	5610.495	20097.29	21934.55	23876.49	25961.11	14633.38	15901.01	17144.94	18501.88
Min	922.48	1000.45	1138.49	1254.54	9740	10730	12254	13639	6307	7317	8246	9319
Max	12827.8	13446.09	15037.48	15729.26	49867	54305	58988	64183	34784	37458	39792	43351
Std Dev	2110.044	2365.057	2640.605	2863.02	8346.213	9018.652	9767.505	10594.79	5850.918	6188.473	6490.298	6842.792
Q1	3075.73	3370.38	3741.07	3895.7	15172	16765	18351.5	19936	11141	12274.5	13189.5	14452.5
Median	4020.66	4389.37	4617.4	4908.55	17395	18808	20420	22034	12769	13811	14900	16160
Q3	5178.495	6039.11	6642.58	7116.265	22506.5	24351.5	26126	28105.5	16217.5	17422	18898	20010

	Guaranteed Livelihood				Min liv allowance ppl in rural			
	w1	w2	w3	w4	w1	w2	w3	w4
Average	5.590323	5.117204	4.315054	3.498925	166.6548	158.0355	145.5419	130.6538
Min	0.1	0.1	0.1	0.1	3	3	3.1	3.4
Max	27	27	26.8	25.6	477.1	458.9	455.3	422.9
Std Dev	6.346196	5.976201	5.138448	4.099108	134.6026	126.4659	115.0918	101.7701
Q1	0.95	0.9	0.9	0.85	50.55	50.55	50.95	48.35
Median	3.3	2.8	2.4	2.1	131.8	122.5	119.9	117.1
Q3	7.7	6.55	5.95	4.9	247.8	229.5	204.7	181.9

	Min Liv. Allownce ppl in urban				Avg wage of employees			
	w1	w2	w3	w4	w1	w2	w3	w4
Average	60.66989	54.38817	47.76667	40.3043	55880.16	61332.41	67377.2	73748.9
Min	4.6	3.6	3.3	2.9	38301	42179	45403	49505
Max	183.6	173.4	156.4	134.5	111390	119935	131700	145766
Std Dev	43.27831	39.56736	35.21259	29.96035	14809.61	16405.28	17848.13	19247.88
Q1	18.35	17.2	16.55	15	47057.5	52078.5	56635.5	61818
Median	55	47.6	40.2	34.8	52332	57270	61973	68533
Q3	87.8	81.75	71.35	57.25	58344	63139	69262.5	76785.5

Table 14. Summary statistics of miscellaneous factors.

	No of unemployed				Social Organisation				No. of NGOs				Basic Pension Holders				ppl with 15+ age			
	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4	W1	W2	W3	W4
Average	24.89355	25.3871	25.54516	25.59355	19451.18	21117.97	22789.27	24455.41	9416.71	10558.66	11722.28	12958.32	1616.449	1628.241	1640.577	1661.218	211444.8	211771.5	211920.9	30859.15
Min	1.6	1.7	1.8	1.8	572	572	572	604	17	17	23	41	78.3	78.3	78.8	78.7	1976	1976	2119	2119
Max	54.6	56.3	56.3	56.3	80385	84094	87024	93061	44705	48534	51225	55090	4855.2	4893.7	5010.2	5082.5	1412318	1412318	1412318	77295
Std Dev	13.94401	14.21177	14.22832	14.22174	14793.53	16352.12	17479.25	18515.86	8258.246	9330.381	10207.21	11050.2	1278.367	1285.709	1293.258	1305.488	329117.9	328950.2	328875	19206.8
Q1	13.55	14.2	14.4	14.5	9643.5	10569.5	10933	12347	3828.5	4713	5076	5661.5	545.65	550.5	581.15	637.3	20803.5	21265.5	21397.5	17881.5
Median	24.5	25.1	25.8	26	15360	17642	19699	21928	6945	7783	8977	9518	1473	1480.4	1489.1	1493.7	40579	41225	41739	27129
Q3	37.75	37.45	37.1	37.1	23983.5	26357	28013	29463	11917.5	13715.5	15120.5	16529.5	2342.9	2337.15	2336.75	2336.75	337330	337330	337330	41986.5

Appendix (C)

Health factors
Window-2:

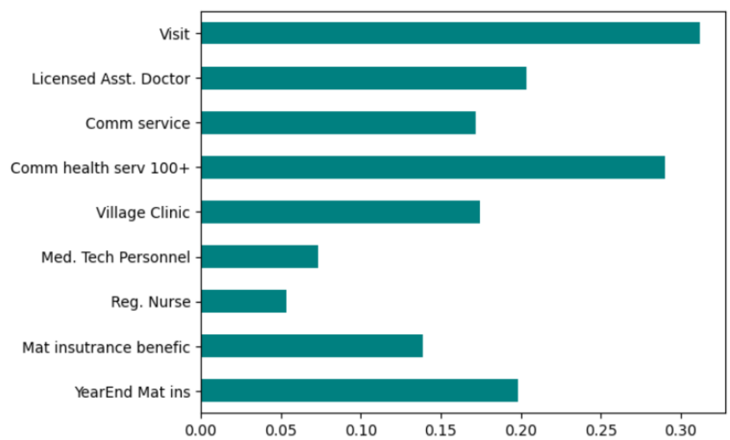


Figure 8(a). Information gain values of relevant health factors (Window-2).

Choose: Visit, 100+bed, Asst Doc, Year End Mat, Village clinic, community service, mat insurance benefic.

	YearEnd Mat ins	Mat insutrance benefic	Reg. Nurse	Med. Tech Personnel	Village Clinic	Comm health serv 100+	Comm service	Licensed Asst. Doctor	Visit	W1-Cluster
YearEnd Mat ins	1.000000	0.781613	0.218997	0.143534	0.237237	0.378102	0.920224	0.173901	0.529121	-0.239473
Mat insutrance benefic	0.781613	1.000000	0.353659	0.269764	0.128950	0.464491	0.686541	0.286086	0.525905	-0.186741
Reg. Nurse	0.218997	0.353659	1.000000	0.915153	-0.254841	0.364994	0.130737	0.781926	0.026029	0.123163
Med. Tech Personnel	0.143534	0.269764	0.915153	1.000000	-0.345278	0.346140	0.042492	0.938990	-0.144318	0.125350
Village Clinic	0.237237	0.128950	-0.254841	-0.345278	1.000000	-0.008705	0.363082	-0.317131	0.661476	-0.168112
Comm health serv 100+	0.378102	0.464491	0.364994	0.346140	-0.008705	1.000000	0.370409	0.447101	0.228970	-0.005238
Comm service	0.920224	0.686541	0.130737	0.042492	0.363082	0.370409	1.000000	0.064651	0.602179	-0.232925
Licensed Asst. Doctor	0.173901	0.286086	0.781926	0.938990	-0.317131	0.447101	0.064651	1.000000	-0.175423	0.076574
Visit	0.529121	0.525905	0.026029	-0.144318	0.661476	0.228970	0.602179	-0.175423	1.000000	-0.245298
W1-Cluster	-0.239473	-0.186741	0.123163	0.125350	-0.168112	-0.005238	-0.232925	0.076574	-0.245298	1.000000

Figure 8(b). Correlation coefficient values for relevant health factors (Window-2).

Choose: Visit, 100+ bed, Asst.Doc only. Rest all are highly correl.

