Journal of Intelligent Systems and Internet of Things
Vol. 14, No. 02, PP. 165-177, 2025
DOI: https://doi.org/10.54216/JISIoT.140214

Forecasting for Vaccinated COVID-19 Cases using Supervised Machine Learning in Healthcare Sector

Ali Khraisat[1,*], Mohd Khanapi Abd Ghani[1]

[1]Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Department of Software Engineering, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia

Emails: p032010018@student.utem.edu.my ; khanapi@utem.edu.my

Abstract

Machine learning (ML)-based forecasting techniques have demonstrated significant value in predicting postoperative outcomes, aiding in improved decision-making for future tasks. ML algorithms have already been applied in various fields where identifying and ranking risk variables are essential. To address forecasting challenges, a wide range of predictive techniques is commonly employed. Research indicates that ML-based models can accurately predict the impact of COVID-19 on Jordan's healthcare system, a concern now recognized as a potential global health threat. Specifically, to determine COVID-19 risk classifications, this study utilized three widely adopted forecasting models: support vector machine (SVM), least absolute shrinkage and selection operator (LASSO), and linear regression (LR). The findings reveal that applying these techniques in the current COVID-19 outbreak scenario is a viable approach. Results indicate that LR outperforms all other models tested in accurately forecasting death rates, recovery rates, and newly reported cases, with LASSO following closely. However, based on the available data, SVM exhibits lower performance across all predictive scenarios.

Received: March 26, 2024 Revised: July 09, 2024 Accepted: October 25, 2024

Keywords: COVID-19; Supervised machine learning; Future forecasting; Least absolute shrinkage and selection operator; Support vector machine

1. Introduction

Machine learning (ML) has made remarkable progress in addressing complex, real-world challenges across diverse fields over the last decade, including gaming, climate modelling, autonomous vehicles, healthcare, natural language processing, robotics, and more [1]. Unlike traditional algorithms that rely heavily on if-else statements, ML algorithms often employ a trial-and-error approach, refining their performance based on data patterns and outcomes. Forecasting and prediction have emerged as pivotal applications of ML, where a variety of algorithms provide guidance for future actions across domains such as stock market analysis, disease prognosis, and weather prediction [2]. Studies have demonstrated the efficacy of ML methods like regression and neural network models in predicting outcomes for several diseases, including coronary artery disease, breast cancer, and cardiovascular disease [3]. These advancements underscore ML's transformative potential for disease prognosis and prediction, ultimately driving improvements across the broader healthcare sector.

The new coronavirus, SARS-CoV-2, has been formally dubbed COVID-19 by the World Health Organization (WHO). The objective of the project is to develop an early forecast model for the spread of the virus [4]. First identified in late 2019 in Wuhan, China, COVID-19 poses a major global danger to human life. There are similarities between its symptoms and those of pneumonia. Among the many symptoms the virus produces in humans are multiple-organ failure and serious acute respiratory syndrome, both of which have a rapid death rate. The epidemic has affected hundreds of thousands of people worldwide and is causing thousands of deaths every day. The virus mostly spreads by respiratory droplets, contact with infected surfaces, and intimate physical contact between people. The virus has challenges in spreading when persons are in the asymptomatic carrier condition, where they may have the infection for an extended period of time without showing any symptoms. Because of this, several countries have implemented complete or partial lockdowns in the affected cities and areas. Global researchers are putting a lot of effort into developing a vaccine and cure for the disease. Governments are highlighting the need of taking preventive action because there is currently no medication that can totally eliminate the virus. Understanding everything there is to know about COVID-19 is considered crucial among the several safety precautions. Numerous scholars are contributing to this data by investigating different facets of the epidemic and developing ways to assist individuals.

Coronaviruses have been extensively studied since they were first identified more than half a century ago. In 2004, Tobinick conducted research on the impact of H5N1 influenza and SARS coronavirus on severe immune-based lung injury. According to the study, using biologic Tumour Necrosis Factor (TNF) inhibitors, such etanercept, may be a more effective and focused approach than corticosteroids to lessen the severe alveolar damage caused by infection [5]. The word "corona," which virologists used to refer to characteristic projections outside the virus, was modelled after the solar corona. Coronaviruses are the cause of Severe Acute Respiratory Syndrome (SARS), which is often fatal. Due to their vulnerability to mutation, single-stranded RNA viruses are extremely varied and have a width of around 120 nm. In 1968, the first coronaviruses identified in humans, 229E and OC43, mostly infected non-human and human birds and animals. The Middle Eastern respiratory syndrome (MERS) epidemic in 2012, the severe acute respiratory syndrome coronavirus 1 (SARS-COV-1) in 2003, and the SARS coronavirus 2 (SARS-COV-2 or COVID-19) in 2019 all led to serious human illnesses [6]. COVID-19 has been spreading quickly over the world since it was discovered in late [7]. It is believed that the virus that causes COVID-19 (SARS-COV-2) first infected bats before infecting people by contaminating meat that was sold in China's meat markets with the excrement of wild animals. Spike glycoproteins, which are required for the viruses to enter host cells, are the cause of coronavirus syndrome. S1 and S2, the spike's two subunits, attach to a receptor on the host cell's surface and fuse with the cell membrane, respectively.

