Sigma J Eng Nat Sci, Vol. 42, No. 5, pp. 1797−1805, December, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00032

Research Article

Evaluation of COVID-19 mortality using machine learning regression methods based on health system indicators

Canan BULUT[1], Tuğba SÖNAL[2], Dilek KOLCA[3], Fatih TARLAK[4,*]

[1]*İstanbul Kültür University, Department of Medical Services and Techniques, İstanbul, 34158, Türkiye*
[2]*Health Ministry of Turkey, Ataşehir District Health Directorate, İstanbul, 34100, Türkiye*
[3]*İstinye University, Department of Health Management, İstanbul, 34010, Türkiye*
[4]*Gebze Technical University, Deparment of Bioengineering, Kocaeli, 41400, Türkiye*

ARTICLE INFO

Article history
Received: 11 September 2023
Revised: 22 October 2023
Accepted: 29 December 2023

Keywords:
COVID-19; Health Indicator; Prediction; Machine Learning

ABSTRACT

In the global health crisis caused by the COVID-19 pandemic, countries have faced significant challenges in combating the outbreak in terms of healthcare systems and economies. Evaluating the performance of healthcare systems in dealing with pandemics has become a priority for policymakers, healthcare providers, and the public alike. Assessing the performance of healthcare systems during the pandemic is crucial for preparedness and improvements in similar situations in the future. By identifying complex patterns and relationships, machine learning algorithms aim to uncover the relationship between healthcare system indicators and deaths due to the COVID-19 pandemic, using large and intricate datasets. These algorithms utilize various datasets containing demographic information and medical factors to reveal hidden relationships between various variables and disease severity. The objective of this study is to predict COVID-19 death rates for 27 OECD (Organisation for Economic Co-operation and Development) countries spanning the period from 2006 to 2019 using various machine learning regression methods. Healthcare system indicators, comprising accessibility, healthcare financing, and healthcare workforce, have been aggregated into three dimensions. The dataset includes COVID-19 death counts per a million-population due to the pandemic. Random forest regression, neural network regression, and Gaussian process regression were employed to forecast COVID-19 death rates, and the predictive capabilities of machine learning regression methods were evaluated using k-fold cross validation. The suitability of the algorithms was assessed using statistical measures such as the coefficient of determination (R^2) and root mean square error (RMSE). A high R^2 value and a low RMSE indicate that Gaussian process regression (GPR) can effectively predict COVID-19 death rates, taking various health indicators into account. Machine learning regression methods have revolutionized our understanding of COVID-19 death rates. Through prediction models, machine learning has empowered healthcare professionals with the ability to forecast death risks for individual patients, guiding decision-making processes and resource allocation. According to the research findings, to enhance the performance of healthcare systems in coping with global pandemics, there is a need to prioritize community-based healthcare services, adopt a social policy approach, encourage the use of advanced technology, ensure the trust of the public and healthcare workers, enhance social support opportunities, emphasize the importance of measures by leaders, and support global governance. Additionally, flexible supply chain plans for the procurement of personal protective equipment have been identified as necessary.

Cite this article as: Bulut C, Sönal T, Kolca D, Tarlak F. Evaluatıon of COVID-19 mortality using machine learning regression methods based on health system indicators. Sigma J Eng Nat Sci 2024;42(5):1797−1805.

***Corresponding author.**
*E-mail address: ftarlak@gtu.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

INTRODUCTION

The outbreak of the novel coronavirus disease (COVID-19) in late 2019 has rapidly spread worldwide, leading to a global health crisis. The virus has exhibited a wide range of clinical outcomes, from mild symptoms to severe complications and death in some cases. As COVID-19 continues to spread aggressively, it poses a serious threat to public health and has resulted in the depletion of medical resources and strain on healthcare workers [1-3].

The urgent need to identify factors associated with SARS-CoV-2 transmission, predictors of COVID-19 severity, and effective treatments has become paramount [4]. Resource-limited areas faced challenges in triaging life-saving therapies due to the high number of infections, emphasizing the importance of identifying patients requiring intensive care or at high risk of mortality [5, 6]. Furthermore, early administration of specific treatments has shown promise in reducing hospitalization duration and decreasing COVID-19 mortality, making it crucial to predict patients at high risk of disease progression and poor outcomes [7]. SARS-CoV-2, the virus responsible for COVID-19, is highly infectious and has spread rapidly across the globe. Its transmission dynamics, including asymptomatic cases and the transmission potential of individuals without symptoms, present unique challenges in controlling the pandemic [8]. Estimating the prevalence of COVID-19 is essential for effective pandemic management [9].