Window-3:

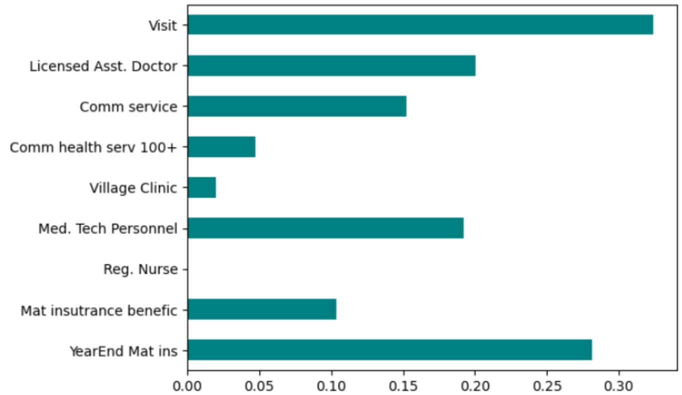


Figure 9(a). Information gain values of relevant health factors (Window-3).

	YearEnd Mat ins	Mat insutrance benefic	Reg. Nurse	Med. Tech Personnel	Village Clinic	Comm health serv 100+	Comm service	Licensed Asst. Doctor	Visit	W1-Cluster
YearEnd Mat ins	1.000000	0.834312	0.229925	0.145183	0.227221	0.393927	0.916806	0.169913	0.522915	-0.290211
Mat insutrance benefic	0.834312	1.000000	0.309066	0.218786	0.143549	0.489008	0.736565	0.235396	0.540320	-0.294691
Reg. Nurse	0.229925	0.309066	1.000000	0.899699	-0.250630	0.329161	0.133021	0.755499	0.053477	-0.020658
Med. Tech Personnel	0.145183	0.218786	0.899699	1.000000	-0.349782	0.282138	0.029336	0.939869	-0.130722	0.037964
Village Clinic	0.227221	0.143549	-0.250630	-0.349782	1.000000	0.041584	0.369386	-0.323113	0.640965	-0.096731
Comm health serv 100+	0.393927	0.489008	0.329161	0.282138	0.041584	1.000000	0.386634	0.366512	0.282584	-0.128061
Comm service	0.916806	0.736565	0.133021	0.029336	0.369386	0.386634	1.000000	0.045574	0.598916	-0.230265
Licensed Asst. Doctor	0.169913	0.235396	0.755499	0.939869	-0.323113	0.366512	0.045574	1.000000	-0.167876	0.032612
Visit	0.522915	0.540320	0.053477	-0.130722	0.640965	0.282584	0.598916	-0.167876	1.000000	-0.189379
W1-Cluster	-0.290211	-0.294691	-0.020658	0.037964	-0.096731	-0.128061	-0.230265	0.032612	-0.189379	1.000000

Figure 9(b). Correlation coefficient values for relevant health factors (Window-3).

Window-4:

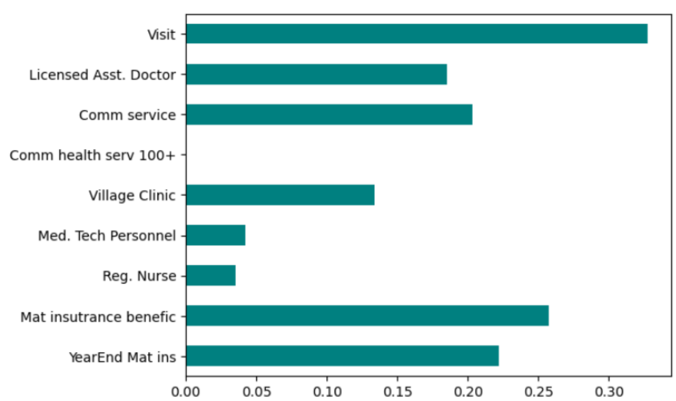


Figure 10(a). Information gain values of relevant health factors (Window-4).

	YearEnd Mat ins	Mat insutrance benefic	Reg. Nurse	Med. Tech Personnel	Village Clinic	Comm health serv 100+	Comm service	Licensed Asst. Doctor	Visit	W1-Cluster
YearEnd Mat ins	1.000000	0.894506	0.252871	0.153402	0.226307	0.415703	0.914685	0.162973	0.530433	-0.244959
Mat insutrance benefic	0.894506	1.000000	0.275879	0.176199	0.160716	0.520692	0.796796	0.185635	0.561821	-0.266091
Reg. Nurse	0.252871	0.275879	1.000000	0.861180	-0.245376	0.300790	0.139904	0.678452	0.107469	0.076158
Med. Tech Personnel	0.153402	0.176199	0.861180	1.000000	-0.354340	0.222203	0.016133	0.933587	-0.111217	0.095337
Village Clinic	0.226307	0.160716	-0.245376	-0.354340	1.000000	0.086376	0.384327	-0.324665	0.618521	-0.097050
Comm health serv 100+	0.415703	0.520692	0.300790	0.222203	0.086376	1.000000	0.407250	0.274859	0.325511	-0.147385
Comm service	0.914685	0.796796	0.139904	0.016133	0.384327	0.407250	1.000000	0.024576	0.602381	-0.177542
Licensed Asst. Doctor	0.162973	0.185635	0.678452	0.933587	-0.324665	0.274859	0.024576	1.000000	-0.161698	0.074861
Visit	0.530433	0.561821	0.107469	-0.111217	0.618521	0.325511	0.602381	-0.161698	1.000000	-0.121049
W1-Cluster	-0.244959	-0.266091	0.076158	0.095337	-0.097050	-0.147385	-0.177542	0.074861	-0.121049	1.000000

Figure 10(b). Correlation coefficient values for relevant health factors (Window-4).

Educational Factors
Window-2:

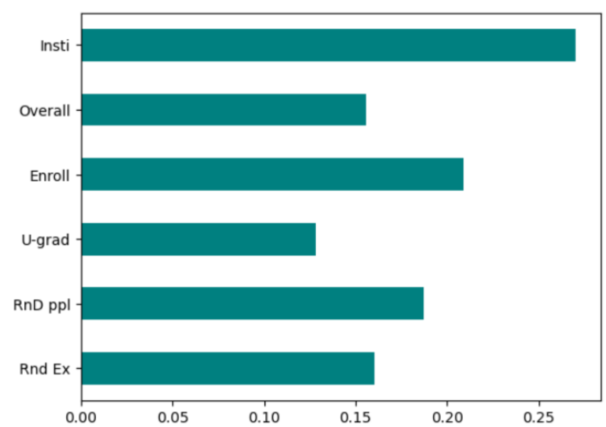


Figure 11(a). Information gain values of relevant educational factors (Window-2).

	Rnd Ex	RnD ppl	U-grad	Enroll	Overall	Insti	W1-Cluster
Rnd Ex	1.000000	0.966833	0.742949	0.729262	0.744715	0.716258	-0.208162
RnD ppl	0.966833	1.000000	0.721700	0.710686	0.721549	0.699472	-0.264118
U-grad	0.742949	0.721700	1.000000	0.979432	0.994921	0.954186	-0.203794
Enroll	0.729262	0.710686	0.979432	1.000000	0.984321	0.964318	-0.184876
Overall	0.744715	0.721549	0.994921	0.984321	1.000000	0.954082	-0.208392
Insti	0.716258	0.699472	0.954186	0.964318	0.954082	1.000000	-0.220362
W1-Cluster	-0.208162	-0.264118	-0.203794	-0.184876	-0.208392	-0.220362	1.000000

Figure 11(b). Correlation coefficient values for relevant educational factors (Window-2).