Located in the centre of the Middle East, the Hashemite Kingdom of Jordan has been negotiating the complex obstacles presented by the worldwide COVID-19 epidemic [8]. The virus's arrival caused unheard-of disruptions to daily life, which had a big impact on social structures, economics, and public health. In this regard, the Jordanian healthcare system became a vital frontline organization, charged with the enormous responsibility of adapting to the changing pandemic dynamics. The healthcare system has changed significantly since the first COVID-19 case was discovered in Jordan in order to control and lessen the virus's effects. Jordan launched immunization operations to vaccinate its citizens against the virus as part of the international response. These initiatives sought to reduce the burden on healthcare systems, stop the spread of COVID-19, and eventually prepare the path for things to return to normal [9]. A significant portion of Jordan's population has received at least one dose of the COVID-19 vaccine thanks to the country's successful immunization efforts. However, it is becoming more and more important to understand and predict the course of vaccinated COVID-19 cases because to the virus's dynamic nature and the appearance of new variations.

Jordan has responded to the COVID-19 pandemic in a variety of ways, including vaccine programs, extensive testing, contact tracing, and strict public health measures. Recognizing the importance of immunizations in containing the virus, the country launched a massive vaccination campaign to safeguard its inhabitants and nationals [9]. A sizeable percentage of the population has sought immunization as a result of the distribution of several vaccine formulations and public awareness initiatives. Despite these initiatives, problems still exist. A planned and flexible reaction is required due to the virus's unpredictable nature and the appearance of new strains. Tools that can predict and evaluate models associated to vaccinated COVID-19 patients are desperately needed as Jordan's healthcare system works to control the continuing epidemic. These tools can provide insights that can guide proactive decision-making. Technology has been essential in enhancing healthcare capacities among the pandemic's obstacles. Specifically, a type of artificial intelligence called machine learning has shown promise in transforming a number of healthcare domains, including diagnosis, treatment planning, and resource efficiency. In the context of predicting and verifying COVID-19 instances, supervised machine learning, in which models are trained on labelled datasets to generate predictions without explicit programming, shows potential.

Machine learning's flexibility and data-driven approach make it especially applicable in the rapidly changing pandemic environment [10]. Healthcare practitioners can learn more about the possible paths of vaccinated COVID-19 patients by using algorithms to examine complicated records. A more dynamic and responsive approach to public health issues is made possible by this analytical method, which is based on machine learning principles and provides a supplementary strategy to conventional epidemiological methodologies. The capacity to predict the trajectory of vaccinated COVID-19 patients becomes critical in the quest for an effective response to the continuing epidemic. Healthcare authorities may plan resource allocation, optimize immunization campaigns, and proactively respond to any surges in cases with the use of forecasting models customized for the Jordanian setting. Additionally, these models' correctness and dependability are guaranteed by validation, which serves as a basis for evidence-based decision-making [11]. Machine learning's incorporation into the healthcare industry's prediction toolkit makes sense as Jordan continues to struggle with the pandemic's effects. In the particular context of the Jordanian healthcare system, this study aims to investigate the use of supervised machine learning in forecasting and verifying models for vaccinated COVID-19 cases. The main contributions of this study are:

- This study applies three widely adopted machine learning models are Support Vector Machine (SVM), Least Absolute Shrinkage and Selection Operator (LASSO), and Linear Regression (LR) to classify COVID-19 risk, contributing valuable insights into predictive techniques that can support healthcare decision-making for pandemic response.
- This study provide identification of optimal predictive model through comparative analysis, the study identifies Linear Regression as the most effective model for accurately predicting key COVID-19 metrics, such as death rates, recovery rates, and new case numbers. This finding helps prioritize predictive models based on performance, enhancing the accuracy and reliability of health outcome forecasting.
- Our study show impact assessment of covid-19 on Jordan’s healthcare system by focusing on Jordan’s healthcare system, this research provides localized insights into the pandemic’s impact, which can serve as a basis for health policy improvements and as a model for other regions aiming to forecast and mitigate COVID-19-related healthcare challenges.

2. Literature Review

A critical lens through which we might examine the complex dynamics of the pandemic in the Hashemite Kingdom is the epidemiological backdrop of COVID-19 in Jordan. Jordan has been actively involved in negotiating the difficulties presented by the new coronavirus since the beginning of the worldwide health catastrophe. Gaining insight into Jordan's experience with the pandemic requires an understanding of the country's transmission patterns, public health effects, and response strategy efficacy. This investigation dives into the particular elements affecting the epidemiological landscape, providing a detailed viewpoint on how public health, disease dynamics, and the socioeconomic fabric interplay in Jordan's particular setting.

2.1 Overview of the COVID-19 pandemic in Jordan.

The Hashemite Kingdom of Jordan, which is tucked away in the center Middle East, encountered hitherto unheard-of difficulties when the COVID-19 epidemic struck. With 250,000 confirmed cases reported nationwide as of January 15, 2024 [12], there is a chance to examine the virus's many impacts on public health, the healthcare system, and the country's larger socioeconomic structure.

Jordan's experience with the COVID-19 pandemic started as they became more aware of the new coronavirus's threat to the entire world. The Jordanian government was quick to identify the possible risks and took a number of aggressive steps to stop the virus from spreading inside its boundaries. These precautions included harsh curfews, travel restrictions, and lockdowns during the early stages of the pandemic [13].

In addition to safeguarding public health, the initial reaction sought to make sure the healthcare system would be resilient to future spikes in COVID-19 cases. In order to reduce the burden on healthcare resources and buy time to strengthen its healthcare infrastructure, Jordan aimed to flatten the curve by putting these measures into place early on. A number of variables have combined to influence the dynamics of COVID-19 transmission in Jordan, reflecting both the country's distinct features and worldwide trends. The danger of community spread was higher in urban areas because of their denser populations and more social connections. Furthermore, a key factor in the virus's spread was human migration both within and across regions. It has become essential to comprehend the mechanisms of illness transmission in order to direct focused therapies. Strategic lockdowns in regions with higher incidence rates, contact tracking, and extensive testing have all been used in an effort to reduce the risk of transmission. The government's dedication to making decisions based on data has been crucial in modifying response plans in light of the pandemic's changing terrain. The COVID-19 pandemic has put tremendous strain on Jordan's healthcare system. Despite their best efforts to care for COVID-19 patients, hospitals and other healthcare facilities have encountered difficulties with staffing, allocating resources, and maintaining standard medical services. The dynamic nature of the pandemic has necessitated adaptive strategies to manage the increased demand for healthcare services" [14].

