Environmental Science

AIMS Environmental Science, 11(5): 703–722.
DOI: 10.3934/environsci.2024035
Received: 23 June 2024
Revised: 05 August 2024
Accepted: 22 August 2024
Published: 27 August 2024

http://www.aimspress.com/journal/environmental

Research article

Measuring the impact of technological innovation, green energy, and sustainable development on the health system performance and human well-being: Evidence from a machine learning-based approach

Alaeddine Mihoub[1], Montassar Kahia[2,*] and Mohannad Alswailim[1]

[1] Department of Management Information Systems, College of Business and Economics, Qassim University, Saudi Arabia
[2] Department of Economics, College of Business and Economics, Qassim University, Saudi Arabia

* **Correspondence:** Email: m.kahia@qu.edu.sa.

Abstract: Health performance and well-being are crucial elements of Saudi Arabia's Vision 2030, aiming to improve the overall quality of life and promote a prosperous community. Within this context, this study intended to examine the impact of recent innovations, logistical measures, Information and Communication Technology (ICT) diffusion, environmental quality improvements, economic growth, and green (renewable) energy exploitation on health performance and well-being, in Saudi Arabia from 1990 to 2022, by implementing machine learning models (random forest and gradient boosting) and regression algorithms (ridge and lasso). Overall, the findings of machine learning models indicate a strong impact of digital connectivity on health spending by internet users, with scores of 0.673 and 0.86. Further, economic growth also influences health costs but to a lesser extent, with scores of 0.145 and 0.082. Mobile user penetration and CO_2 emissions have moderate to low importance, suggesting nuanced interactions with health expenditure. Patent applications and logistics performance show minimal impact, indicating a limited direct influence on health costs within this study. Similarly, the share of renewable energy is negligible, reflecting its minimal impact on the analyzed data. Finally, regression analyses using ridge and lasso models confirmed similar trends, further validating these findings. Limitations and several policy implications are also debated.

Keywords: health expenditure; innovation technology; renewable energy; sustainable growth; ICT access; economic well-being

1. Introduction

Promoting well-being and providing adequate healthcare for individuals of all ages are essential to sustainable development. Without ensuring health and well-being, eradicating poverty and fostering economic growth are unattainable goals, as these factors contribute significantly to vibrant and thriving communities [1,2]. Health, inherently a human right, enables individuals to realize their full potential, aligning with global health standards and practices [3–5]. In alignment with Sustainable Development Goal 3, Saudi Arabia has advanced its healthcare systems through a series of policy implementations and infrastructure developments under the Vision 2030 initiative, aiming to ensure healthy lives and promote well-being for all at all ages [6–8].

Saudi Arabia has significantly invested in its healthcare infrastructure by constructing new facilities and expanding existing hospitals to enhance service delivery. These efforts are complemented by advancements in healthcare quality through provider skill development and the integration of new technologies and treatments, highlighting a major innovation in healthcare technology [9–11]. Embracing the digital health trend, the Kingdom is improving the quality and efficiency of healthcare services through advanced e-health technologies like telemedicine and electronic health records, aimed at optimizing resource use and reducing treatment costs [12–14]. Additionally, Saudi Arabia has launched several initiatives to combat non-communicable diseases by establishing guidelines for healthy living, health education programs, and facilities dedicated to noncommunicable disease (NCD) care, significantly enhancing public health measures [15–17]. The government is also focused on improving maternal and child healthcare, implementing measures to uplift pregnancy care and setting up specialized facilities and education programs targeted at mothers and children [12,13]. Furthermore, the country is proactively addressing other public health challenges like infectious diseases through initiatives aimed at popularizing vaccination programs and building specialized medical institutions to better manage these diseases [18,19].

Despite considerable advancements in various health domains, there is still a research gap regarding the full understanding of the collective impact of recent innovations, logistical measures, ICT diffusion, environmental quality improvements, economic growth, and green energy exploitation on health performance in Saudi Arabia. While these elements are acknowledged as critical determinants of healthcare outcomes, empirical evidence detailing their integrated effects on service quality and accessibility is limited [6]. This study aims to bridge this gap by exploring the specific contributions and interactions of these factors within the Saudi healthcare system, thus offering a more detailed understanding of the mechanisms that drive improvements under the Saudi Vision initiative. Additionally, although traditional econometric tools have been widely utilized to analyze health outcomes, they often do not fully capture the complex, nonlinear interactions within healthcare data due to their reliance on predefined assumptions. Machine learning, capable of handling large datasets and identifying hidden patterns without such constraints, presents a promising alternative. However, empirical applications of machine learning in assessing the impact of these variables on health performance are still rare in Saudi Arabia. By employing advanced machine learning algorithms, this paper seeks to provide deeper and more reliable insights, transitioning from simple empirical analysis to more thoroughly researched realities.

The rest of this paper is divided into four parts. Section 2 provides an overview of the literature review, Section 3 specifies the implemented methodology of this investigation, and Section 4 details and discusses the empirical outcomes. Finally, Section 5 provides the conclusion with several practical policy recommendations.

2. Literature review

The connection between innovation, ICT diffusion, green energy technologies, logistic measures, economic growth, environmental quality, and health performance has not been extensively examined in literature reviews pertaining to Saudi Arabia's economy. Also, findings have been equivocal, prompting researchers to recommend further investigations.

Initially, the influence of environmental degradation as well as green energy on household and health revenue in Middle East and North Africa (MENA) economies was analyzed by Alharthi et al. [20]. The findings demonstrate that clean energy greatly benefits human health and drastically decreases pollution levels. Further, health spending, carbon emissions, GDP growth, and natural resource availability are all factors that Ampon-Wireko et al. [21] analyzed for developing countries between 2000 and 2018. PMG and DOLS estimations were used in the investigation and revealed that rising gross domestic product (GDP), pollution, and the exploitation of natural resources all contribute to higher healthcare costs at the panel level. Nevertheless, better sanitation lowers healthcare costs. Chen et al. [22] confirmed that various types of green and clean energy contribute to health improvement for a sample of an Asian economy. The study by da Silva et al. [23] confirmed that technological development of the public health sector has a crucial role in enhancing public health outcomes across various contexts. Further, both developing and developed countries have difficulties with healthcare logistics during the recent pandemic, as confirmed by Fathollahi-Fard et al. [24]. An effectively crafted architecture for strong and sustainable healthcare logistics would boost health performance in response to such a pandemic. In addition, the study conducted by Jiang et al. [25] demonstrated that the process of digitization had a positive impact on long-term lifespan in the nations of the BRICS group, excluding Brazil. The implementation of green technology in Russia and China has been observed to exhibit a favorable link with increased lifespan over a prolonged period. However, its immediate effects on health outcomes are found to be negligible. Besides, the study conducted by Karaaslan and Çamkaya [26] demonstrated a favorable connection between GDP as well as NREC with a rise in CO_2 levels. Conversely, the research findings indicated that health and clean energy negatively correlated with CO_2, suggesting their potential to moderate CO_2 emission. In their study, Li et al. [27] showed that emissions of CO_2, energy consumption, and GDP harm mortality rates and life expectancy. This research focused on the five major carbon-emitting countries using time-series data. It was further shown that performance logistics indices for 27 EU economies all had a strong reliance on CO_2 emission and oil product usage [28]. Paul et al. [29] pointed out further that within healthcare organizations, demand is rising for wide-ranging digital technology systems that help to improve operations.