Window-3:

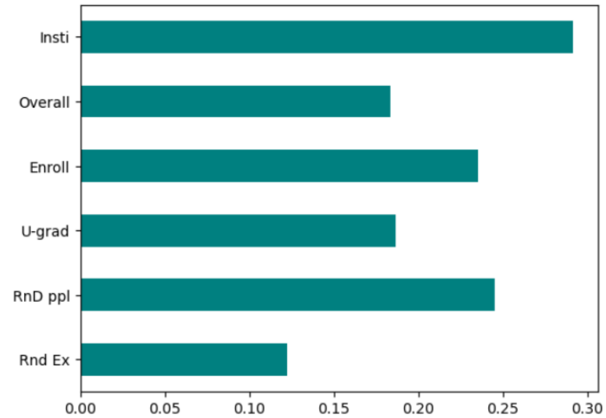


Figure 12(a). Information gain values of relevant educational factors (Window-3).

	Rnd Ex	RnD ppl	U-grad	Enroll	Overall	Insti	W1-Cluster
Rnd Ex	1.000000	0.965902	0.751742	0.736417	0.740934	0.726512	-0.286690
RnD ppl	0.965902	1.000000	0.711987	0.707413	0.699661	0.705258	-0.305305
U-grad	0.751742	0.711987	1.000000	0.983462	0.995646	0.949569	-0.180188
Enroll	0.736417	0.707413	0.983462	1.000000	0.981173	0.965371	-0.186841
Overall	0.740934	0.699661	0.995646	0.981173	1.000000	0.946755	-0.191002
Insti	0.726512	0.705258	0.949569	0.965371	0.946755	1.000000	-0.215943
W1-Cluster	-0.286690	-0.305305	-0.180188	-0.186841	-0.191002	-0.215943	1.000000

Figure 12(b). Correlation coefficient values for relevant educational factors (Window-3).

Window-4:

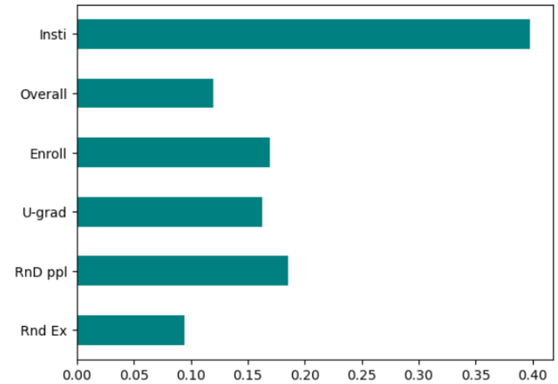


Figure 13(a). Information gain values of relevant educational factors (Window-4).

	Rnd Ex	RnD ppl	U-grad	Enroll	Overall	Insti	W1-Cluster
Rnd Ex	1.000000	0.966164	0.747348	0.739211	0.733340	0.731446	-0.284551
RnD ppl	0.966164	1.000000	0.688691	0.692122	0.677389	0.690029	-0.265683
U-grad	0.747348	0.688691	1.000000	0.980802	0.994873	0.947583	-0.132390
Enroll	0.739211	0.692122	0.980802	1.000000	0.980644	0.966148	-0.116468
Overall	0.733340	0.677389	0.994873	0.980644	1.000000	0.944219	-0.131137
Insti	0.731446	0.690029	0.947583	0.966148	0.944219	1.000000	-0.126767
W1-Cluster	-0.284551	-0.265683	-0.132390	-0.116468	-0.131137	-0.126767	1.000000

Figure 13(b). Correlation coefficient values for relevant educational factors (Window-4).

Public Health Factors
Window-2:

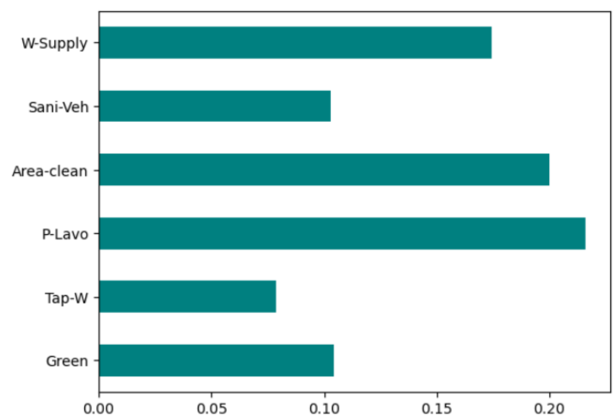


Figure 14(a). Information gain values of relevant public health factors (Window-2).

	Green	Tap-W	P-Lavo	Area-clean	Sani-Veh	W-Supply	W1-Cluster
Green	1.000000	0.132227	0.155125	0.352465	0.225846	0.046499	0.198901
Tap-W	0.132227	1.000000	0.235614	0.248577	0.335556	0.088039	0.018972
P-Lavo	0.155125	0.235614	1.000000	0.784939	0.805749	0.538740	-0.183565
Area-clean	0.352465	0.248577	0.784939	1.000000	0.811412	0.532732	-0.128741
Sani-Veh	0.225846	0.335556	0.805749	0.811412	1.000000	0.495015	0.028440
W-Supply	0.046499	0.088039	0.538740	0.532732	0.495015	1.000000	-0.212861
W1-Cluster	0.198901	0.018972	-0.183565	-0.128741	0.028440	-0.212861	1.000000

Figure 14(b). Correlation coefficient values for relevant public health factors (Window-2).

Window-3:

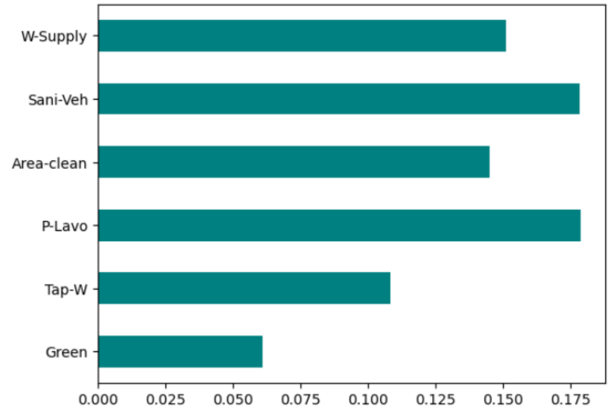


Figure 15(a). Information gain values of relevant public health factors (Window-3).

	Green	Tap-W	P-Lavo	Area-clean	Sani-Veh	W-Supply	W1-Cluster
Green	1.000000	0.252782	0.211812	0.364169	0.263077	0.098508	-0.153337
Tap-W	0.252782	1.000000	0.251045	0.202116	0.286238	0.174951	-0.137895
P-Lavo	0.211812	0.251045	1.000000	0.803708	0.781448	0.545464	-0.264102
Area-clean	0.364169	0.202116	0.803708	1.000000	0.842276	0.569350	-0.216090
Sani-Veh	0.263077	0.286238	0.781448	0.842276	1.000000	0.520915	-0.060893
W-Supply	0.098508	0.174951	0.545464	0.569350	0.520915	1.000000	-0.135266
W1-Cluster	-0.153337	-0.137895	-0.264102	-0.216090	-0.060893	-0.135266	1.000000

Figure 15(b). Correlation coefficient values for relevant public health factors (Window-3).