The knowledge gained from the public health and governmental reactions to the COVID-19 pandemic serves as a basis for future readiness. It emphasizes the significance of flexible tactics, evidence-based judgment, and a concerted global reaction. Continuous cooperation, investments in healthcare infrastructure, and a dedication to guaranteeing fair access to resources, such as vaccinations, are necessary to address the global health issues brought on by pandemics .

2.2 Vaccination Strategies and Campaigns in Jordan

Breathlessness has been observed by many patients, and it often appears to be a flu symptom. Other significant symptoms of COVID-19 include fever (98%), cough (76%), and diarrhea (3%), which are frequently more severe in older persons with chronic conditions. The global epidemic was declared as an Emergency in Public Health of global significance and is contracted by direct contact with an infected person's bodily fluids, like sneezing or blowing. Despite a significant global effort by researchers and experts to develop a treatment and an authorized vaccine, COVID-19 does not yet have either [15]. However, there are now antiviral medications being utilized in COVID-19 clinical studies, such as [16]. and the COVID-19 pandemic is being managed with medicinal herbs believed to have preventive properties on respiratory tract disorders.

In addition, a lack of testing tools and postponed or incorrect diagnoses from cases with no symptoms risk individuals, guests, and medical staff to the the 2019-n Co virus. This poses a major danger for the medical and business sectors. It has grown clear that nonclinical techniques like data exploration, machine learning, trained systems, and various other AI technologies must be applied in order to identify and manage the COVID-19 epidemic. Non-therapeutic approaches offer the most reliable and accurate diagnosis techniques for 2019-nCoV while also alleviating a significant load on health care systems. The application of machine learning (ML), one of the increasingly advanced concepts in artificial intelligence (AI), provides a systematic approach to developing automated, complex, and objective algorithmic techniques for analyzing data related to biology or heterogeneous and multivariate arithmetic [17]. The machine learning techniques may retrieve and alter their framework according to visible information by maximizing around a cost function or an objective.

In order to detect flaws and adjust the model based on the outcomes, the methods can supply input data results with a suitable training procedure and examine outcomes with predictions and real-life results.

When the training dataset is neither labeled or categorized, machine learning is employed as an unsupervised learning approach. The techniques for learning derive an equation from the data set with no labels in order to find hidden information or patterns.. In order to uncover hidden patterns in the unlabeled dataset, the strategy eliminates observations rather than identifying the correct output.

Methods to learning which fall halfway among directed and unstructured are known as semi-supervised methods. Unlabeled samples are used for training, whereas approaches to unsupervised learning are identified. Learning methods often take into account a larger unlabeled dataset and a smaller labeled dataset. When learning or acquiring knowledge with a labeled dataset calls for appropriate and competent assets, the learning procedures are preferred and may be modified to get more accuracy.

In order to identify mistakes, reinforcement learning strategies engage with the learning environment. In order to improve the model's performance, reinforcement learning approaches often use trial-and-error searches and delayed incentives. These strategies are used to determine the optimal behavior in a given scenario.

Following that, real-time PCR (RT-PCR) is regarded as a diagnostic gold standard. In order to diagnose COVID-19, the time frame and kind of specimen obtained for RT-PCR are crucial. Serum was negative for the virus in the early stages, while respiratory models were positive. Additionally, rather than in the latter stages of viral infection, it has been proposed that patients have elevated virus levels in the early stages [18]. It is crucial to collect and evaluate specimens quickly, and a laboratory professional should provide guidance. For COVID-19, there are now seven fast RT-PCR diagnostic kits on the market. Only one of these kits, from the Beijing Genome Institute (BGI) in China, is authorized for clinical diagnosis; the other six are for research. Furthermore, two kits (BGI, China, and Veredus, Singapore) may use sequencing and microarray technologies, respectively, to identify COVID-19 and another pathogen.

In order to comprehend how the early ML models were put up and the extent of their validity, the unsupervised ML was included in the table. This kind of MLS is often used to categorize and analyze data in order to execute a model, not for prediction. The remaining papers in the table are owned by supervised multilingual search engines. This method's discrimination was said to be an effective one. In order to aid in the prediction process, [19] from India used logistic regression and naïve Bayes to categorize data up to 96% accuracy.

In the healthcare industry, predictive modeling is a game-changing strategy that uses data and sophisticated analytics to predict and control a range of patient health issues. In the domain of infectious illnesses, where the capacity to predict disease outcomes and patterns has significant public health ramifications, this trend has been especially noticeable. Predictive algorithms to forecast the spread of COVID-19 have been actively developed in recent years by researchers like [20]. A wide range of fields, including early warning systems, tracking and prediction systems, data dashboards, diagnostics, prognostics, treatments, cures, and social control tactics, are incorporating artificial intelligence (AI) into the pandemic response. The lack of data has been a recurring issue, nevertheless, preventing the development of reliable early prediction models in spite of the encouraging uses of AI in this field.