Machine learning, a subfield of artificial intelligence (AI), plays a crucial role in enabling computers to learn and make predictions or decisions without explicit programming [10]. In datasets with numerous independent variables, there might be complex and non-linear relationships that traditional statistical methods struggle to capture. Machine learning algorithms can learn and model intricate patterns within the data. Additionally, they are good with large datasets and a high number of variables, handling them faster than traditional statistical methods [11]. The process of machine learning involves training algorithms on labelled data, where each data point is associated with corresponding target values. Through this training process, algorithms adjust their internal parameters to minimize the disparity between their predicted outputs and the true target values [12]. The objective is to develop models that generalize well and can accurately predict or make decisions on new, unseen data. Machine learning algorithms have wide-ranging ap-plications and have demonstrated success in various domains, including healthcare, finance, natural language processing, computer vision, and recommendation systems [11, 13]. They have the potential to uncover insights and make accurate predictions in complex and data-rich environments, ultimately driving advancements and improving decision-making processes across industries [12-14].

In this study, the primary aim was to identify the relationship between COVID-19 death rates and healthcare system indicators among 27 countries within the OECD (Organisation for Economic Co-operation and Development) using machine learning regression methods. Healthcare system indicators spanning from 2006 to 2019, encompassing three dimensions accessibility, healthcare financing, and healthcare workforce, were considered as ten different variable datasets. A com-prehensive dataset containing healthcare indicators and COVID-19 death rates for a 14-year period for OECD countries was collected from various databases. To enhance the analysis, a standardization process was applied. Three machine learning regression methods, namely Random Forest Regression (RFR), Neural Network Regression (NNR), and Gaussian Process Regression (GPR), were employed to predict the relationship between COVID-19 death rates and healthcare system indicators. The performance of these methods was evaluated using k-fold cross-validation, and statistical measures such as the coefficient of determination (R^2) and root mean square error (RMSE) were utilized to assess their suitability.

MATERIALS AND METHODS

The study was conducted in four main steps. Firstly, healthcare indicators for OECD countries between 2006 and 2019 were collected from various databases such as World in Data, Worldometer, IHME-GHDx, and Eurostat. Secondly, data processing, especially standardization, was carried out using Matlab 8.3.0.532 (R2014a) software by MathWorks Inc. (Natick, MA, USA). Thirdly, data evaluation techniques such as correlation heat map and variable importance determination were applied. Finally, various machine learning-based regression methods were employed to explore the relationship between COVID-19 death rates and healthcare system indicators. The evaluation of machine learning-based regression methods involved assessing their predictive capabilities using multiple criteria, including the coefficient of determination and root mean square error. Figure 1 provides a visual representation of the research process, illustrating the main steps followed in this study. Subsequent subsections provide comprehensive explanations of each stage in this research.

Data Collection

Healthcare indicators for OECD countries spanning the years 2006 to 2019 were collected from databases such as World in Data, Worldometer, IHME-GHDx, and Eurostat. The used health status indicators, risk factors indicators and service coverage indicators with their explanations were given in Table 1. The study aims to evaluate COVID-19 mortality using a machine learning-based regression method based on health system indicators. Ten indicators, derived from the World Health Organization's 100 essential health indicators and grouped under the dimensions of health workforce, health access, and health financing, were utilized to determine these health system indicators. In the health workforce dimension, indicators such as healthcare

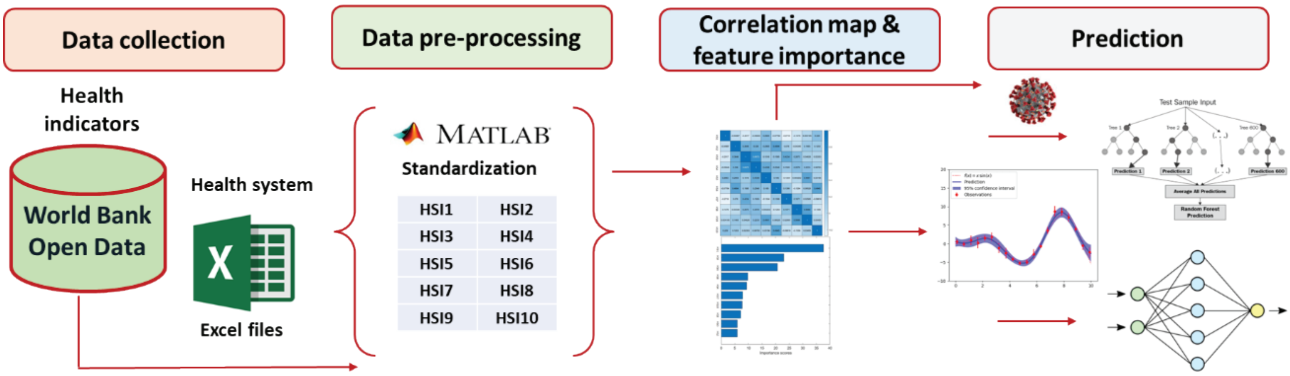

Figure 1. The flow chart outlining the main steps followed in the present study.