Window-4:

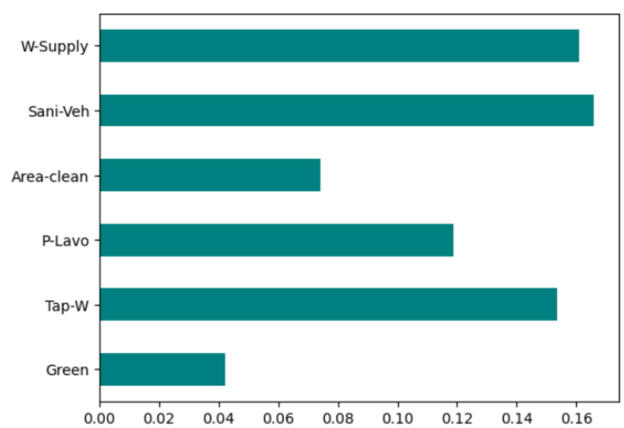


Figure 16(a). Information gain values of relevant public health factors (Window-4).

	Green	Tap-W	P-Lavo	Area-clean	Sani-Veh	W-Supply	W1-Cluster
Green	1.000000	0.291834	0.236292	0.366459	0.266182	0.145436	-0.136802
Tap-W	0.291834	1.000000	0.244916	0.200525	0.276620	0.173421	0.160948
P-Lavo	0.236292	0.244916	1.000000	0.778267	0.783555	0.547274	-0.191467
Area-clean	0.366459	0.200525	0.778267	1.000000	0.858809	0.572568	-0.209347
Sani-Veh	0.266182	0.276620	0.783555	0.858809	1.000000	0.525039	-0.017848
W-Supply	0.145436	0.173421	0.547274	0.572568	0.525039	1.000000	-0.098488
W1-Cluster	-0.136802	0.160948	-0.191467	-0.209347	-0.017848	-0.098488	1.000000

Figure 16(b). Correlation coefficient values for relevant public health factors (Window-4).

Wealth Indicators
Window-2:

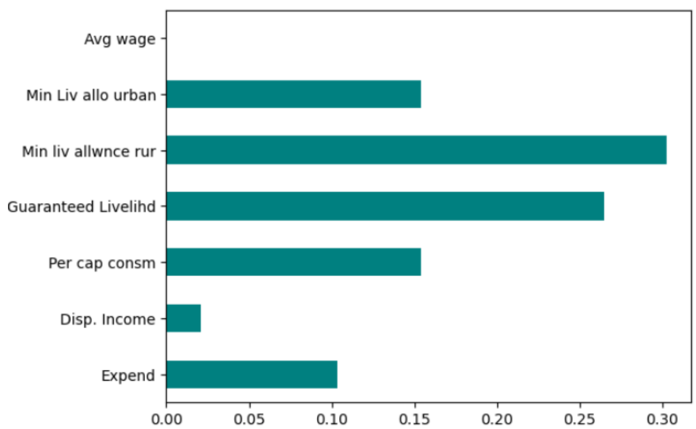


Figure 17(a). Information gain values of relevant wealth indicators (Window-2).

	Expend	Disp. Income	Per cap consm	Guaranteed Livelihd	Min liv allwnce rur	Min Liv allo urban	Avg wage	W1-Cluster
Expend	1.000000	0.376589	0.340266	0.490274	0.286054	0.196244	0.108016	-0.260546
Disp. Income	0.376589	1.000000	0.988904	-0.218327	-0.488212	-0.429397	0.807468	-0.039026
Per cap consm	0.340266	0.988904	1.000000	-0.257881	-0.511798	-0.430664	0.813659	-0.044189
Guaranteed Livelihd	0.490274	-0.218327	-0.257881	1.000000	0.561830	0.630534	-0.318705	-0.271006
Min liv allwnce rur	0.286054	-0.488212	-0.511798	0.561830	1.000000	0.651283	-0.495514	-0.104317
Min Liv allo urban	0.196244	-0.429397	-0.430664	0.630534	0.651283	1.000000	-0.537190	-0.120117
Avg wage	0.108016	0.807468	0.813659	-0.318705	-0.495514	-0.537190	1.000000	0.078036
W1-Cluster	-0.260546	-0.039026	-0.044189	-0.271006	-0.104317	-0.120117	0.078036	1.000000

Figure 17(b). Correlation coefficient values for relevant wealth indicators (Window-2).

Window-3:

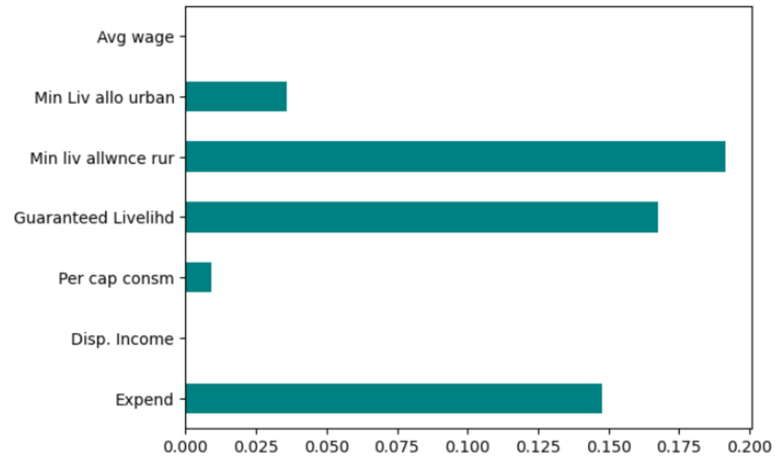


Figure 18(a). Information gain values of relevant wealth indicators (Window-3).

	Expend	Disp. Income	Per cap consm	Guaranteed Livelihd	Min liv allwnce rur	Min Liv allo urban	Avg wage	W1-Cluster
Expend	1.000000	0.402178	0.387777	0.382940	0.254515	0.118911	0.098544	-0.309030
Disp. Income	0.402178	1.000000	0.987859	-0.231601	-0.491342	-0.418705	0.750522	-0.137825
Per cap consm	0.387777	0.987859	1.000000	-0.257656	-0.504452	-0.412956	0.743743	-0.149339
Guaranteed Livelihd	0.382940	-0.231601	-0.257656	1.000000	0.558571	0.613662	-0.330486	-0.051974
Min liv allwnce rur	0.254515	-0.491342	-0.504452	0.558571	1.000000	0.669758	-0.492028	-0.062845
Min Liv allo urban	0.118911	-0.418705	-0.412956	0.613662	0.669758	1.000000	-0.536864	-0.038319
Avg wage	0.098544	0.750522	0.743743	-0.330486	-0.492028	-0.536864	1.000000	-0.022012
W1-Cluster	-0.309030	-0.137825	-0.149339	-0.051974	-0.062845	-0.038319	-0.022012	1.000000

Figure 18(b). Correlation Coefficient values for relevant Wealth Indicators (Window-3).