Medical experts and governmental officials were quite supportive of the early efforts to model and stop the development of COVID-19. The discussion was broadened by [21], who talked about combining cloud computing capabilities with improved mathematical modeling and machine learning. Based on the analytical questions that were part of their investigation. In order to get the best results, their study highlighted how crucial it is to match predictive algorithms with the particular subtleties of the analytical issues. In summary, the COVID-19 pandemic response is a prime example of the complex interactions of socioeconomic factors, public health initiatives, and epidemiological models. While admitting the limits and uncertainties inherent in predicting during a fast shifting crisis, the use of predictive models highlights the need of data-driven decision-making. The lessons learnt from the COVID-19 pandemic highlight the necessity of a comprehensive and flexible strategy, combining scientific knowledge with socioeconomic factors to create robust public health institutions, as the globe navigates future health emergencies.

3. Methodology

COVID-19 predictions, another name for the new coronavirus, are the subject of this investigation. The COVID-19 virus has shown itself to be a current hazard to people life. It claims the lives of millions of people worldwide, and the fatality rate is increasing every day. With the goal to assist manage the risk of a pandemic, this research attempts to predict the death toll, the amount of each day proven patients with the virus, and the amount of recuperation instances within the Jordanian healthcare system during the following 10 days. Three machine learning techniques that are suitable for this situation were used to complete the predicting. The total amount of cases confirmed, deaths, and cures during the previous amount of days before the pandemic started are shown in the hourly time series summary tables in the study's dataset. To get the worldwide statistics of the daily number of fatalities, confirmed cases, and recoveries, the dataset was first preprocessed for this study.

Following the first stage of data preparation, the dataset was split into two subsets: a testing set (10 days) and a training set (56 days) for model training. In this work, learning models including SVM, LR, and LASSO were employed. The days and fresh confirmed cases, recovery, and death trends have been used to train these models. The results were then given after the learning models were evaluated using key indicators such as MSE, RMSE, and MAE. A simplified schematic of the proposed technique used in the study is displayed in Figure 1.

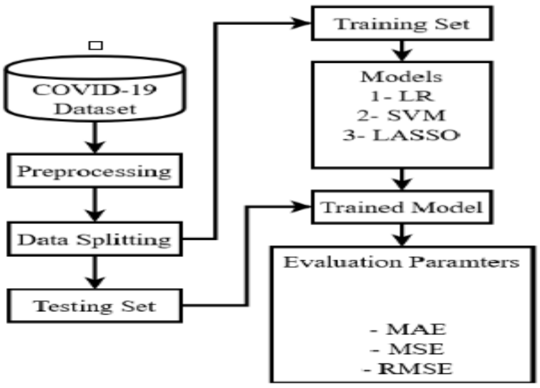

Figure 1. The proposed model

3.1 supervised machine learning

A supervised algorithm for learning is built to produce a prediction given an unidentified input example. To learn the model of regression using this method of learning, the learning algorithm uses a dataset including input instances and their matching regressor. A projection on the experimental dataset or unanticipated input data is then generated by the trained model. Methods for classification and regression techniques can be applied in this learning approach to construct predictive models. In this research of COVID-19 future forecasting, three regression models Linear Regression, LASSO Regression, and Support Vector Machine have been applied.

i) Linear Regression

Understanding and predicting results in the fields of machine learning and predictive modeling frequently depend on the connection between a target class and independent [22]. A popular statistical method known for its interpretability and efficacy in predictive analysis within machine learning, linear regression is a key tool for investigating this connection.

Finding the The variable that is dependent (y) and the variable that is not dependent (x) have a relationship that is linear. is the goal of linear regression. Since linear regression aims to measure the effect of changes in the independent variable on the dependent variable, every observation in this model depends on these two values. Equations (1) and (2), which capture the core of this connection, reflect the model:

$$y = \beta_0 + \beta_{1x} + \varepsilon \tag{1}$$

or equivalently

$$E(y) = \beta_0 + \beta_{1x} \tag{2}$$

The error term is shown by ε, the coefficient for the independent variable 'x' is represented by β_1, and the intercept or constant term that has to be determined is represented by β_0. The equations demonstrate how the linear regression model seeks to choose the best coefficients in order to reduce the total prediction error.

Adding to the mathematical underpinnings, further equations clarify certain facets of linear regression. Equation (2) emphasizes the idea of the expected value (E(y)), which is the mean or average value of the dependent variable 'y' for a given 'x'. An essential metric for comprehending the primary tendency of anticipated results is this expected value.

The expression for the slope (β_1) in the context of linear regression is shown as follows to further enhance the conversation in Equations (3) and (4):

$$\text{minimize } \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2 \tag{3}$$

$$g = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2 \tag{4}$$

A cost function, denoted by g, is the root mean square of the actual value of y (y_i) and the projected value of y ($pred_i$), where n is the total number of data points.

ii) LASSO

With its distinctive use of shrinkage, the Least Absolute Shrinkage and Selection Operator (LASSO) is a well-known regression model in the larger linear regression approach [23]. In this application, shrinkage refers to bringing a data sample's extreme values closer to its central values in order to improve the stability of the model and lower errors [24].

The uniqueness of LASSO is demonstrated by its skillful management of multi-collinearity scenarios, a frequent problem in regression modeling. The penalty is equivalent to the size of the coefficients when using L1 regularization. By automatically punishing superfluous characteristics, this regularization strategy streamlines the regression process. It is possible to efficiently minimize possibly to zero features that have a negligible impact on the regression findings. Every variable that is available is used in a conventional multivariate regression, and each one is given a regression coefficient. By introducing features gradually, LASSO takes a more judicious approach. A new feature is essentially set to zero if it does not considerably enhance the fit when accounting for the penalty term. This special feature of LASSO emphasizes the inclusion of characteristics that really improve predicted accuracy, leading to a more condensed model.