Rahman and Alam [30] looked into the relationship between good governance, growth, clean energy, and cities with a high population at birth for selected economies. The research utilized yearly data from 1996 through 2019 and examined the relationship between these variables using various econometric methods. Results of their research show that these variables are markedly positively correlated to citizens' life expectancy at birth. In addition, Saleem et al. [31] found a positive link between energy output and healthcare expenditures, based on a balanced panel of data from 38 OECD member states during 2008–2018. The studies conducted by Shin and Kang (2020) and Small et al. [32,33] demonstrated the significant impact of technological innovation and digital technology on improving health outcomes and cleanliness. The studies conducted by Vicente, J. Zhang et al. and Q. Zhang et al. [34–36] established that the dissemination of information and communication technology (ICT) diffusion displays positive influences on health outcomes. This is achieved through two main mechanisms—first, by raising macro-level public healthcare spending, and second, by fostering micro-level literacy in healthcare. In addition, the proliferation of ICT diffusion substantially influences decreasing population death rates. In their recent study, Priyan et al. [37] disclosed that green technology promotion may significantly assist the management of healthcare sector concerns. Further, Omri et al. [6] investigated the association between environmental sustainability, measured by the environmental performance index, and human health outcomes, specifically disability-adjusted life-years and life expectancy, in Saudi Arabia from 2000 to 2020. The researchers also considered the influence of policy variables, namely ICT and technological innovation (TI). The results suggested that the presence of advanced and sophisticated technology enhances the impact of environmental performance on human health outcomes. Likewise, Ijiga et al. [38] investigated the impact of technology advancements in addressing health issues in New York City. The analysis indicated that the incorporation of smart city technology, such as sophisticated data analytics, Internet of Things (IoT) devices, and artificial intelligence (AI)-based solutions, has greatly enhanced the provision of healthcare services and emergency response.

Based on our comprehensive analysis, we found no empirical research in the existing body of literature that examines the impact of recent innovations, logistical measures, ICT diffusion, environmental quality improvements, economic growth, and green energy exploitation on health performance and well-being, specifically in Saudi Arabia. The present study makes a significant contribution to the existing knowledge by addressing the gap in understanding the dynamics of innovations, logistical measures, ICT diffusion, environmental quality improvements, economic growth, and green energy exploitation on health performance and well-being and their influence by implying several machine learning and regression models. In fact, this approach can facilitate a more comprehensive and precise data analysis, enabling the identification of patterns and trends that may not be discernible by conventional approaches to develop practical and effective policies.

3. Methodology

3.1. Dataset description

This study utilizes a dataset composed of various economic, environmental, technological, and logistical indicators to assess their impact on health expenditure in Saudi Arabia. The dataset was compiled from multiple authoritative sources, especially the World Bank website, ensuring a robust basis for the analysis. The data encompasses a period from 1990 to 2022, allowing for a comprehensive study of trends over more than three decades.

The compiled dataset captures eight key variables that reflect the interplay between economic activity, environmental practices, technological advancement, and logistical efficiency within Saudi Arabia. These variables are:

- GDP: Gross domestic product, representing the economic activity and output of Saudi Arabia. A key indicator of economic health.
- CO_2: Carbon dioxide emissions, used to assess environmental impact, particularly relevant in the context of sustainable development.
- HE: Health expenditure, used as a proxy for health system performance.
- RE: Renewable energy share, indicating the share of energy from renewable sources, reflecting the commitment to green energy.
- Mob: Number of mobile users, showing the penetration of mobile technology in the population.
- Log: Logistics performance index, evaluating the efficiency of logistics systems within the country.
- Int: Number of internet users, to measure the extent of internet connectivity and digital inclusion.
- Pat: Patent applications, serving as a proxy for technological innovation through the volume of patents filed.

These indicators have been specifically chosen for their potential influence on health expenditure (HE), which represents our target variable in this study, making them crucial for assessing the impact of technological innovation and sustainable development on health system performance. This choice is confirmed by our initial correlation study as shown in both Figure 1 and Table 1, where the correlation between the target variable HE and the remaining indicators ranges between 0.97 and 0.65 for all variables except for RE, which was equal to 0.31. In addition, the p-values for all variables are significant at the 0.01 level except for RE. These high coefficients demonstrate the relevance of the chosen variables for this study. Further impact studies will be discussed in the following sections.

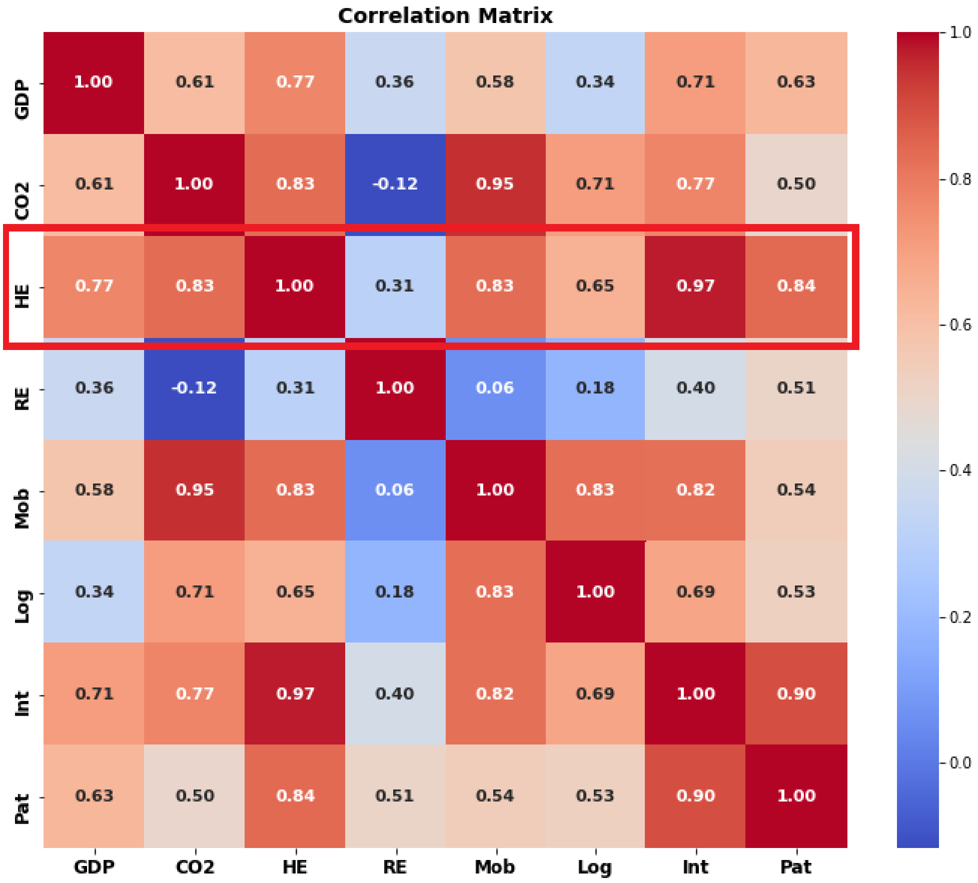

Figure 1. Correlation matrix between all study variables.