Window-4:

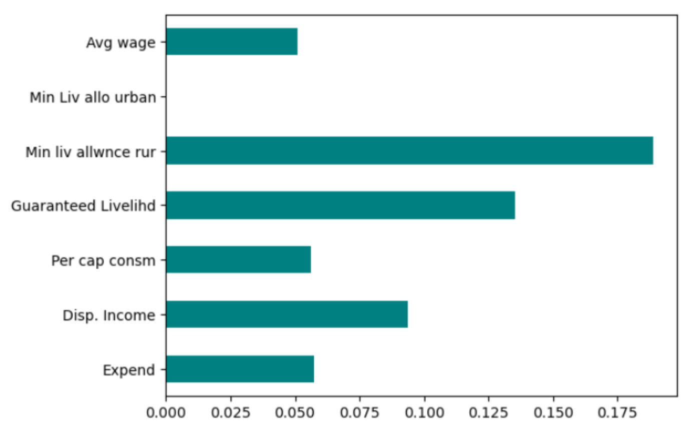


Figure 19(a). Information gain values of relevant wealth indicators (Window-4).

	Expend	Disp. Income	Per cap consm	Guaranteed Livelihd	Min liv allwnce rur	Min Liv allo urban	Avg wage	W1-Cluster
Expend	1.000000	0.404142	0.390937	0.367878	0.250500	0.098111	0.121684	-0.245978
Disp. Income	0.404142	1.000000	0.987724	-0.227383	-0.498558	-0.404898	0.766032	-0.062410
Per cap consm	0.390937	0.987724	1.000000	-0.251893	-0.505123	-0.391187	0.756479	-0.064368
Guaranteed Livelihd	0.367878	-0.227383	-0.251893	1.000000	0.541013	0.600703	-0.332468	-0.159979
Min liv allwnce rur	0.250500	-0.498558	-0.505123	0.541013	1.000000	0.668215	-0.488610	-0.040306
Min Liv allo urban	0.098111	-0.404898	-0.391187	0.600703	0.668215	1.000000	-0.524556	-0.074334
Avg wage	0.121684	0.766032	0.756479	-0.332468	-0.488610	-0.524556	1.000000	-0.125313
W1-Cluster	-0.245978	-0.062410	-0.064368	-0.159979	-0.040306	-0.074334	-0.125313	1.000000

Figure 19(b). Correlation coefficient values for relevant wealth indicators (Window-4).

Miscellaneous Factors
Window-2:

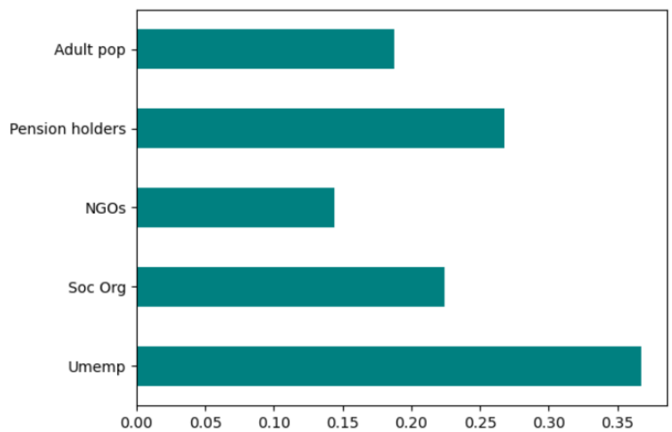


Figure 20(a). Information gain values of relevant miscellaneous factors (Window-2).

	Umemp	Soc Org	NGOs	Pension holders	Adult pop	W1-Cluster
Umemp	1.000000	0.646244	0.669481	0.696568	0.350305	-0.194501
Soc Org	0.646244	1.000000	0.982604	0.606662	0.433116	-0.262250
NGOs	0.669481	0.982604	1.000000	0.585623	0.449770	-0.241886
Pension holders	0.696568	0.606662	0.585623	1.000000	0.354142	-0.227074
Adult pop	0.350305	0.433116	0.449770	0.354142	1.000000	-0.172475
W1-Cluster	-0.194501	-0.262250	-0.241886	-0.227074	-0.172475	1.000000

Figure 20(b). Correlation coefficient values for relevant miscellaneous factors (Window-2).

Window-3:

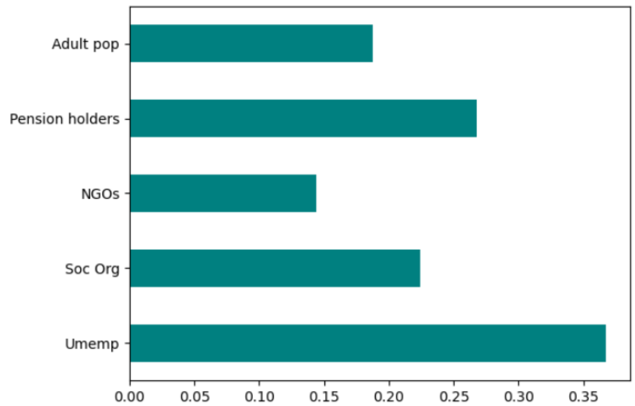


Figure 21(a). Information gain values of relevant miscellaneous factors (Window-3).

	Umemp	Soc Org	NGOs	Pension holders	Adult pop	W1-Cluster
Umemp	1.000000	0.589905	0.635555	0.683278	0.319382	-0.144992
Soc Org	0.589905	1.000000	0.976938	0.583664	0.269131	-0.291667
NGOs	0.635555	0.976938	1.000000	0.589851	0.249901	-0.277095
Pension holders	0.683278	0.583664	0.589851	1.000000	0.341397	-0.137226
Adult pop	0.319382	0.269131	0.249901	0.341397	1.000000	0.085040
W1-Cluster	-0.144992	-0.291667	-0.277095	-0.137226	0.085040	1.000000

Figure 21(b). Correlation coefficient values for relevant miscellaneous factors (Window-3).

Window-4:

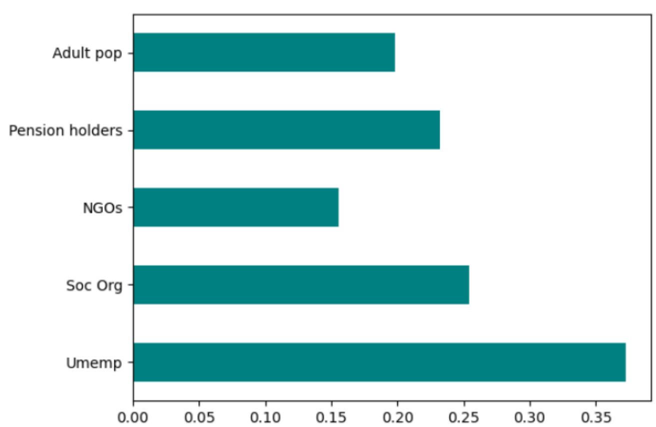


Figure 22(a). Information gain values of relevant miscellaneous factors (Window-4).

	Umemp	Soc Org	NGOs	Pension holders	Adult pop	W1-Cluster
Umemp	1.000000	0.589761	0.647085	0.691727	0.815999	-0.106883
Soc Org	0.589761	1.000000	0.974188	0.588403	0.828172	-0.278471
NGOs	0.647085	0.974188	1.000000	0.605404	0.843874	-0.238932
Pension holders	0.691727	0.588403	0.605404	1.000000	0.866483	-0.089319
Adult pop	0.815999	0.828172	0.843874	0.866483	1.000000	-0.167398
W1-Cluster	-0.106883	-0.278471	-0.238932	-0.089319	-0.167398	1.000000

Figure 22(b). Correlation coefficient values for relevant miscellaneous factors (Window-4).