In terms of mathematics, LASSO adds a regularization component to the objective function of conventional linear regression. The following is the expression for the LASSO optimization problem in Equations (5)

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \tag{5}$$

When λ|slope| is a penalty term, it determines the coefficient, which may be understood as min (sum of square residuals + λ|slope|).

iii) Support Vector Machine (SVM)

A supervised machine learning approach for regression and classification is called (SVM) [25]. SVM analysis is a non-parametric technique that depends on a number of mathematical processes. The kernel is a group of components that changes the incoming data into the required format. When dealing with non-linear regression issues, SVM transfers the input vector (x) to an n-dimensional space known as a feature space (z). SVM uses a linear technique to handle regression issues. Irregular mapping strategies are utilized to conclude this mapping once linear regression has been applied to space. Utilizing (y_n) as a collection of reported reactions and a multiple-variate training collection (x_n) with N occurrences, the idea is placed in an ML framework. Equation (6) illustrates the linear function as follows:

$$f(x) = x'\beta + b \quad (6)$$

3.2 Dataset

Predicting the future spread of COVID-19 is the main goal of this study, with particular attention to estimating the number of new positive cases, fatalities, and recoveries. A carefully selected dataset has been obtained from the GitHub repository linked to the Epidemiology Centre dataset files in order to do this. According to [1], the dataset is a thorough compilation of daily time series summary tables that includes important variables including the number of confirmed cases, fatalities, and recoveries.

As a trustworthy source of epidemiological data, this repository guarantees that the study's basis is based on current and correct data. A thorough examination of trends and patterns over time is made possible by the dataset's time series format, which offers a granular picture of the changing COVID-19 scenario. Notably, every piece of data in this repository comes from daily case reports, guaranteeing the dataset's timeliness and usefulness in capturing the pandemic's present situation. One of the most important aspects of dataset integrity is the frequency of data changes. The dataset in this study is updated every day to reflect the dynamic nature of the COVID-19 scenario. The dedication to regular updates guarantees that the models created in this research are trained and assessed using the most up-to-date and precise data available. In a rapidly changing pandemic, where prompt insights can guide preemptive public health interventions, this temporal accuracy is especially important [25].

4. Results & Discussion

4.1 Evaluation Parameters

The indicators that follow are used in this research to evaluate each learning model's performance: R2-squared (R2) score, adjusted R-squared score (R2adjusted), the mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE).

i) R-squared Score

Regression model performance is assessed statistically using the R-squared (R2) score [26][27]. The statistic displays the percentage of variance in the dependent variable that determines the independent variable overall. Using a handy 0–100% scale, it assesses the degree of correlation between the dependent variable and regression models. The R2score may be used to assess the goodness-of-fit of trained models after the regression model has been trained. The R2score, also known as the coefficient of determination, measures how dispersed the data points are around the regression line. Its score ranges from 0% to 100% at all times. A score of 0% indicates that the response variable's variability around its mean is not explained by the model, while a score of 100% indicates that the variability around its mean is fully explained. The trained model's quality is demonstrated by its strong R2 score. The proportion of variance in the independent variable is explained by the R2 linear model. It is located as Equations (7):

$$R^2 = \frac{Varianceexplainedbymodel}{Totalvariance} \quad (7)$$

ii) Adjusted R-squared Score

A modified version of R2, the Adjusted R-squared (R2adjusted) additionally indicates how well the data points fit the curve. R2 and R2adjusted vary primarily in that the latter accounts for the quantity of features in a prediction model. If the newly additional characteristics are beneficial to the prediction model, the addition of more features may result in a rise in R2adjusted. Its worth will drop, nevertheless, if the recently added features are pointless. One definition of the R2adjusted is in Equations (8)

$$R^2_{adjusted} = 1 - (1 - R^2)\frac{n-1}{n-(k+1)} \quad (8):$$

In this case, k is the number of independent variables in the regression equation, and n is the sample size.

iii) Mean Absolute Error (MAE)

The mean absolute error is the average magnitude of the errors in the set of model forecasts. [28]. This is an average of the test data, where each unique difference is given equal weight, between the model predictions and the actual data. It is also known as negatively-oriented scores since its matrix value ranges from 0 to infinity, with lower score values indicating the quality of learning models [29]-[31] as Equations (9).

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_i - \hat{y}_j| \tag{9}$$

iv) Mean Square Error (MSE)

Regression model performance may also be assessed using mean square error [28]. MSE squares the data points' separation from the regression line. Because it eliminates the value's negative sign and provides greater weight to significant discrepancies, squaring is required. The closer you are to locating the line of best fit, the less the mean squared error. MSE can be computed as Equations (10):

$$MSE = \frac{1}{n}\sum_{j=1}^{n} |y_i - \hat{y}_j|^2 \tag{10}$$

v) Root Mean Square Error (RMSE)

The standard deviation of the prediction mistakes is known as the root mean square error. Prediction errors, often referred to as residuals, are the separation between the actual data points and the best fit line. As a result, RMSE quantifies the degree of concentration of the actual data points around the optimal fit line. This is the error rate as determined by the square root of the mean square error (MSE)[32]-[34] as Equations (11).

$$RMSE = \sqrt{\frac{1}{n}\sum_{I=1}^{n} |y_i - \hat{y}_i|^2} \tag{11}$$

4.2 Death Rate Future Forecasting

The study makes predictions about the death rate, and the findings show that LR outperforms all other models, with LASSO having a good R2score. SVM, on the other hand, does the worst in this scenario. Table 1 presents the findings.