Table 1. The used health indicators with their explanations

Code	Health indicators	Explanation
HSI1	Health workforce	Healthcare worker density and distribution - 1 (Medical doctor)
HSI2		Healthcare worker density and distribution - 2 (Dentist)
HSI3		Healthcare worker density and distribution - 3 (Pharmacist)
HSI4		Healthcare worker density and distribution – 4 (Nurses and midwives)
HSI5		Medical graduates, per 100 000 population
HSI6		Dentist graduates, per 100 000 population
HSI7		Pharmacists graduates, per 100 000 population
HSI8		Nursing and midwives graduates, per 100 000 population
HSI9	Health access	Total hospital beds per 1 000 population
HSI10	Health financing	Total current expenditure on health (% of gross domestic product)

worker density and distribution (medical doctors, dentists, pharmacists, nurses and midwives, medical graduates, dentist graduates, pharmacist graduates, nurses and midwives graduates) were considered. The health access dimension incorporated the indicator of total hospital beds, while the health financing dimension included the indicator of total current expenditure on health as a percentage of gross domestic product (Table 1).

Data Pre-processing

Standardization is a crucial data preprocessing step that involves transforming data to adhere to a standard normal distribution. In a standard normal distribution, the data has a mean of 0 and a standard deviation of 1 [15]. By transforming the data into a standard normal distribution with a mean of 0 and a standard deviation of 1, it becomes more suitable to comparison and analysis across different variables. By applying standardization, the data is rescaled to have a mean of 0 and a standard deviation of 1, facilitating easier comparison and analysis across different variables. This process eliminates the influence of varying scales and units, resulting in data that is more interpretable and suitable for certain statistical techniques and machine learning algorithms [16]. One common method for standardization is z-score standardization, also known as standard score standardization. This approach involves calculating the z-score for each data point, which indicates the number of standard deviations that the data point deviates from the mean [17]. By utilizing the z-score standardization method, the data was transformed into a standard normal distribution in this study. Matlab 8.3.0.532 (R2014a) software was employed for implementing this standardization technique.

Data Evaluation

Ten health indicator variables were subjected to correlation analysis in order to determine if there is any correlation between indicator variables. For this purpose, pearson correlation values [18] were obtained using corr command using Matlab 8.3.0.532 (R2014a). Feature selection and feature importance determination using F-test is a method commonly employed in statistical analysis and machine learning to identify the most relevant features in a dataset [19]. The F-test assesses the significance of the relationship between the target variable and each feature individually, allowing for the selection of features that are most informative for predicting the target variable. The F-test calculates the F-statistic, which is then compared to the F-distribution to determine the significance level. Features with high

F-statistic values and low p-values are considered more relevant for the model and are thus selected for further analysis or model building. On the other hand, features with low F-statistic values and high p-values may be considered less relevant and can be excluded from the analysis to simplify the model and avoid overfitting. It is important to note that the F-test assumes certain underlying assumptions, such as the normality of the data and the homogeneity of variances [20]. Violations of these assumptions can affect the reliability of the results. Therefore, it is crucial to interpret the results of the F-test and to consider the specific characteristics of the dataset and the context of the analysis.

Machine Learning Regression and Assessment

The predictive capability of machine learning models can be influenced by two important factors: data bias and data variance [21]. Data bias refers to the systematic errors or inaccuracies present in the training data used to build the machine learning models. If the training data is biased, meaning it does not accurately represent the true underlying patterns in the target variable, the models may produce predictions that are skewed or biased as well. It is crucial to address data bias and ensure that the training data is representative and unbiased to achieve reliable predictions [22]. Data variance, on the other hand, refers to the sensitivity of the machine learning models to fluctuations or noise in the training data. Models with high variance are overly complex and tend to overfit the training data, capturing noise or random fluctuations instead of the true underlying patterns. Such models may perform well on the training data but fail to generalize effectively to new, unseen data. It is essential to strike a balance between model complexity and generalizability to minimize variance and achieve accurate predictions [23].

The process of k-fold cross-validation involves dividing the available dataset into k subsets or folds of approximately equal size [24]. The model is then trained and evaluated k times, with each fold being used as the validation set once while the remaining folds are used as the training set. This ensures that every data point is used for both training and validation, reducing the potential for bias in model evaluation. During each iteration of k-fold cross validation, the model is trained on the training set and evaluated on the validation set. The evaluation metrics, such as accuracy, mean squared error, or area under the curve, are recorded for each iteration. The final performance of the model is typically obtained by averaging the evaluation metrics across all iterations. A 10-fold cross validation method was employed in this study [25].