References

2014 National Healthcare Quality and Disparities Report: Chartbook on Women's Health Care [Internet]. Rockville (MD): Agency for healthcare research and quality (US); 2015. AHRQ Pub. No. 15-0007-10-EF. Sep. Available from: https://www.ncbi.nlm.nih.gov/books/NBK581138/".

Ali, A.I., & Seiford, L.M. (1990). Translation invariance in data envelopment analysis. *Operations Research Letters*, *9*(6), 403-405. https://doi.org/10.1016/0167-6377(90)90061-9.

Asmild, M., Paradi, J.C., Aggarwall, V., & Schaffnit, C. (2004). Combining DEA window analysis with the Malmquist index approach in a study of the Canadian banking industry. *Journal of Productivity Analysis*, *21*, 67-89. https://doi.org/10.1023/b:prod.0000012453.91326.ec.

Attaran, M., & Gunasekaran, A. (2019). Healthcare and life sciences. *Applications of Blockchain Technology in Business: Challenges and Opportunities* (pp. 53-57). Springer. https://doi.org/10.1007/978-3-030-27798-7.

Badri, M.A., & Hollingsworth, J. (1993). A simulation model for scheduling in the emergency room. *International Journal of Operations & Production Management*, *13*(3), 13-24. https://doi.org/10.1108/01443579310025989.

Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1031-1142. https://doi.org/10.1287/mnsc.30.9.1078.

Bekkerman, R., El-Yaniv, R., Tishby, N., & Winter, Y. (2003). Distributional word clusters vs. words for text categorization. *Journal of Machine Learning Research*, *3*, 1183-1208.

Belhadi, A., Kamble, S., Jabbour, C.J.C., Gunasekaran, A., Ndubisi, N.O., & Venkatesh, M. (2021). Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries. *Technological Forecasting and Social Change*, *163*, 120447. https://doi.org/10.1016/j.techfore.2020.120447.

Bretthauer, K.M., Heese, H.S., Pun, H., & Coe, E. (2011). Blocking in healthcare operations: a new heuristic and an application. *Production and Operations Management*, *20*(3), 375-391. https://doi.org/10.1111/j.1937-5956.2011.01230.x.

Brohman, M.K., Piccoli, G., Martin, P., Zulkernine, F., Parasuraman, A., & Watson, R.T. (2009). A design theory approach to building strategic network-based customer service systems. *Decision Sciences*, *40*(3), 403-430.

Cai, J., Coyte, P.C., & Zhao, H. (2017). Decomposing the causes of socioeconomic-related health inequality among urban and rural populations in China: a new decomposition approach. *International Journal for Equity in Health*, *16*(128), 1-14. https://doi.org/10.1186/s12939-017-0624-9.

Campbell, D., & Frei, F. (2010). Cost structure, customer profitability, and retention implications of self-service distribution channels: Evidence from customer behavior in an online banking channel. *Management Science, 56*(1), 4-24. https://doi.org/10.1287/mnsc.1090.1066.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

Charles, V., Aparicio, J., & Zhu, J. (2019). The curse of dimensionality of decision-making units: a simple approach to increase the discriminatory power of data envelopment analysis. *European Journal of Operational Research*, *279*(3), 929-940. https://doi.org/10.1016/j.ejor.2019.06.025.

Charnes, A., & Cooper, W.W. (1984). Preface to topics in data envelopment analysis. *Annals of Operations Research*, *2*(1), 59-94. https://doi.org/10.1007/bf01874733.

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444. https://doi.org/10.1016/0377-2217(78)90138-8.

Chen, K., Ren, X.T., Yang, G.I., & Qin, H.B. (2022). The other side of the coin: the declining of Chinese social science. *Scientometrics, 127*(1), 127-143. https://doi.org/10.1007/s11192-021-04208-2.

Chen, W., Zhang, Z.G., & Chen, X. (2020). On two-tier healthcare system under capacity constraint. *International Journal of Production Research, 58*(12), 3744-3764. https://doi.org/10.1080/00207543.2019.1634846.

Co, H.C., & Chew, K.S. (1997). Performance and R&D expenditures in American and Japanese manufacturing firms. *International Journal of Production Research*, *35*(12), 3333-3348. https://doi.org/10.1080/002075497194101.

Cooper, W.W., Seiford, L.M., & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*. Springer, New York. ISBN: 978-0-387-45283-8. https://doi.org/10.1007/978-0-387-45283-8.

DeLellis, N.O., & Ozcan, Y.A. (2013). Quality outcomes among efficient and inefficient nursing homes: a national study. *Health Care Management Review, 38*(2), 156-165. https://doi.org/10.1097/hmr.0b013e31824bec38.

Dhillon, I.S., Mallela, S., & Kumar, R. (2003). A divisive information theoretic feature clustering algorithm for text classification. *The Journal of Machine Learning Research*, *3*, 1265-1287.

Du, T. (2017). Performance measurement of healthcare service and association discussion between quality and efficiency: evidence from 31 provinces of mainland China. *Sustainability, 10*(1), 74. https://doi.org/10.3390/su10010074.

Dwyer-Lindgren, L., Stubbs, R.W., Bertozzi-Villa, A., Morozoff, C., Callender, C., Finegold, S.B., Shirude, S., Flaxman, A.D., Lawrent, A., Kern, E., Duchin, J.S., Fleming, D., Mokdad, A.H., & Murray, C.J.L. (2017). Variation in life expectancy and mortality by cause among neighbourhoods in King County, WA, USA, 1990–2014: a census tract-level analysis for the Global Burden of Disease Study 2015. *The Lancet Public Health, 2*(9), e400-e410. https://doi.org/10.1016/s2468-2667(17)30165-2.

Eggleston, K., Ling, L., Qingyue, M., Lindelow, M., & Wagstaff, A. (2008). Health service delivery in China: a literature review. *Health Economics, 17*(2), 149-165. https://doi.org/10.1002/hec.1306.

Ersoy, K., Kavuncubasi, S., Ozcan, Y.A., & Harris II, J.M. (1997). Technical efficiencies of Turkish hospitals: DEA approach. *Journal of Medical Systems, 21*, 67-74.https://doi.org/10.1023/a:1022801222540.

Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253-290. https://doi.org/10.2307/2343100.

Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, *120*(3), 253-281.

Feng, Y., Zhang, H., Chiu, Y.H, & Chang, T.H. (2021). Innovation efficiency and the impact of the institutional quality: a cross-country analysis using the two-stage meta-frontier dynamic network DEA model. *Scientometrics, 126*(4), 3091-3129. https://doi.org/10.1007/s11192-020-03829-3.

Ferreira, D.C., Nunes, A.M., & Marques, R.C. (2018). Doctors, nurses, and the optimal scale size in the Portuguese public hospitals. *Health Policy*, *122*(10), 1093-1100. https://doi.org/10.1016/j.healthpol.2018.06.009.

Fetter, R.B., & Thompson, J.D. (1965). The simulation of hospital systems. *Operations Research*, *13*(5), 689-711. https://doi.org/10.1287/opre.13.5.689.