Table1 : Models' performance in predicting the mortality rate in the future.

Model	R^2 Score	R^2 Adjusted	MSE	MAE	RMSE
LR	0.94	0.93	851242.25	701.23	901.36
LASSO	0.84	0.81	3345176.81	1242.25	1532.1
SVM	0.54	0.38	15125311.56	3028.21	3875.54

The performance of the LR, LASSO, and SVM models is displayed as graphs in Figures 2, 3, and 4, respectively. All of the numbers' graphs indicate that the death rate is expected to rise in the days ahead, which is quite concerning. Plotting the current mortality rate in Figures 2,3, and 4 graph demonstrates that the models' projections were accurate.

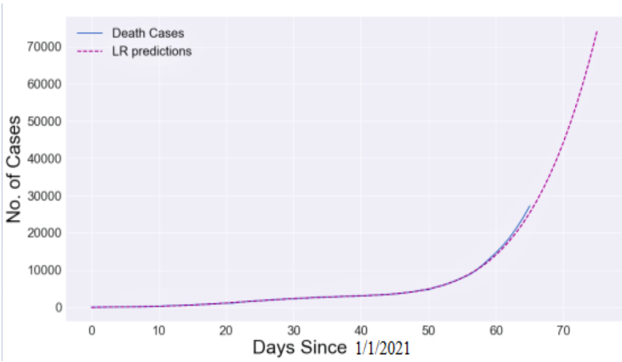

Figure 2. LR's death forecast over the next ten days.

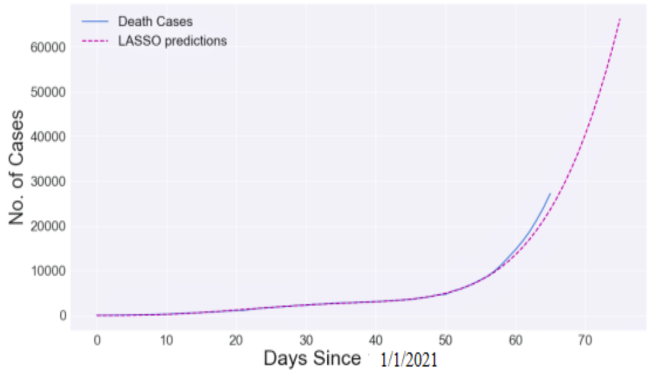

Figure 3. LASSO's death forecast over the next ten days.

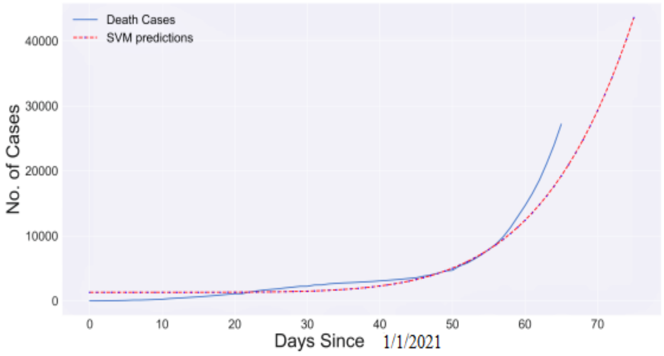

Figure 4. SVM-based death prediction for the next ten days.

4.3 New Infected Confirmed Cases Future Forecasting

The predicting outcomes of the models employed in this investigation are displayed in Table 2. SVM performs quite poorly across all assessment measures, while LASSO leads the table in terms of performance and LR does well as well. Figures 5, 6, and 7 display the learning models' predictions as graphs.

Table 2: Performance of models for future predictions of newly confirmed infected patients.

Model	R^2 Score	R^2 Adjusted	MSE	MAE	RMSE
LR	0.81	0.78	1481856504.96	31145.55	37281.51
LASSO	0.97	0.96	244538560.99	11658.97	14521.11
SVM	0.57	0.48	5851882869.30	60165.90	76821.28