Table 1. Correlation (with related p-value) between HE and the rest of the variables (Pearson coefficient). Note that ***, **, and * denote significance levels of 0.01, 0.05, and 0.1, respectively.

Correlation	GDP	CO2	RE	Mob	Log	Int	Pat
HE	0.77***	0.83***	0.31*	0.83***	0.65***	0.97***	0.84***
P-values	1.67e-07	2.27e-09	7.63e-02	1.70e-09	4.44e-05	1.27e-21	9.13e-10

3.2. Data preprocessing

The preprocessing of the dataset was meticulously executed to address several key issues: missing data, outliers, and range anomalies. These challenges were systematically tackled to ensure the integrity and utility of the data for complex analyses. Two main operations were applied, as shown in the methodology diagram (Figure 2), specifically missing data handling and data normalization.

Missing data handling: Missing values in the dataset were managed using both backward and forward-filling methods. This imputation strategy ensures that no data points are left blank, preserving the continuity and reliability of time-series data across the study period.

Data normalization: To achieve uniformity and eliminate scale biases during machine learning analysis, a StandardScaler was applied to all input features. This transformation standardizes each feature to have a mean of zero and a standard deviation of one. By normalizing the data, we reduce the impact of outliers, which can skew analysis and lead to inaccurate models. The standardized distribution of features ensures that the measurement scale is consistent across different variables, enhancing the reliability of multivariate analysis. This step is crucial for ensuring that the models developed are not only accurate but also generalizable across different sets of data.

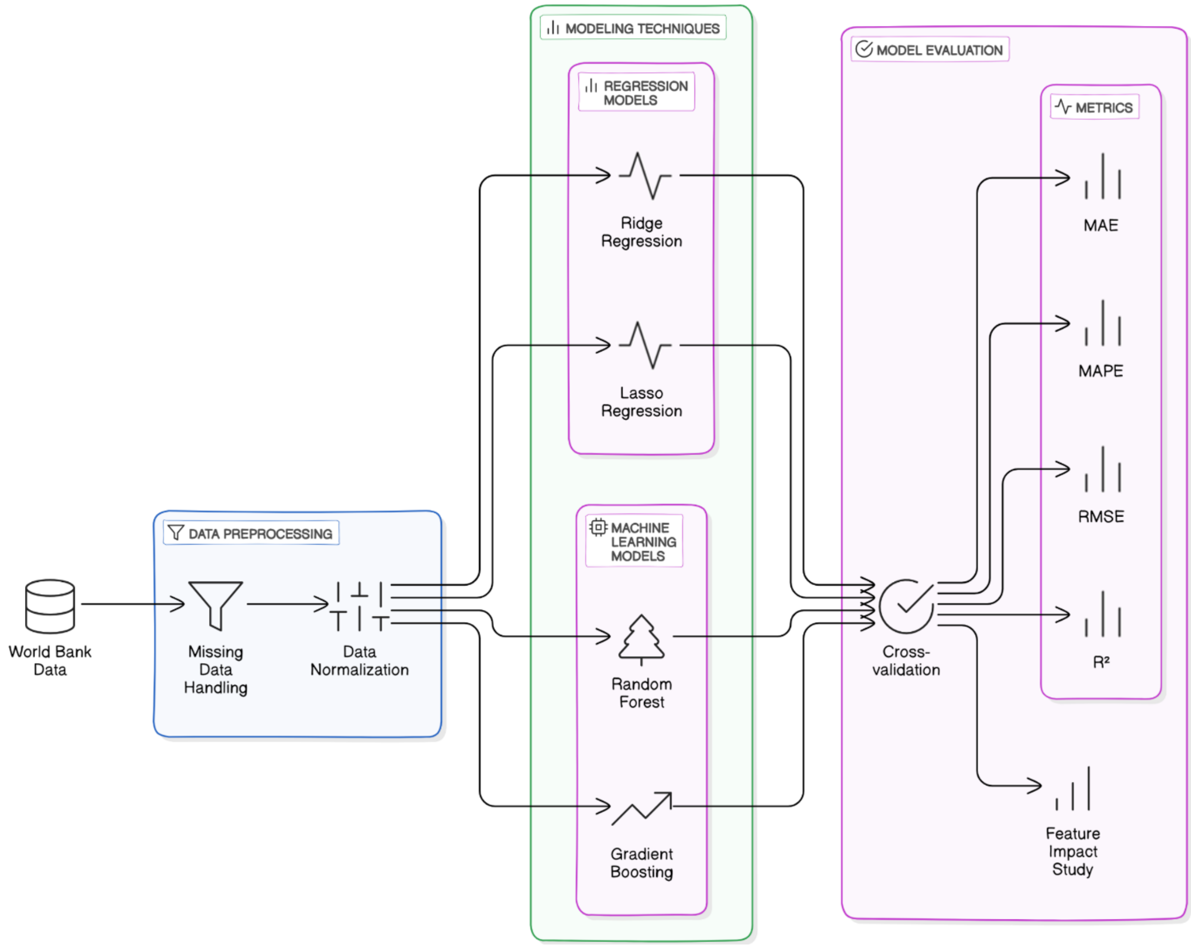

Figure 2. The adopted methodology from data preprocessing to model evaluation.

3.3. Modeling techniques

The primary objective of our modeling efforts is to estimate the target variable, health expenditure (HE), based on seven key input features: GDP, CO_2 emissions, renewable energy share, number of mobile users, logistics performance index, number of internet users, and patent applications. We aim to thoroughly investigate the influence of each factor on health expenditure, offering insights into how technological innovation and sustainable practices impact health system performance.

For this analysis, as shown in the methodology diagram (Figure 2), two principal categories of models were employed: (1) regression models, specifically ridge and lasso [39], and (2) machine learning models, namely random forest [40] and gradient boosting [41]. These methodologies were chosen to address the complexities inherent in the dataset, particularly the potential presence of multicollinearity among variables (independent variables are highly correlated).

Ridge and lasso regression models are well-suited for scenarios where multicollinearity might otherwise distort the estimates of classic linear regression. Both models apply a penalty to the coefficients of the regression variables, thus mitigating the issue of multicollinearity. Similarly, advanced machine learning techniques, such as random forest and gradient boosting, are inherently designed to handle large datasets with complex and nonlinear relationships without being heavily influenced by multicollinearity. This makes them exceptionally useful for predictive accuracy and model robustness in empirical research where traditional methods might fail or yield biased results.