Finkler, M.D., & Wirtschafter, D.D. (1993). Cost-effectiveness and data envelopment analysis. *Health Care Management Review*, *18*(3), 81-88. https://doi.org/10.1097/00004010-199301830-00010.

Fone, D., Hollinghurst, S., Temple, M., Round, A., Lester, N., Weightman, A., Robert, K., Coyle, E., Bevan, G., & Palmer, S. (2003). Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health*, *25*(4), 325-335. https://doi.org/10.1093/pubmed/fdg075.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, *3*, 1289-1305. https://doi.org/10.1162/153244303322753670.

Geng, N., Xie, X., & Zhang, Z. (2019). Addressing healthcare operational deficiencies using stochastic and dynamic programming. *International Journal of Production Research*, *57*(14), 4371-4390. https://doi.org/10.1080/00207543.2017.1397789.

Gholami-Zanjani, S.M., Pishvaee, M.S., & Torabi, S.A. (2018). OR models for emergency medical service (EMS) management. In: Kahraman, C., Topcu, Y. (eds) *Operations Research Applications in Health Care Management*, Springer, Cham, pp. 395-421. https://doi.org/10.1007/978-3-319-65455-3_16.

Gray, R.M. (2011). *Entropy and information theory*. Springer Science & Business Media, New York. ISBN: 978-1-4419-7969-8(p), 978-1-4419-7970-4(e). https://doi.org/10.1007/978-1-4419-7970-4.

Han, J., & Kamber, M. (2001). *Data mining concepts and techniques*. Morgan Kaufmann Publishers, San Francisco, CA.

Henrique, D.B., Filho, M.G., Marodin, G., Jabbour, A.B.L.D.S., & Chiappetta Jabbour, C.J. (2021). A framework to assess sustaining continuous improvement in lean healthcare. *International Journal of Production Research*, *59*(10), 2885-2904. https://doi.org/10.1080/00207543.2020.1743892.

Hill, K.D., Vu, M., & Walsh, W. (2007). Falls in the acute hospital setting-impact on resource utilisation. *Australian Health Review*, *31*(3), 471-477. https://doi.org/10.1071/ah070471.

Ho, M.H.C., & Liu, J.S. (2013). The motivations for knowledge transfer across borders: the diffusion of data envelopment analysis (DEA) methodology. *Scientometrics*, *94*(1), 397-421. https://doi.org/10.1007/s11192-012-0705-x.

Hollingsworth, B. (2003). Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science*, *6*(4), 203-218. https://doi.org/10.1023/a:1026255523228.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, *17*(10), 1107-1128. https://doi.org/10.1002/hec.1391.

Hollingsworth, B., Dawson, P.J., & Maniadakis, N. (1999). Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Management Science*, *2*(3), 161-172. https://doi.org/10.1023/a:1019087828488.

Hood, C.M., Gennuso, K.P., Swain, G.R., & Catlin, B.B. (2016). County health rankings: relationships between determinant factors and health outcomes. *American Journal of Preventive Medicine*, *50*(2), 129-135. https://doi.org/10.1016/j.amepre.2015.08.024.

Hotchkiss, D.R., & Jacobalis, S. (1999). Indonesian heath care and the economic crisis: is managed care the needed reform? *Health Policy*, *46*(3), 195-216. https://doi.org/10.1016/s0168-8510(98)00063-3.

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. *Health Care Management Science*, *4*(2), 103-115. https://doi.org/10.1023/a:1011453526849.

Jha, A.K., Li, Z., Orav, E.J., & Epstein, A.M. (2005). Care in US hospitals—the hospital quality alliance program. *New England Journal of Medicine, 353*(3), 265-274. https://doi.org/10.1056/nejmsa051249.

Jiang, S., Min, R., & Fang, P.Q. (2017). The impact of healthcare reform on the efficiency of public county hospitals in China. *BMC Health Services Research, 17*, 838. https://doi.org/10.1186/s12913-017-2780-4.

Kamble, S.S., Gunasekaran, A., Goswami, M., & Manda, J. (2019). A systematic perspective on the applications of big data analytics in healthcare management. *International Journal of Healthcare Management*, *12*(3), 226-240. https://doi.org/10.1080/20479700.2018.1531606.

Kaplan, R.S., & Porter, M.E. (2011). How to solve the cost crisis in health care. *Harvard Business Review*, *89*(9), 46-52. https://pubmed.ncbi.nlm.nih.gov/21939127/.

Kohl, S., Schoenfelder, J., Fügener, A., & Brunner, J.O. (2019). The use of data envelopment analysis (DEA) in healthcare with a focus on hospitals. *Health Care Management Science, 22*, 245-286. https://doi.org/10.1007/s10729-018-9436-8.

Kusiak, A. (2006). Data mining: manufacturing and service applications. *International Journal of Production Research, 44*(18-19), 4175-4191. https://doi.org/10.1080/00207540600632216.

LaGanga, L.R., & Lawrence, S.R. (2012). Appointment overbooking in health care clinics to improve patient service and clinic performance. *Production and Operations Management, 21*(5), 874-888. https://doi.org/10.1111/j.1937-5956.2011.01308.x.

Li, B., Mohiuddin, M., & Liu, Q. (2019). Determinants and differences of township hospital efficiency among Chinese provinces. *International Journal of Environmental Research and Public Health, 16*(9), 1601. https://doi.org/10.3390/ijerph16091601.

Lu, M., Jin, M., & Quan, H. (2008). From a national, centrally planned health system to a system based on the market: lessons from China, 2008. *Health Affairs, 27*(4), 937-948. https://hdl.handle.net/1783.1/31859.

Martínez-Caro, E., Cegarra-Navarro, J.G., García-Pérez, A., & Fait, M. (2018). Healthcare service evolution towards the internet of things: an end-user perspective. *Technological Forecasting and Social Change, 136*, 268-276. https://doi.org/10.1016/j.techfore.2018.03.025.

Navarro-Espigares, J.L., & Torres, E.H. (2011). Efficiency and quality in health services: a crucial link. *The Service Industries Journal*, *31*(3), 385-403. https://doi.org/10.1080/02642060802712798.

Nayar, P., Ozcan, Y.A., Yu, F., & Nguyen, A.T. (2013). Benchmarking urban acute care hospitals: efficiency and quality perspectives. *Health Care Management Review, 38*(2), 137-145. https://doi.org/10.1097/hmr.0b013e3182527a4c.

Newhouse, J.P. (1970). Toward a theory of nonprofit institutions: an economic model of a hospital. *The American Economic Review, 60*(1), 64-74.

Ozcan, Y.A. (2014). *Health care benchmarking and performance evaluation an assessment using data envelopment analysis* (*DEA*). Springer, New York. ISBN: 978-1-4899-7471-6(p), 978-1-4899-7472-3(e) https://doi.org/10.1007/978-1-4899-7472-3.

Pan, J., Ding, S., Wu, D., Yang, S., & Yang, J. (2019). Exploring behavioural intentions toward smart healthcare services among medical practitioners: a technology transfer perspective. *International Journal of Production Research, 57*(18), 5801-5820. https://doi.org/10.1080/00207543.2018.1550272.

Pandey, U., & Singh, S. (2021). Environmental performance evaluation of European farms by assessing polluting factors in joint production. *Journal of Cleaner Production*, *328*, 129457. https://doi.org/10.1016/j.jclepro.2021.129457.