In the context of the study, three machine learning regression methods were employed: random forest regression (RFR), neural network regression (NNR), and Gaussian process regression (GPR) [26]. Random forest regression (RFR) is an ensemble method that combines multiple decision trees to make predictions. It is known for its robustness against overfitting and ability to handle complex relationships in the data [27]. Neural network regression (NNR) is a type of machine learning model inspired by the structure and function of the human brain. It consists of interconnected nodes, or neurons, organized in layers. NNR has the capability to capture nonlinear relationships and handle large amounts of data [28]. Gaussian process regression (GPR) is a probabilistic machine learning method that models the underlying relationship between input variables and output variables. It assumes that the data follows a Gaussian process, allowing for uncertainty estimation in the predictions. GPR is particularly effective when dealing with small datasets or when uncertainty estimation is important [29].

These three machine learning regression methods were chosen for their unique strengths and capabilities in predicting COVID-19 mortality rates. The selection of these methods allows for a comprehensive evaluation of their prediction capabilities, considering different aspects such as model complexity, interpretability, and uncertainty estimation. By com-paring their performance using statistical measures such as the coefficient of determination (R2) and root mean square error (RMSE), the suitability of each method can be assessed, providing valuable insights into the predictive capabilities of machine learning models for COVID-19 mortality estimation.

To compare the performance of the models, two metrics were utilized, including the coefficient of determination (R2), root mean square error (RMSE), using the equations (1) and (2), respectively [30]:

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{n}\left(y_{obs} - y_{pre}\right)^2}{\sum_{i=1}^{n}\left(y_{obs} - \overline{y_{obs}}\right)^2}\right] \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{obs} - y_{pre}\right)^2}{n}} \tag{2}$$

where y_{obs} is the observed COVID-19 death counts per a million-population, y_{pre} is the predicted COVID-19 death counts per a million-population, $\overline{y_{obs}}$ is the average of observed COVID-19 death counts per a million-population and n is the observation number.

RESULTS AND DISCUSSION

The outbreak of the COVID-19 pandemic in late 2019 presented an unprecedented global health crisis, affecting millions of people worldwide and overwhelming healthcare systems. In response to this crisis, machine learning has emerged as a powerful tool for analysing vast amounts of data and extracting meaningful insights. By identifying complex patterns and correlations, machine learning algorithms have played a crucial role in predicting COVID-19 mortality rates and informing critical healthcare decisions.

To investigate potential correlations between ten health indicator variables, a correlation analysis was conducted. The analysis focused on computing Pearson correlation coefficients between the indicator variables (Fig. 2). Fig.2 shows that there is no correlation health indicator among variables, meaning that all health indicator variables can contribute to prediction of COVID-19 mortality rates in machine learning methodology. Subsequently, an F-test was performed to evaluate the significance of the connection between the target variable and each feature individually, facilitating the identification of the most informative features for predicting the target variable (Fig. 3). F-test results show that all ten-health indicator are important for prediction of COVID-19 mortality rates.

Machine learning techniques are well-suited for uncovering complex and non-linear relationships. COVID-19 is a multifaceted disease with interdependencies among numerous factors, and traditional statistical methods may not fully capture the intricacies of these relationships. Machine learning algorithms excel at identifying hidden

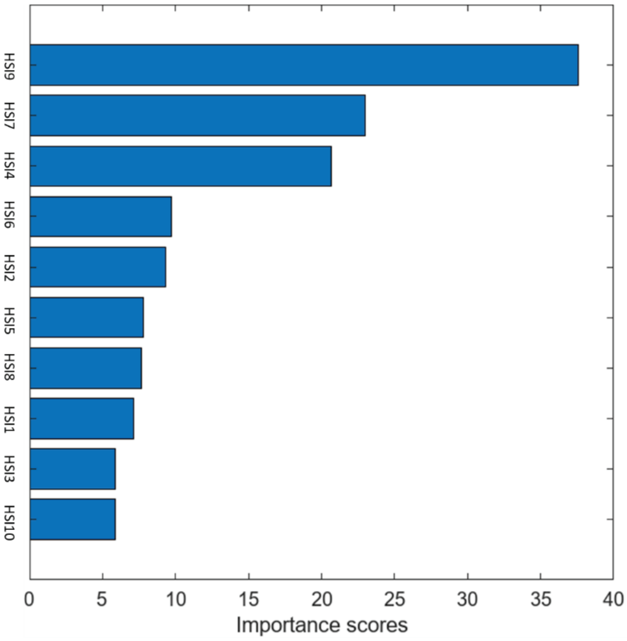

Figure 3. The importance scores of health care indicators.