3.3.1. Regression models

Regression models were chosen for their ability to provide quantifiable insights into the relationships between independent variables and the dependent variable. Regression techniques, such as ridge and lasso, offer robust methods for handling multicollinearity and model overfitting. Ridge regression helps in reducing the model complexity by imposing a penalty on the size of coefficients, whereas lasso regression is effective in feature selection by reducing the coefficients of less significant variables to zero. Further details are given in the following subsections.

Ridge regression

Ridge regression, a variant of regularized linear regression, effectively handles multicollinearity by adding a penalty proportional to the sum of the squares of the coefficients. This approach moderates large coefficients, thus stabilizing solutions and enhancing generalizability. Differing from traditional ordinary least squares (OLS) regression, which minimizes the sum of the squared residuals to fit the model, ridge regression integrates an additional penalty term ($\alpha||\beta||^2$) to curb overfitting. The penalty term, governed by the tuning parameter α, ensures a balanced approach to maintaining model accuracy and complexity. The objective function for ridge regression, aiming for minimization, is defined as:

$$\text{Minimize: } ||Y - X\beta||^2 + \alpha||\beta||^2 \quad (1)$$

Where:

- Y is the response vector, i.e., the actual observed values we aim to predict.
- X is the matrix of input features (predictors), where each column is a feature, and each row represents one data instance.
- β represents the coefficients or parameters of the model, which quantify the influence of each predictor X on the response Y.
- α is the regularization parameter controlling the penalty strength.

The L2 norm penalty ($\alpha||\beta||^2$) ensures that the coefficients are shrunk toward zero but typically keeps all variables in the model, hence preserving the full interpretability of the predictors.

Lasso regression

Lasso regression is another type of linear regression that uses regularization mechanisms. Following ridge regression, lasso regression also modifies the traditional OLS approach by incorporating a regularization penalty. However, unlike ridge, which uses an L2 penalty ($\alpha||\beta||^2$), lasso applies an L1 penalty ($\alpha|\beta|$). This penalty encourages sparsity in the coefficient matrix by allowing some coefficients to shrink entirely to zero. Consequently, lasso not only helps reduce overfitting but also performs feature selection, identifying variables that make the most significant contribution to predicting the output. The objective function for lasso regression, which aims to balance model accuracy with complexity, is structured to minimize the following:

$$\text{Minimize: } \left||Y - X\beta|\right|^2 + \alpha|\beta| \quad (2)$$

Hence, this formulation forces some coefficients to zero, effectively selecting a simpler, more interpretable model that still captures essential patterns in the data.

3.3.2. Machine learning models

The machine learning models were selected due to their ability to model complex nonlinear relationships and interactions between variables that are not easily captured by traditional regression techniques. Models such as random forest and gradient boosting regressors are particularly valuable in this context. Random forest leverages an ensemble of decision trees to improve prediction accuracy and control overfitting, providing a comprehensive measure of feature importance. Gradient boosting, on the other hand, sequentially builds trees to correct the predecessors' errors, which enhances the model's performance. These techniques are adept at handling complex datasets, making them ideal for extracting nuanced patterns and predictive insights from the data. Further details are given in the following subsections.

Random forest

Random forest is an ensemble learning method that combines predictions from multiple decision trees, built on different subsamples of the dataset. By averaging the results of individual trees, random forest reduces the variance of the predictions, leading to improved accuracy and robustness against overfitting. The general equation for a prediction using a random forest model can be described as the average of all the individual decision tree predictions:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} f_b(X) \quad (3)$$

Where:

- $\hat{y}$ is the value predicted by the model,
- B is the number of trees,
- f_b is the prediction function of the b-th tree,
- X represents the input features.

Random forest models are particularly valuable when dealing with datasets with high-dimensional feature spaces and complex interaction structures among features.

Gradient boosting

Gradient boosting builds models incrementally using an ensemble of weak learning models, typically decision trees. It improves model predictions by focusing on errors of previous models and minimizing a loss function iteratively. The equation for gradient boosting, which is used to incrementally improve model predictions, can be represented as:

$$F(x) = F_0(x) + \sum_{m=1}^{M} \gamma_m h_m(x) \quad (4)$$

Where:

- $F_0(x)$ represents the initial model, which provides a basic starting point for the boosting process.
- M represents the total number of boosting stages or the number of sequential trees to be added. Each tree tries to improve on the areas where the previous model residuals were high.
- $h_m(x)$ denotes the output of the m-th base learner applied to input x. Base learners are typically decision trees, particularly shallow trees (i.e., trees with only a few levels), which are weak learners. These are trained on the residuals—differences between observed and predicted values—from the preceding model.
- γ_m is a coefficient or weight for the m-th base learner in the ensemble.

This method is particularly effective in scenarios where the relationship between predictors and the response variable is complex and nonlinear, requiring adaptive fitting procedures.

By integrating both regression and machine learning models, our approach not only underscores the significance of each predictor but also ensures a robust and comprehensive analysis, thereby enabling more accurate predictions and strategic insights into health expenditure trends.

3.4. Model evaluation

The performance of the models was rigorously assessed using several metrics to evaluate the quality of the prediction task. These metrics help quantify how well the models predict health expenditure, focusing on various aspects of prediction accuracy and reliability. These metrics are MAE, MAPE, RMSE, and R^2. The description of each metric is as follows:

Mean absolute error (MAE): Measures the average magnitude of the errors in a set of predictions, without considering their direction. It is calculated as:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}| \tag{5}$$

Where y_i is the actual value and $\hat{y}_i$ is the predicted value for the i-th observation, and n is the number of observations.

Mean absolute percentage error (MAPE): This metric provides a measure of prediction accuracy expressed as a percentage, offering a view of the error relative to the actual values. It is computed as:

$$\mathrm{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{6}$$

Root mean squared error (RMSE): Provides a measure of the magnitude of the error by squaring the average errors to avoid canceling out negative values, thereby emphasizing larger errors. It is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2} \tag{7}$$

R-squared (R^2): Indicates the proportion of variance in the dependent variable that is predictable from the independent variables. Knowing that $\bar{y}$ is the mean of the observed data, it is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

These metrics collectively provide a comprehensive view of model performance, highlighting both the accuracy and the effectiveness of the models in capturing the variability of the target variable.

3.5. Application

In our modeling of health expenditure, we remember that we first applied a series of preprocessing steps to address missing values and normalize the input features. We designated all variables as features, except HE (health expenditure), which was the target variable. To train and test our models, we employed 5-fold cross-validation, a method that enhances the robustness and generalizability of the results by systematically rotating each subset of the dataset (20%) as a test set while training on the remaining data (80%). This approach not only prevents overfitting but also ensures that each segment of the dataset is used for both training and validation, providing a comprehensive assessment of model performance.