Pandey, U., & Singh, S. (2022). Data envelopment analysis in hierarchical category structure with fuzzy boundaries. *Annals of Operations Research*, *315*(2), 1517-1549. https://doi.org/10.1007/s10479-020-03854-8.

Pazzani, M., & Billsus, D. (1997). Learning and revising user profiles: the identification of interesting web sites. *Machine Learning*, *27*, 313-331. https://doi.org/10.1023/a:1007369909943.

Porter, M.E. (2011). Competitive advantage of nations: creating and sustaining superior performance. Free Press. ISBN: 9781451651492.

Priya, M., & Kumar, R.P. (2015). A novel intelligent approach for predicting atherosclerotic individuals from big data for healthcare. *International Journal of Production Research*, *53*(24), 7517-7532. https://doi.org/10.1080/00207543.2015.1087655.

Puertas, R., Marti, L., & Coque, J.M.G.A. (2023). Are female researchers more efficient? An analysis of gender in a Spanish technological university. *Scientometrics*, *128*(12), 6611-6632. https://doi.org/10.1007/s11192-023-04844-w.

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81-106. https://doi.org/10.1007/bf00116251.

Quinlan, J.R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, *27*(3), 221-234. https://doi.org/10.1016/s0020-7373(87)80053-6.

Ramadevi, D., Gunasekaran, A., Roy, M., Rai, B.K., & Senthilkumar, S.A. (2016). Human resource management in a healthcare environment: framework and case study. *Industrial and Commercial Training*, *48*(8), 387-393. https://doi.org/10.1108/ict-03-2016-0014.

Salzarulo, P.A., Bretthauer, K.M., Côté, M.J., & Schultz, K.L. (2011). The impact of variability and patient information on health care system performance. *Production and Operations Management*, *20*(6), 848-859. https://doi.org/10.1111/j.1937-5956.2010.01210.x.

Sampson, S.E., & Froehle, C.M. (2006). Foundations and implications of a proposed unified services theory. *Production and Operations Management*, *15*(2), 329-343. https://doi.org/10.1111/j.1937-5956.2006.tb00248.x.

Sanmartin, C., Houle, C., Tremblay, S., & Berthelot, J.M. (2002). Changes in unmet health care needs. *Health Reports*, *13*(3), 15-21.

Sarmiento, R., & Shukla, V. (2011). Zero-sum and frontier trade-offs: an investigation on compromises and compatibilities amongst manufacturing capabilities. *International Journal of Production Research*, *49*(7), 2001-2017. https://doi.org/10.1080/00207540903555544.

Schmenner, R.W., & Swink, M.L. (1998). On theory in operations management. *Journal of Operations Management*, *17*(1), 97-113. https://doi.org/10.1016/s0272-6963(98)00028-x.

Senot, C., Chandrasekaran, A., & Ward, P.T. (2016). Role of bottom-up decision processes in improving the quality of health care delivery: a contingency perspective. *Production and Operations Management*, *25*(3), 458-476. https://doi.org/10.1111/poms.12404.

Sherman, H.D., & Gold, F. (1985). Bank branch operating efficiency: evaluation with data envelopment analysis. *Journal of Banking & Finance*, *9*(2), 297-315. https://doi.org/10.1016/0378-4266(85)90025-1.

Shirazi, F., Wu, Y., Hajli, A., Zadeh, A.H., Hajli, N., & Lin, X. (2021). Value co-creation in online healthcare communities. *Technological Forecasting and Social Change*, *167*, 120665. https://doi.org/10.1016/j.techfore.2021.120665.

Singh, S., & Ranjan, P. (2018). Efficiency analysis of non-homogeneous parallel sub-unit systems for the performance measurement of higher education. *Annals of Operations Research*, *269*(1-2), 641-666. https://doi.org/10.1007/s10479-017-2586-0.

Smith, E.A., & Warner, H.R. (1971). Simulation of a multiphasic screening procedure for hospital admissions. *Simulation*, *17*(2), 57-64. https://doi.org/10.1177/003754977101700202.

Smith-Daniels, V.L., Schweikhart, S.B., & Smith-Daniels, D.E. (1988). Capacity management in health care services: Review and future research directions. *Decision Sciences*, *19*(4), 889-919.

Song, X., Wei, Y., Deng, W., Zhang, S., Zhou, P., Liu, Y., & Wan, J. (2019). Spatio-temporal distribution, spillover effects and influences of China's two levels of public healthcare resources. *International Journal of Environmental Research and Public Health, 16*(4), 582. https://doi.org/10.3390/ijerph16040582.

Talluri, S., Kim, M.K., & Schoenherr, T. (2013). The relationship between operating efficiency and service quality: are they compatible? *International Journal of Production Research, 51*(8), 2548-2567. https://doi.org/10.1080/00207543.2012.737946.

Tone, K. (2011). Slacks-Based measure of efficiency. In: Cooper, W., Seiford, L., Zhu, J. (eds) *Handbook on Data Envelopment Analysis*. Springer, Boston, MA, pp. 195-209. https://doi.org/10.1007/978-1-4419-6151-8_8.

Torkkola, K. (2003). Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research, 3*, 1415-1438.

Tortorella, G.L., Saurin, T.A., Fogliatto, F.S., Rosa, V.M., Tonetto, L.M., & Magrabi, F. (2021). Impacts of healthcare 4.0 digital technologies on the resilience of hospitals. *Technological Forecasting and Social Change, 166*, 120666. https://doi.org/10.1016/j.techfore.2021.120666.

Tsolas, I.E., Charles, V., & Gherman, T. (2020). Supporting better practice benchmarking: a DEA-ANN approach to bank branch performance assessment. *Expert Systems with Applications*, *160*, 113599. https://doi.org/10.1016/j.eswa.2020.113599.

Wagstaff, A., Yip, W., Lindelow, M., & Hsiao, W.C. (2009). China's health system and its reform: a review of recent studies. *Health Economics, 18*(S2), S7-S23. https://doi.org/10.1002/hec.1518.

Wang, Y., Kung, L., & Byrd, T.A. (2018). Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3-13. https://doi.org/10.1016/j.techfore.2015.12.019.

Wu, S., Wang, C., & Zhang, G. (2015). Has China's new health care reform improved efficiency at the provincial level? Evidence from a panel data of 31 Chinese provinces. *Journal of Asian Public Policy, 8*(1), 36-55.

Yang, J., & Zeng, W. (2014). The trade-offs between efficiency and quality in the hospital production: some evidence from Shenzhen, China. *China Economic Review, 31*, 166-184. https://doi.org/10.1016/j.chieco.2014.09.005.

Zemtsov, S., & Kotsemir, M. (2019). An assessment of regional innovation system efficiency in Russia: the application of the DEA approach. *Scientometrics*, *120*(2), 375-404. https://doi.org/10.1007/s11192-019-03130-y.

Zhang, X., Zhao, L., Cui, Z., & Wang, Y. (2015). Study on equity and efficiency of health resources and services based on key indicators in China. *PLoS One*, *10*(12), e0144809. https://doi.org/10.1371/journal. pone.0144809.

Zhang, Y., Wang, Q., Jiang, T., & Wang, J. (2018). Equity and efficiency of primary health care resource allocation in mainland China. *International Journal for Equity in Health, 17*(1), 140. https://doi.org/10.1186/s12939-018-0851-8.

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.