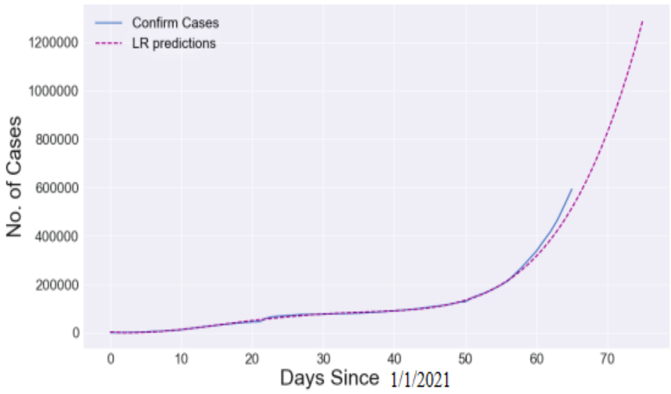

Figure 5. LR's projection of new infected confirm cases during the next ten days

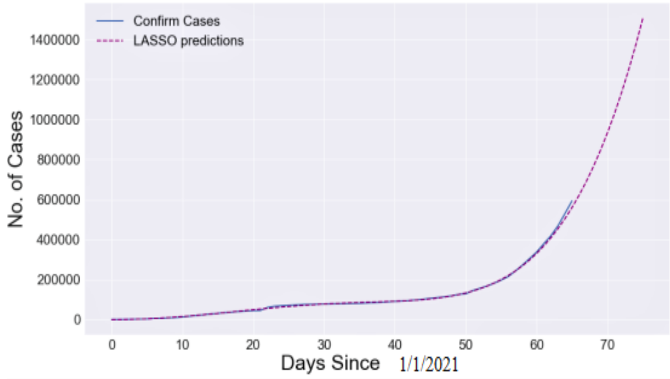

Figure 6. LASSO's projection of new infected confirm cases for the next ten days

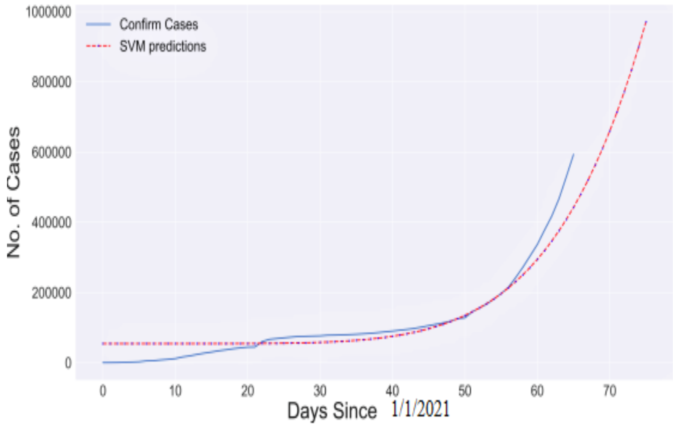

Figure 7. SVM predicts new infected confirm cases within the next ten days

4.4 Recovery Rate Future Forecasting

Once more, the ES outperforms all other models in recovery rate future projection. Because of the nature of the time-series data that is available, ES performs the best, followed by LR, LASSO, and SVM. All other models perform badly. Figures 8, 9, and 10 display the patterns of the predictions for the next several days. Table 3 below displays the learning models' performance outcomes:

Table 3: Models' performance in predicting the recovery rate in the future.

Model	R^2 Score	$R^2_{Adjusted}$	MSE	MAE	RMSE
LR	0.38	0.20	481011814.51	16926.08	221219.95
LASSO	0.28	0.07	1553244344.82	31215.27	37137.99
SVM	0.23	0.02	13212148615.72	107629.82	109847.58

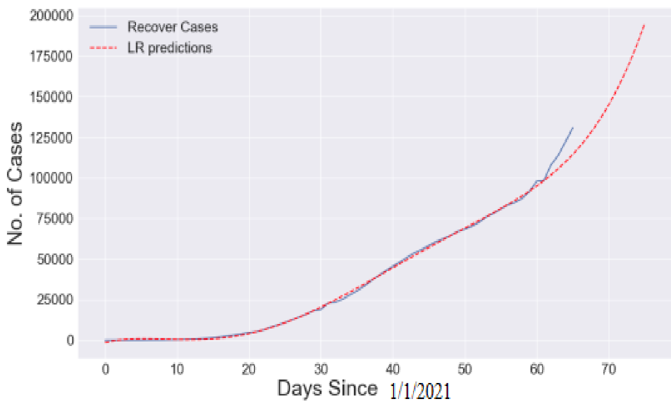

Figure 8. Recovery rate prediction by LR for the upcoming 10 days.

Figure 9. Shows the LASSO recovery rate projection for the next 10 days.

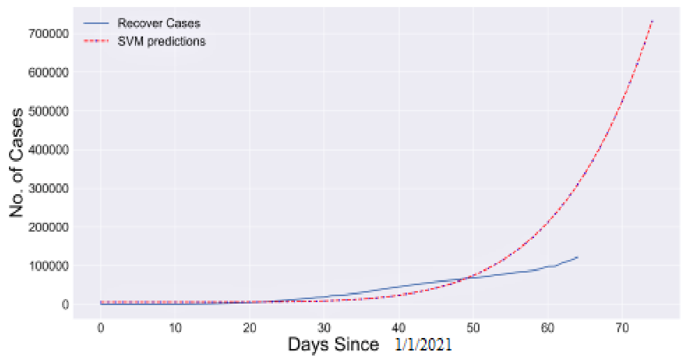

Figure 10. SVM-predicted recovery rate over the next ten days

5. Conclusion

A major worldwide catastrophe was sparked by the COVID-19 pandemic's precariousness. A significant section of the Jordanian health system was impacted by the epidemic. An ML-based prediction method has been suggested in this work to forecast the risk of COVID-19 outbreaks for health sector personnel. The technology uses machine learning algorithms to forecast future days after analyzing a dataset that includes day-by-day real historical data. Considering the size and composition of the dataset, the study's conclusions show that LR excels in the current forecasting field. To a certain degree, LR and LASSO are also effective in forecasting in order to confirm cases and predict the death rate. The results of these two models suggest that in the days ahead, the death rate will increase and the rate of recovery would slow down. Because the numbers in the dataset fluctuate, SVM consistently yields subpar results. It was somewhat difficult to draw an exact hyperplane between the numbers in the dataset. All things considered, we conclude that the model's predictions given the current situation are correct, which might help us comprehend what lies ahead. As a result, the study's predictions can also greatly aid authorities in making judgments and acting promptly to limit the COVID-19 pandemic.

Funding: “This research received no external funding”
Conflicts of Interest: “The authors declare no conflict of interest”.

References

[1] Salman, A. O., & Geman, O. (2023). Evaluating Three Machine Learning Classification Methods for Effective COVID-19 Diagnosis. International Journal of Mathematics, Statistics, and Computer Science, 1, 1–14. https://doi.org/10.59543/ijmscs.v1i.7693

[2] Arif, Z. H., & Cengiz, K. (2023). Severity Classification for COVID-19 Infections based on Lasso-Logistic Regression Model. International Journal of Mathematics, Statistics, and Computer Science, 1, 25–32. https://doi.org/10.59543/ijmscs.v1i.7715

[3] S. Y. Siddiqui et al., "Modelling, simulation, and optimization of diagnosis cardiovascular disease using computational intelligence approaches," Journal of Medical Imaging and Health Informatics, vol. 10, no. 5, pp. 1005-1022, 2020.