For all models, extensive hyperparameter tuning was conducted, testing various parameters across different configurations to identify the optimal settings for each model and ensure peak performance. It is important to note that in regression models, the alpha parameter (α) significantly influences the regularization process. This process, as previously described, is a mechanism used to prevent overfitting by adding a penalty to the loss function. Specifically, ridge regression utilizes an alpha of 0.01 for L2 regularization, helping to prevent overfitting while retaining important predictors. Similarly, lasso regression employs an alpha of 0.01 for L1 regularization to enhance sparsity and feature selection. This effectively reduces the coefficients of less influential variables to zero, which is ideal for models that benefit from feature reduction.

Additionally, two advanced machine learning techniques were deployed: a random forest regressor, which was configured with 100 estimators to effectively balance nonlinear relationships and control variance, and a gradient boosting regressor. The gradient boosting model was meticulously set with 100 boosting stages and a learning rate of 0.1 to ensure precise error correction across sequences, thereby minimizing the risk of overfitting. The efficacy of these models was systematically assessed using the previously mentioned metrics: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and R-squared (R^2). These indicators have provided valuable insights into the accuracy and predictive capabilities of the models.

Please note that for technical implementation, all data operations and modeling in this study were carried out using Python, known for its powerful data science utilities. We employed key libraries including NumPy for numerical computations, Pandas for efficient data handling, and Scikit-learn for implementing and evaluating regression and machine learning models, ensuring a robust analytical framework.

4. Results and discussion

4.1. Model performance

In this study, we employed a list of regression and machine learning models to evaluate health expenditure based on a set of predictors. The regression models included ridge regression and lasso regression, while the machine learning models comprised random forest regressor and gradient boosting regressor. These models were assessed using four key performance metrics: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and R-squared (R^2). The following table summarizes the performance of each model across the four metrics:

Table 2. The four models metrics.

Model/Metric	MAE	MAPE	RMSE	R^2
Ridge	41.44	10.91	50.94	0.94
Lasso	41.74	11.00	51.26	0.94
Random forests	43.94	10.52	59.37	0.91
Gradient boosting	44.06	10.01	65.82	0.85

The results of the study show that both ridge and lasso regression models performed similarly, with ridge having a slight edge, with an R^2 of 0.94, MAE of 41.44, MAPE of 10.91, and RMSE of 50.94, indicating its robustness against multicollinearity among predictors due to its regularization techniques. The comparable R^2 values for these models suggest they explained similar variances in health expenditure, effectively capturing the linear relationships among the predictors. On the other hand, the random forest model, despite having a slightly poorer performance on MAE of 43.94 and RMSE of 59.37, also exhibited a smaller R^2 of 0.91 compared to the regression models, yet it maintained a competitive MAPE of 10.52, indicating its strength in capturing nonlinear relationships and complex interactions between predictors, which might not be fully recognized by linear models. Lastly, the gradient boosting model, while showing the highest MAE and RMSE of 44.06 and 65.82, respectively, recorded the best MAPE of 10.01, indicating its effectiveness in predictive accuracy relative to the actual values. However, its significantly lower R^2 of 0.85 suggests that it might require further tuning of parameters such as learning rate or tree depth to optimize performance and enhance its predictive accuracy in the context of health expenditure forecasting.

Overall, the high R^2 values across all models confirm that the selected input features—such as the number of internet users, GDP, or CO_2 emissions—effectively explain health expenditure, validating our choice of both selected variables and applied models. This strong model fit not only highlights the relevance of these features in capturing the complex relationships between economic activities, environmental factors, and health outcomes but also confirms the suitability of the modeling approaches for this analysis.

4.2. Variables' impact

From the previous results, it is evident that the selected indicators significantly influence health expenditure predictions, providing crucial insights for targeted policymaking. A key advantage of machine learning models is their ability to not only offer performance metrics but also to reveal the significance of each feature in the prediction process. The feature importances for both random forest and gradient boosting models are highlighted in the following table, underscoring the differential impact of each variable on health expenditure.

Table 3. Feature importance coefficient for machine learning models.

Feature/model	Random forest	Gradient boosting
Int	0.673	0.860
GDP	0.145	0.082
Mob	0.070	0.034
CO2	0.065	0.018
Pat	0.035	0.003
Log	0.011	0.002
RE	0.000	0.000

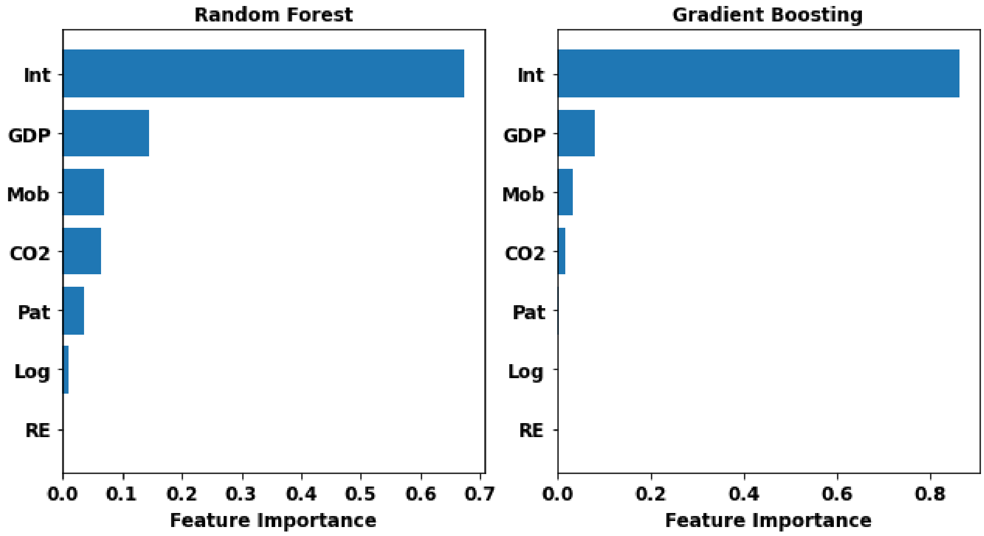

Figure 3. Features importance for machine learning models.

As shown in Table 3 and Figure 3, feature importance analysis from the machine learning models clearly identifies the key variables affecting health expenditure. Internet users (Int) is the most significant predictor in both random forest and gradient boosting models, with importance scores of 0.673 and 0.860, respectively, highlighting the strong impact of digital connectivity on health spending. GDP also influences health costs, but to a lesser extent, with scores of 0.145 and 0.082. Mobile user penetration (Mob) and CO_2 emissions (CO2) have moderate to low importance, suggesting nuanced interactions with health expenditure. Patent applications (Pat) and logistics performance (Log) show minimal impact, indicating a limited direct influence on health costs within this study. Similarly, renewable energy (RE) share has negligible importance, reflecting its minimal impact within the analyzed data.