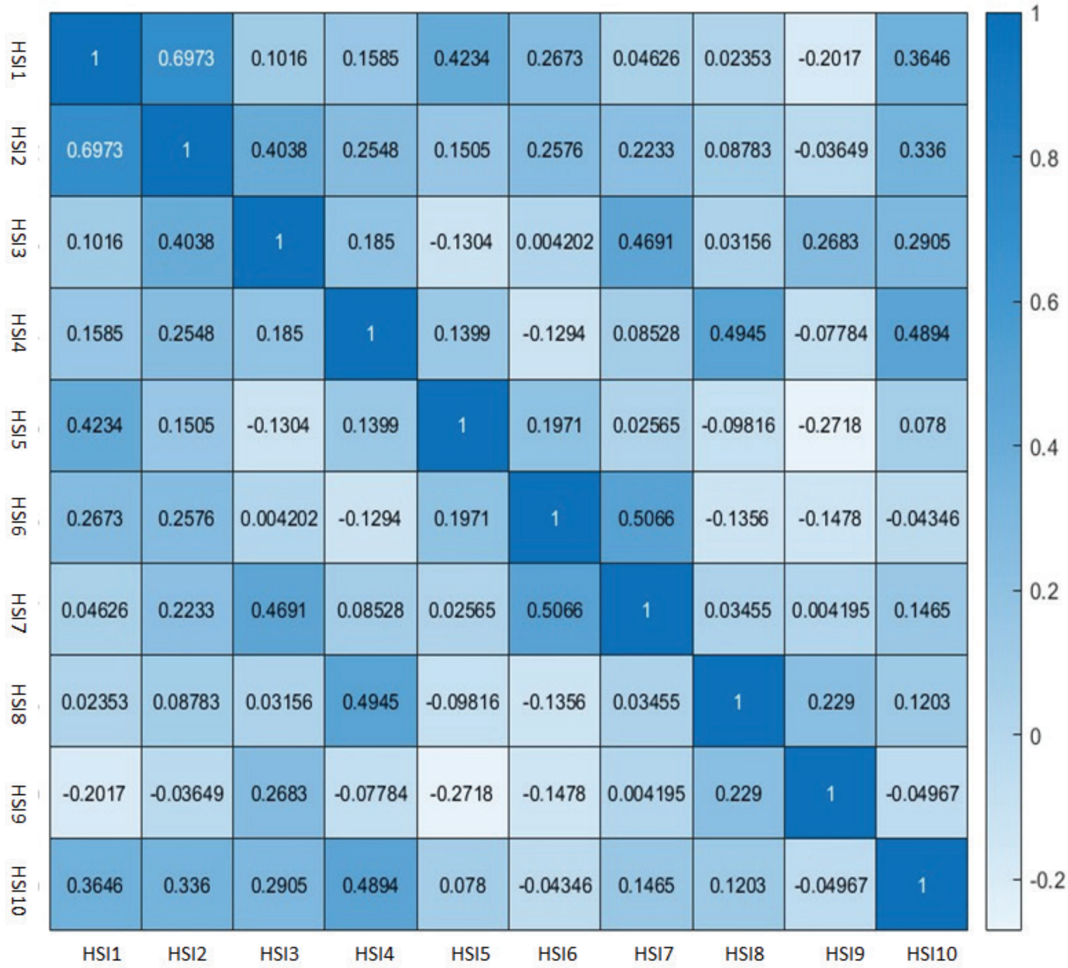

Figure 2. Correlation map of main predictor variables (the used health indicators).