[4] World Health Organization (WHO), "Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020," [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
[5] E. Tobinick, "TNF-α inhibition for potential therapeutic modulation of SARS coronavirus infection," Current Medical Research and Opinion, vol. 20, no. 1, pp. 39-40, 2004.
[6] A. García-Salido, "Narrative review of the immune response against coronavirus: An overview, applicability for SARS-COV-2, and therapeutic implications," Anales de Pediatría (English Edition), vol. 93, no. 1, pp. 60.e1-60.e7, 2020.
[7] M. L. Jibril and U. S. Sharif, "Power of artificial intelligence to diagnose and prevent further COVID-19 outbreak: A short communication," arXiv preprint arXiv:2004.12463, 2020.
[8] P. McKenzie, "Life in the Time of COVID-19: The everyday impacts of the pandemic in Amman, Jordan," Doctoral dissertation, University of South Carolina, 2022.
[9] Jordan Ministry of Health, "Jordan's vaccination campaigns: Combating COVID-19 and restoring normalcy," 2022.
[10] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review," Chaos, Solitons & Fractals, vol. 139, p. 110059, 2020.
[11] K. U. Rehman et al., "Blockchain-enabled smart agriculture: Enhancing data-driven decision making and ensuring food security," Journal of Cleaner Production, vol. 427, p. 138900, 2023.
[12] S. Lupieri, "The Making of a 'Crisis': Syrian refugees and the politics of health in Jordan," Doctoral dissertation, 2021.
[13] Q. Zhang et al., "Inborn errors of type I IFN immunity in patients with life-threatening COVID-19," Science, vol. 370, no. 6515, p. eabd4570, 2020.
[14] S. Albayati and A. Altamimi, "Maintaining routine medical services amidst the COVID-19 pandemic: Jordan's perspective," 2023.
[15] D. M. Morens, P. Daszak, and J. K. Taubenberger, "Escaping Pandora's box—another novel coronavirus," New England Journal of Medicine, vol. 382, no. 14, pp. 1293-1295, 2020.
[16] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," Scientific Reports, vol. 10, no. 1, pp. 1-12, 2020.
[17] P. Sajda, "Machine learning for detection and diagnosis of disease," Annual Review of Biomedical Engineering, vol. 8, no. 1, pp. 537-565, 2006.
[18] J. Pang et al., "Potential rapid diagnostics, vaccine, and therapeutics for 2019 novel coronavirus (2019-nCoV): A systematic review," Journal of Clinical Medicine, vol. 9, no. 3, p. 623, 2020.
[19] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. M. U. Din, "Machine learning-based approaches for detecting COVID-19 using clinical text data," International Journal of Information Technology, vol. 12, no. 3, pp. 731-739, 2020.
[20] W. Naudé, "Artificial intelligence against COVID-19: An early review," IZA Discussion Papers, no. 13110, 2020.
[21] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," Internet of Things, vol. 11, p. 100222, 2020.
[22] T. T. Cai and P. Hall, "Prediction in functional linear regression," 2006.
[23] R. J. Tibshirani, "The lasso problem and uniqueness," 2013.
[24] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, vol. 12, no. 1, pp. 55-67, 1970.
[25] W. Aljedaani, I. Abuhaimed, F. Rustam, M. W. Mkaouer, A. Ouni, and I. Jenhani, "Automatically detecting and understanding the perception of COVID-19 vaccination: A Middle East case study," Social Network Analysis and Mining, vol. 12, no. 1, p. 128, 2022.
[26] J. Lupón et al., "Biomarker-assist score for reverse remodeling prediction in heart failure: The ST2-R2 score," International Journal of Cardiology, vol. 184, pp. 337-343, 2015.
[27] J. H. Han and S. Y. Chi, "Consideration of manufacturing data to apply machine learning methods for predictive manufacturing," in 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 109-113, IEEE, July 2016.
[28] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Climate Research, vol. 30, no. 1, pp. 79-82, 2005.
[29] A. Lasalvia et al., "Psychological impact of COVID-19 pandemic on healthcare workers in a highly burdened area of north-east Italy," Epidemiology and Psychiatric Sciences, vol. 30, e1, 2021.
[30] S. Baran and D. Nemoda, "Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting," Environmetrics, vol. 27, no. 5, pp. 280-292, 2016.

[31] F. S. Mohamad, M. Iqtait, and F. Alsuhimat, "Age prediction on face features via multiple classifiers," in 2018 4th International Conference on Computer and Technology Applications (ICCTA), pp. 161-166, IEEE, May 2018.

[32] H. Abu Owida, G. AlMahadin, J. I. Al-Nabulsi, N. Turab, S. Abuowaida, and N. Alshdaifat, "Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach," International Journal of Electrical & Computer Engineering, vol. 14, no. 3, 2024.

[33] A. Al Sharah, H. A. Owida, F. Alnaimat, and S. Abuowaida, "Application of machine learning in chemical engineering: outlook and perspectives," International Journal of Artificial Intelligence, vol. 13, no. 1, pp. 619-630, 2024.

[34] M. Iqtait, F. S. Mohamad, and M. Mamat, "Feature extraction for face recognition via active shape model (ASM) and active appearance model (AAM)," in IOP Conference Series: Materials Science and Engineering, vol. 332, no. 1, p. 012032, IOP Publishing, March 2018.