For the feature importance study related to regression models, as shown in Table 4 and Figure 4, internet users emerge as the most impactful predictor, with ridge assigning a coefficient of 265.64 and lasso of 270.80, highlighting the substantial costs associated with expanding digital health services. In fact, this correlation can be attributed to several factors. First, the expansion of digital health services requires significant investments in infrastructure such as broadband and digital platforms, which results in high initial and operational costs. Second, as access to the Internet expands, the demand for e-health services increases, and this initially results in increased health expenditures as the system expands to accommodate new care modalities. In addition, this investment stimulates economic activity in the IT, telecommunications, and health innovation sectors, which can boost overall economic growth and, subsequently, health spending. However, this initial cost increase is an investment in transforming healthcare delivery into a more efficient, effective, and accessible system. Over time, these digital health services should lead to savings through better disease management, reduced face-to-face visits, and better allocation of resources, in line with broader goals of health system sustainability and improved patient outcomes. CO_2 emissions follow as the second most influential factor, with coefficients of 115.06 in ridge and 116.02 in lasso, indicating significant health-related expenses due to environmental pollution.

Economically, high CO_2 emissions are closely linked to poor air quality and increased healthcare costs by increasing the prevalence of respiratory and cardiovascular diseases. This, in turn, leads to higher rates of hospitalization and prolonged treatment, which increases overall health expenditures. In addition, preventive health measures and public health campaigns to mitigate the effects of pollution entail substantial costs. Long-term exposure to high CO_2 levels also contributes to chronic diseases, which further increases healthcare expenditures over time. Beyond direct medical costs, high levels of pollution can decrease economic productivity by increasing morbidity and mortality, which will indirectly increase demands for healthcare. Governments often incur significant expenditures to implement and enforce environmental regulations aimed at controlling and reducing CO_2 emissions, and this reflects the broader economic impacts of managing environmental pollutants. These findings highlight the critical need for strong environmental policies and investments in cleaner technologies; these could reduce the burden on healthcare by reducing the adverse health effects of CO_2 emissions.

Mobile user penetration inversely affects health expenditure, demonstrated by coefficients of -106.86 (ridge) and -111.19 (lasso), suggesting that increased mobile access significantly reduces health costs. This reduction is due to enhanced access to health information, efficient disease management via remote monitoring, and broader reach into underserved areas. These factors collectively improve healthcare efficiency and cost-effectiveness, underscoring the strategic value of mobile technology in healthcare optimization.

From the economic side, this impact is attributed to several transformative aspects of mobile technology in healthcare. First, improved access to health information via mobile devices allows individuals to learn about preventive measures and disease management, potentially reducing the need for expensive medical interventions. Second, mobile technology facilitates effective disease management through remote monitoring, which is particularly beneficial for chronic diseases. This reduces the need for frequent hospital visits, travel costs, and time spent by patients and healthcare providers. In addition, the ability of mobile technology to reach underserved areas helps reduce healthcare disparities by providing essential health services to those isolated. This is particularly critical in rural or remote areas. In addition, mobile technology streamlines healthcare operations, reduces administrative burdens, and improves patient data management through electronic medical records accessible via mobile devices. This leads to operational efficiencies and profitability. Overall, integrating mobile technology in healthcare delivery not only improves accessibility and quality of health services but also significantly reduces associated costs, underlining its strategic value in healthcare optimization and economic sustainability. GDP shows a positive influence but to a lesser extent, with coefficients of 25.66 (ridge) and 25.63 (lasso), reflecting the direct relationship between economic growth and health spending. In fact, this relationship indicates that GDP growth is accompanied by increased health expenditure, even if the impact is less pronounced than that of technological advances or environmental impacts. Higher GDP means more public and private funding for health, which expands health infrastructure and access to services. In addition, economic growth increases disposable income, strengthening individuals' ability to afford better and more comprehensive health services. However, the moderate influence of GDP on health expenditure shows that economic growth alone does not dictate levels of health spending, and this suggests a nuanced interaction of various socio-economic factors. Patent applications have a moderately negative impact, where healthcare innovation appears to reduce costs, evidenced by coefficients of -35.80 (ridge) and -39.03 (lasso).

Logistics performance, although less impactful, positively correlates with expenditure, with coefficients of 6.48 (ridge) and 7.58 (lasso), suggesting that enhanced logistics contribute to higher health service costs. This relationship can be attributed to the fact that improved logistics enhance the distribution and availability of medical supplies and services, leading to potentially higher utilization and associated costs. This expansion increases healthcare use as facilities can offer a broader range of treatments and manage larger volumes of medical supplies. In addition, the infrastructure required to support advanced logistics, such as sophisticated transportation and warehousing technologies, represents significant investment costs.

Renewable energy has the least influence, with minor positive coefficients of 5.02 (ridge) and 4.79 (lasso), possibly linked to investments in health-related environmental sustainability. Actually, the marginal positive impact may be ascribed to the initial capital investment in renewable energy infrastructure in healthcare facilities, which is compensated by the long-term cost reductions resulting from decreased energy expenditures. In addition, using renewable energy contributes to improving public health outcomes by reducing emissions and pollution, which could reduce healthcare demands related to pollution-induced diseases over time. However, these environmental health benefits and their economic impacts on health expenditures may not be immediately apparent. Ultimately, although the direct financial influence of renewables on current healthcare costs appears limited, they play a crucial role in promoting environmental sustainability and alignment with broader public health objectives, suggesting that their importance extends beyond the immediate economic impacts. This ordered analysis underscores the critical areas for policy intervention, particularly in digital health technology and environmental management, to effectively manage health costs.

Table 4. Features coefficients for regression models.

Feature	Ridge coefficient	Lasso coefficient
Internet users (Int)	265.64	270.80
CO_2 emissions (CO2)	115.06	116.02
Mobile users (Mob)	-106.86	-111.19
Patent applications (Pat)	-35.80	-39.03
GDP	25.66	25.63
Logistics performance (Log)	6.48	7.58
Renewable energy (RE)	5.02	4.79

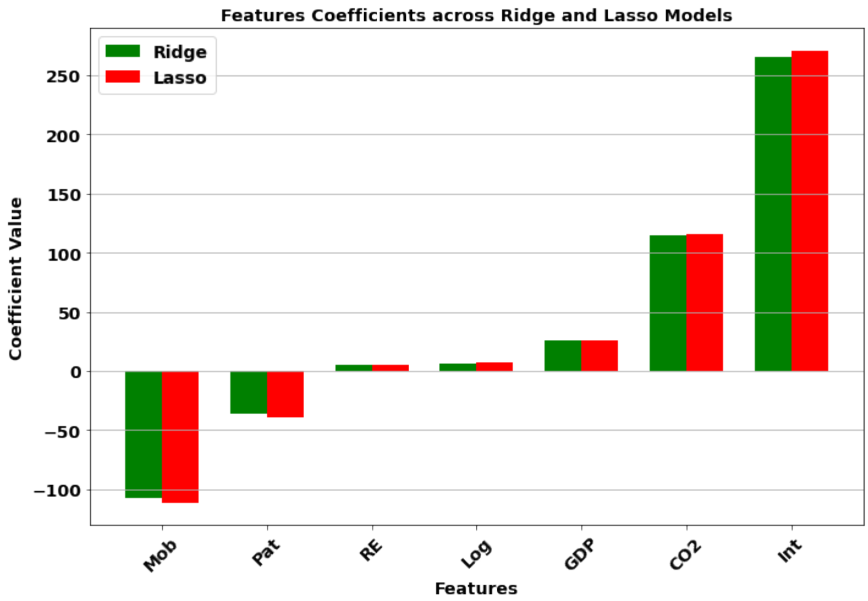

Figure 4. Features coefficients across ridge and lasso models.