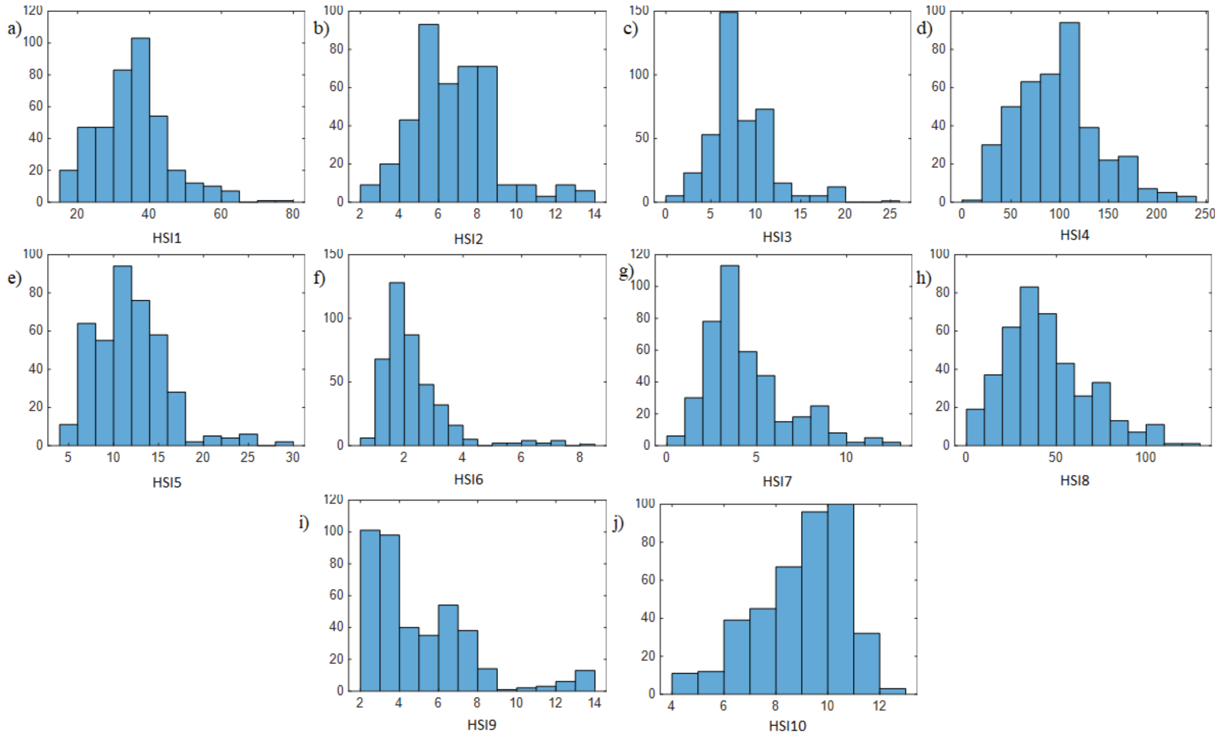

Figure 4. Histograms depicting the variables are shown for: a) HSI1, b) HSI2, c) HSI3, d) HSI4, e) HSI5, f) HSI6, g) HSI7, h) HSI8, j) HSI9, and j) HSI10.

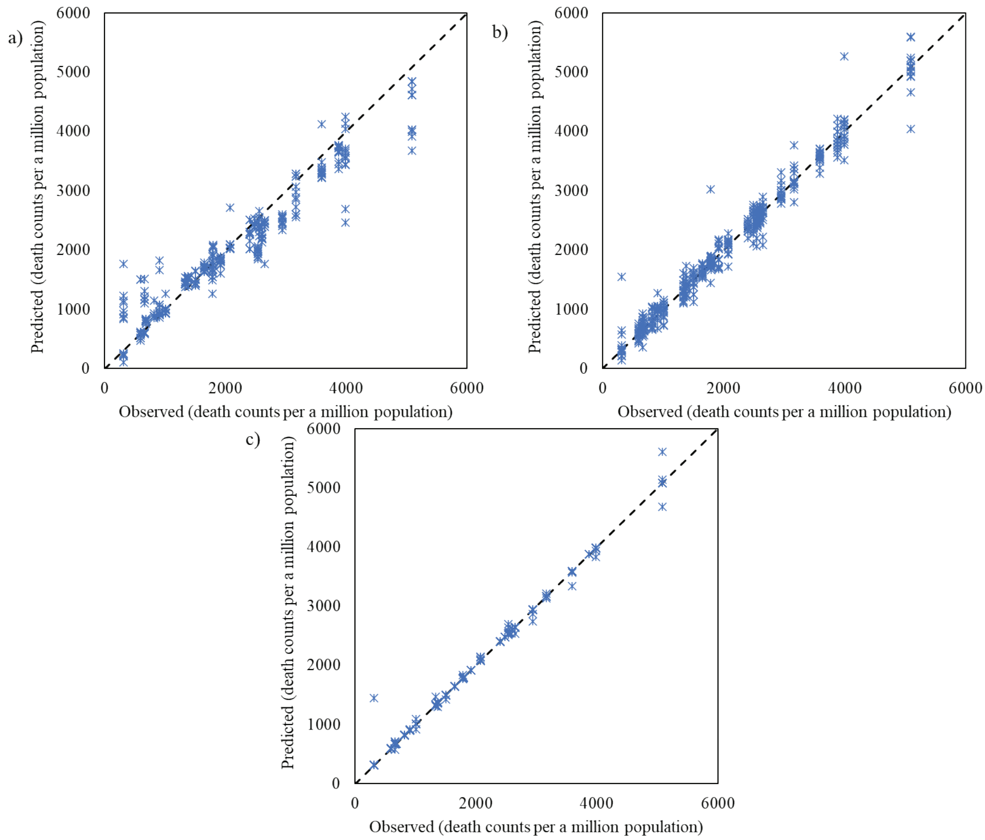

Figure 5. The observed and predicted COVID-19 mortality using a) random forest regression, b) neural network regression, and c) Gaussian process regression.