Further, the different impacts of mobile users (-106.86 in ridge and -111.19 in lasso) and internet users (265.64 in ridge and 270.80 in lasso) on health expenditure can be attributed to how each technology is utilized within the healthcare system. Mobile technology, typically used for preventive measures and basic health management, helps reduce costs by promoting early intervention and efficient care management, particularly in rural or underprivileged areas. Conversely, internet usage, associated with higher expenditures, enables access to more comprehensive online health services, such as telemedicine and advanced diagnostic tools, which increase healthcare spending while enhancing service quality. Additionally, the infrastructure costs associated with integrating internet technologies in hospitals and clinics contribute to this expenditure, as establishing and maintaining robust digital services requires significant investment. Both technologies improve healthcare access but have distinct cost implications due to their different roles and the breadth of services they enable.

5. Conclusion and policy implications

This investigation contributes to the empirical literature by exploring the impact of recent innovations, logistical measures, ICT diffusion, environmental quality improvements, economic growth, and green energy exploitation on the healthcare system and well-being in the context of Saudi Arabia, thereby providing a more nuanced understanding of the mechanisms driving health performance improvements under the Vision 2030 framework. Further, while traditional econometric tools have been extensively used to analyze health outcomes and the effectiveness of various healthcare interventions, there is a notable research gap in applying machine learning techniques within this context. The data encompasses a period from 1990 to 2022, allowing for a comprehensive study of trends over more than three decades. Using the machine learning procedure, we find that internet users have a strong impact of digital connectivity on health spending with scores of 0.673 and 0.86. Further, GDP also influences health costs, but to a lesser extent, with scores of 0.145 and 0.082. Mobile user penetration (Mob) and CO_2 emissions (CO2) have moderate to low importance, suggesting nuanced interactions with health expenditure. Patent applications and logistics performance show minimal impact, indicating a limited direct influence on health costs within this study. Similarly, the share of renewable energy is negligible, reflecting its minimal impact on the analyzed data. Finally, regression models, namely ridge and lasso, also exhibited similar outcomes, reinforcing the conclusions drawn from the machine learning analyses.

These results enable us to recommend several practical and policy implications for the policymakers of Saudi Arabia. First, managing the growing expenses of digital health services by implementing public-private partnerships is one policy outcome that may be considered. Through using the resources and expertise of the private sector, these partnerships can significantly lower costs while providing improved levels of service. The government should also offer grants or subsidies to digital health companies to encourage innovation, spark creativity, and expedite development. Second, it is highly recommended for authorities to strengthen the control of industrial emissions and carry out stronger environmental standards. Moreover, concentrating resources on green infrastructure and promoting the application of renewable energies can also effectively lower pollution levels. Environmental protection departments should extensively promote health education campaigns focused on the risks of air pollution. Increasing public awareness about these risks can help communities adopt preventive measures, ultimately reducing healthcare costs associated with pollution-related health conditions. This approach supports governmental efforts to effectively manage public health expenditures. Finally, to encourage investment in advanced supply chain technology, the authorities should actively sponsor enterprises that enhance efficiency and reduce operating costs for tandem operators. Furthermore, promoting domestic manufacturing of pharmaceuticals and medical supplies not only decreases reliance on costly imports but also strengthens supply chain robustness and stimulates domestic economic growth. Governments can further support this industry by offering subsidies or favorable tax conditions to improve logistics within the healthcare sector, potentially easing financial pressures and enhancing overall industry performance.

To conclude, it is essential to mention that this research does not take into account all factors that influence health performance. Future studies might examine the impact of additional macroeconomic variables on health performance and well-being in various countries and groups of states.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors gratefully acknowledge Qassim University, represented by the Deanship of Graduate Studies and Scientific Research, on the financial support for this research under the number (2023-SDG-l-HSRC-36388) during the academic year 1445 AH / 2023 AD.

Conflict of interest

The authors declare no conflict of interest.

References

1. De Kock JH, Latham HA, Leslie SJ, et al. (2021) A rapid review of the impact of COVID-19 on the mental health of healthcare workers: implications for supporting psychological well-being. *BMC Public Health* 21: 1–18. https://doi.org/10.1186/s12889-020-10070-3
2. Hoogendoorn CJ, Rodríguez ND (2023) Rethinking dehumanization, empathy, and burnout in healthcare contexts. *Curr Opin Behav Sci* 52: 101285. https://doi.org/10.1016/j.cobeha.2023.101285
3. Hee Lee D, Yoon SN (2021) Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. *Int J Environ Res Public Heal* 18: 271. https://doi.org/10.3390/ijerph18010271
4. Monteiro S, Amor-Esteban V, Lemos K, et al. (2023) Are we doing the same? A worldwide analysis of business commitment to the SDGs. *AIMS Environ Sci* 10: 446–466. https://doi.org/10.3934/environsci.2023025
5. Rieiro-García M, Amor-Esteban V, Aibar-Guzmán C, et al. (2023) 'Localizing' the sustainable development goals: a multivariate analysis of Spanish regions. *AIMS Environ Sci* 10: 356–381. https://doi.org/10.3934/environsci.2023021
6. Omri A, kahouli B, Kahia M (2024) Environmental sustainability and health outcomes: Do ICT diffusion and technological innovation matter? *Int Rev Econ Financ* 89: 1–11. https://doi.org/10.1016/j.iref.2023.09.007
7. Omri A, Kahouli B, Afi H, et al. (2023) Impact of Environmental Quality on Health Outcomes in Saudi Arabia: Does Research and Development Matter? *J Knowl Econ* 14: 4119–4144. https://doi.org/10.1007/s13132-022-01024-8
8. Omri A, Kahouli B, Kahia M (2023) Impacts of health expenditures and environmental degradation on health status—Disability-adjusted life years and infant mortality. *Front Public Heal* 11: 1118501. https://doi.org/10.3389/fpubh.2023.1118501
9. Al-Hanawi MK, Qattan AMN (2019) An Analysis of Public-Private Partnerships and Sustainable Health Care Provision in the Kingdom of Saudi Arabia. https://doi.org/101177/1178632919859008.
10. Al Saffer Q, Al-Ghaith T, Alshehri A, et al. (2021) The capacity of primary health care facilities in Saudi Arabia: infrastructure, services, drug availability, and human resources. *BMC Health Serv Res* 21: 1–15. https://doi.org/10.1186/s12913-021-06355-x
11. Alasiri AA, Mohammed V (2022) Healthcare Transformation in Saudi Arabia: An Overview Since the Launch of Vision 2030. *Heal Serv Insights* 15. https://doi.org/10.1177/11786329221121214
12. Alkhamis A, Miraj SA (2021) Access to Health Care in Saudi Arabia: Development in the Context of Vision 2030. *Handb Healthc Arab World* 1629–1660. https://doi.org/10.1007/978-3-030-36811-1_83
13. Alluhidan M, Tashkandi N, Alblowi F, et al. (2020) Challenges and policy opportunities in nursing in Saudi Arabia. *Hum Resour Health* 18: 1–10. https://doi.org/10.1186/s12960-020-00535-2
14. Alnowibet K, Abduljabbar A, Ahmad S, et al. (2021) Healthcare Human Resources: Trends and Demand in Saudi Arabia. *Healthc* 9: 955. https://doi.org/10.3390/healthcare9080955