Table 2. Performance evaluation of various regression methods for validation process

Process	Training		Validation		Testing	
Regression methods	R^2	RMSE	R^2	RMSE	R^2	RMSE
Random forest regression	0.940	314	0.835	467	0.837	522
Neural network regression	0.980	138	0.841	459	0.884	441
Gaussian process regression	0.997	11	0.951	255	0.971	222

patterns and non-linear correlations that may not be immediately apparent through conventional approaches. This provides researchers and healthcare professionals with deeper insights into the complex dynamics of COVID-19 mortality, enabling more informed decision-making and targeted interventions.

The distribution of data frequency for each feature collected in the study is illustrated in Figure 4. The figure provides a visual representation of the number of occurrences or observations for each feature category. This information is valuable for conducting further analysis and drawing meaningful conclusions based on the data collected.

In addition, Supplementary Table displays the average values of each main predictor variable, along with their corresponding standard deviations (σ). These statistics provide an understanding of the central tendency and variability of the data.

The entire dataset was subjected to a random split into two subsets, with 90% allocated for training and 10% for testing. To ensure the robustness of the training process, a 10-fold cross-validation method was utilized. The outcomes of this process, which included both observed and predicted COVID-19 mortality figures, obtained through random forest regression, neural network regression, and Gaussian process regression, are presented in Figure 5. When evaluating the validation results, it becomes evident that the predictions generated by the Gaussian process regression method outperform those produced by random forest regression and neural network regression. In other words, the Gaussian process regression method exhibited a higher level of accuracy and precision in predicting COVID-19 mortality as compared to the other two regression techniques.

The results of the training, validation, and testing processes were displayed in detail in Table 2. Regarding the training results, each of the regression methods demonstrated high predictive performance. The R^2 values obtained from the machine learning-based regression methods (RFR, NNR, and GPR) ranged from 0.835 to 0.971, while the corresponding RMSE values varied from 222 to 522 (Table 2). Among the machine learning-based regression methods, it was observed that the predictions derived from the Gaussian process regression method surpassed those generated by the random forest regression and neural network regression techniques. This indicates that the Gaussian process regression method exhibited a higher degree of accuracy and precision in forecasting COVID-19 mortality compared to the other two regression methodologies.

Machine learning regression methods offer valuable tools for predicting COVID-19 mortality based on healthcare indicators. By analysing extensive datasets and leveraging advanced algorithms, these methods can uncover patterns and relationships between various variables and disease severity. The utilization of machine learning techniques in predicting COVID-19 mortality can provide crucial insights for healthcare professionals, enabling them to make informed decisions regarding patient care, resource allocation, and treatment strategies.

CONCLUSION

The Gaussian process regression method outperformed the other two regression techniques (random forest regression and neural network regression) in predicting COVID-19 mortality. These findings offer valuable insights for policymakers and healthcare providers, suggesting the importance of effective healthcare system practices during future pandemics. The study underscores the need for global, national, and local collaboration during crises, highlighting the significance of global governance. It also recommends a review of financing policies to address unexpected financial burdens and protect individuals' financial well-being in public health emergencies. Investments in both healthcare system capacity and the quality of services and healthcare workers are advised for pandemic management. Planning for the supply chain of essential medical resources should occur at global and national levels. Encouraging the development and use of digital technologies can help reduce healthcare access disparities. Community engagement and support are crucial, promoting compliance with necessary measures and fostering shared behaviours. More detailed studies on how investments in healthcare system capacity, service quality improvements, and better support for healthcare workers can be done and their direct impact on pandemic management can be modelled. Additionally, the optimisation works can be done about direct and indirect outcomes of investments in healthcare system.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Amsalem D, Dixon LB, Neria Y. The coronavirus disease 2019 (COVID-19) outbreak and mental health: current risks and recommended actions. JAMA Psychiatry 2021;78:9−10. [CrossRef]

[2] Ali SA, Baloch M, Ahmed N, Ali AA, Iqbal A. The outbreak of Coronavirus Disease 2019 (COVID-19)-An emerging global health threat. J Infect Public Health 2020;13:644−646. [CrossRef]

[3] Li Y, Xie Z, Lin W, Cai W, Wen C, Guan Y, et al. Efficacy and safety of lopinavir/ritonavir or arbidol in adult patients with mild/moderate COVID-19: an exploratory randomized controlled trial. Med 2020;1:105−113. [CrossRef]

[4] Bilal M, Nazir MS, Rasheed T, Parra-Saldivar R, Iqbal HM. Water matrices as potential source of SARS-CoV-2 transmission-an overview from environmental perspective. Case Stud Chem Environ Eng 2020;2:100023. [CrossRef]

[5] Diaz JV, Riviello ED, Papali A, Adhikari NK, Ferreira JC. Global critical care: moving forward in resource-limited settings. Ann Glob Health 2019;85:3. [CrossRef]

[6] Anantham D, Chai-Lim C, Zhou JX, Phua GC. Operationalization of critical care triage during a pandemic surge using protocolized communication and integrated supportive care. J Intensive Care 2020;8:1−9. [CrossRef]

[7] Hussain Alsayed HA, Saheb Sharif-Askari F, Saheb Sharif-Askari N, Hussain AAS, Hamid Q, Halwani R. Early administration of remdesivir to COVID-19 patients associates with higher recovery rate and lower need for ICU admission: a retrospective cohort study. PLoS One 2021;16:e0258643. [CrossRef]

[8] Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). 2023 Aug 18. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-.

[9] Nakhaeizadeh M, Chegeni M, Adhami M, Sharifi H, Gohari MA, Iranpour A, et al. Estimating the number of COVID-19 cases and ımpact of New COVID-19 variants and vaccination on the population in Kerman, Iran: A mathematical modeling study. Comput Math Methods Med 2022;2022:6624471. [CrossRef]

[10] Lin P, Hazelbaker T. Meeting the challenge of artificial intelligence: what CPAs need to know. CPA J 2019;89:48−52.

[11] Fua YH, Ward MO, Rundensteiner EA. Hierarchical parallel coordinates for exploration of large datasets. In: IEEE. 1999. p. 43−508.

[12] Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S, et al. Mastering atari, go, chess and shogi by planning with a learned model. Nature 2020;588:604−609. [CrossRef]

[13] Wei J, Chu X, Sun XY, Xu K, Deng HX, Chen J, et al. Machine learning in materials science. InfoMat 2019;1:338−358. [CrossRef]

[14] Hammer M, Hammer M. Digitization perspective: impact of digital technologies in manufacturing. In: Management approach for resource-productive operations: design of a time-based and analytics-supported methodology grounded in six sigma. Manhattan, New York City: Springer; 2019. p. 27−68. [CrossRef]

[15] Bisong E, Bisong E. Introduction to Scikit-learn. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. 2019. p. 215−229. [CrossRef]

[16] Khaledian Y, Miller BA. Selecting appropriate machine learning methods for digital soil mapping. Appl Math Model 2020;81:401−418. [CrossRef]

[17] Wang Z, Hong J, Liu P, Zhang L. Voltage fault diagnosis and prognosis of battery systems based on entropy and Z-score for electric vehicles. Appl Energy 2017;196:289−302. [CrossRef]

[18] Yücel Ö, Tarlak F. An intelligent based prediction of microbial behaviour in beef. Food Control 2023;148:109665. [CrossRef]

[19] Siraj MJ, Ahmad T, Ijtihadie RM. Analyzing ANOVA F-test and sequential feature selection for ıntrusion detection systems. Int J Adv Soft Comput Appl 2022;14:185−194. [CrossRef]

[20] Delacre M, Leys C, Mora YL, Lakens D. Taking parametric assumptions seriously: arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. Int Rev Soc Psychol 2019;32:13. [CrossRef]

[21] Mehta P, Bukov M, Wang CH, Day AG, Richardson C, Fisher CK, et al. A high-bias, low-variance introduction to machine learning for physicists. Phys Rep 2019;810:1–124. [CrossRef]

[22] Chen Y, Clayton EW, Novak LL, Anders S, Malin B. Human-centered design to address biases in artificial intelligence. J Med Internet Res 2023;25. [CrossRef]

[23] Li S, Ke L, Pratama K, Tai YW, Tang CK, Cheng KT. Cascaded deep monocular 3d human pose estimation with evolutionary training data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 6173–6183. [CrossRef]

[24] Soper DS. Greed is good: rapid hyperparameter optimization and model selection using greedy k-fold cross validation. Electronics 2021;10:1973. [CrossRef]

[25] Gunturi SK, Sarkar D. Ensemble machine learning models for the detection of energy theft. Electr Power Syst Res 2021;192:106904. [CrossRef]

[26] Choi JC, Liu Z, Lacasse S, Skurtveit E. Leak-off pressure using weakly correlated geospatial information and machine learning algorithms. Geosciences 2021;11:181. [CrossRef]

[27] Cheng Q, Chunhong Z, Qianglin L. Development and application of random forest regression soft sensor model for treating domestic wastewater in a sequencing batch reactor. Sci Rep 2023;13:9149. [CrossRef]

[28] Kampe J, Vysogorets A. Predicting Zeros of the Riemann Zeta Function Using Machine Learning: A Comparative Analysis. 2018.

[29] Manfredi P, Trinchero R. A probabilistic machine learning approach for the uncertainty quantification of electronic circuits based on Gaussian process regression. IEEE Trans Comput Aided Des Integr Circuits Syst 2021;41:2638–2651. [CrossRef]

[30] Ayoobi N, Sharifrazi D, Alizadehsani R, Shoeibi A, Gorriz JM, Moosaei H, et al. Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. Results Phys 2021;27:104495. [CrossRef]