15. Hejazi MM, Al-Rubaki SS, Bawajeeh OM, et al. (2022) Attitudes and Perceptions of Health Leaders for the Quality Enhancement of Workforce in Saudi Arabia. *Healthc* 10: 891. https://doi.org/10.3390/healthcare10050891
16. Qureshi NA, Al-Habeeb AA, Koenig HG (2013) Mental health system in Saudi Arabia: an overview. *Neuropsychiatr Dis Treat* 9: 1121–1135. https://doi.org/10.2147/NDT.S48782
17. Rahman R (2020) The Privatization of Health Care System in Saudi Arabia. https://doi.org/101177/1178632920934497.
18. Rahman R, Al-Borie HM (2021) Strengthening the Saudi Arabian healthcare system: Role of Vision 2030. *Int J Healthc Manag* 1–9. https://doi.org/10.1080/20479700.2020.1788334
19. Rahman R, Qattan A (2021) Vision 2030 and Sustainable Development: State Capacity to Revitalize the Healthcare System in Saudi Arabia. *Inq (United States)* 58. https://doi.org/10.1177/0046958020984682
20. Alharthi M, Hanif I, Alamoudi H (2022) Impact of environmental pollution on human health and financial status of households in MENA countries: Future of using renewable energy to eliminate the environmental pollution. *Renew Energy* 190: 338–346. https://doi.org/10.1016/j.renene.2022.03.118
21. Ampon-Wireko S, Zhou L, Xu X, et al. (2021) The relationship between healthcare expenditure, CO2 emissions and natural resources: evidence from developing countries. 11: 272–286. https://doi.org/10.1080/21606544.2021.1979101
22. Chen J, Su F, Jain V, et al. (2022) Does Renewable Energy Matter to Achieve Sustainable Development Goals? The Impact of Renewable Energy Strategies on Sustainable Economic Growth. *Front Energy Res* 10: 829252. https://doi.org/10.3389/fenrg.2022.829252
23. da Silva RGL, Chammas R, Novaes HMD (2021) Rethinking approaches of science, technology, and innovation in healthcare during the COVID-19 pandemic: the challenge of translating knowledge infrastructures to public needs. *Heal Res Policy Syst* 19: 1–9. https://doi.org/10.1186/s12961-021-00760-8
24. Fathollahi-Fard AM, Ahmadi A, Karimi B (2022) Sustainable and Robust Home Healthcare Logistics: A Response to the COVID-19 Pandemic. *Symmetry* 14: 193. https://doi.org/10.3390/sym14020193
25. Jiang C, Chang H, Shahzad I (2022) Digital Economy and Health: Does Green Technology Matter in BRICS Economies? *Front Public Heal* 9: 827915. https://doi.org/10.3389/fpubh.2021.827915
26. Karaaslan A, Çamkaya S (2022) The relationship between CO2 emissions, economic growth, health expenditure, and renewable and non-renewable energy consumption: Empirical evidence from Turkey. *Renew Energy* 190: 457–466. https://doi.org/10.1016/j.renene.2022.03.139
27. Li J, Irfan M, Samad S, et al. (2023) The Relationship between Energy Consumption, CO2 Emissions, Economic Growth, and Health Indicators. *Int J Environ Res Public Heal* 20: 2325. https://doi.org/10.3390/ijerph20032325
28. Magazzino C, Mele M, Schneider N (2022) A new artificial neural networks algorithm to analyze the nexus among logistics performance, energy demand, and environmental degradation. *Struct Chang Econ Dyn* 60: 315–328. https://doi.org/10.1016/j.strueco.2021.11.018
29. Paul M, Maglaras L, Ferrag MA, et al. (2023) Digitization of healthcare sector: A study on privacy and security concerns. *ICT Express* 9: 571–588. https://doi.org/10.1016/j.icte.2023.02.007

30. Rahman MM, Alam K (2022) Life expectancy in the ANZUS-BENELUX countries: The role of renewable energy, environmental pollution, economic growth and good governance. *Renew Energy* 190: 251–260. https://doi.org/10.1016/j.renene.2022.03.135
31. Saleem H, Khan MB, Shabbir MS, et al. (2022) Nexus between non-renewable energy production, CO2 emissions, and healthcare spending in OECD economies. *Environ Sci Pollut Res* 29: 47286–47297. https://doi.org/10.1007/s11356-021-18131-9
32. Shin H, Kang J (2020) Reducing perceived health risk to attract hotel customers in the COVID-19 pandemic era: Focused on technology innovation for social distancing and cleanliness. *Int J Hosp Manag* 91: 102664. https://doi.org/10.1016/j.ijhm.2020.102664
33. Small GW, Lee J, Kaufman A, et al. (2022) Brain health consequences of digital technology use. 22: 179–187. https://doi.org/10.31887/DCNS.2020.22.2/gsmall
34. Vicente MR (2022) ICT for healthy and active aging: The elderly as first and last movers. *Telecomm Policy* 46: 102262. https://doi.org/10.1016/j.telpol.2021.102262
35. Zhang J, Gong X, Zhang H (2022) ICT diffusion and health outcome: Effects and transmission channels. *Telemat Informatics* 67: 101755. https://doi.org/10.1016/j.tele.2021.101755
36. Zhang Q, Guo X, Vogel D (2022) Examining the health impact of elderly ICT use in China. 28: 451–467. https://doi.org/10.1080/02681102.2022.2048782
37. Priyan S, Matahen R, Priyanshu D, et al. (2024) Environmental strategies for a healthcare system with green technology investment and pandemic effects. *Innov Green Dev* 3: 100113. https://doi.org/10.1016/j.igd.2023.100113
38. Ijiga AC, Peace AE, Idoko IP, et al. (2024) Technological innovations in mitigating winter health challenges in New York City, USA. *Int J Sci Res Arch* 11: 535–551. https://doi.org/10.30574/ijsra.2024.11.1.0078
39. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. *Source J R Stat Soc Ser B* 58: 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
40. Breiman L (2001) Random forests. *Mach Learn* 45: 5–32. https://doi.org/10.1023/A:1010933404324
41. Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *Ann Stat* 29: 1189–1232. https://doi.org/10.1214/aos/1013203451