Journal of Informatics and Web Engineering
Vol. 4 No. 2 (June 2025) eISSN: 2821-370X

Performance Evaluation on COVID-19 Prediction using Machine Learning Models

Obai Ali Abderlahman[1], Naveen Palanichamy[2*], Su-Cheng Haw[3], Subhashini Gopal[4]
[1,2,3]Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Malaysia
[4]St. Joseph's Institute of Technology, OMR, Chennai, Tamil Nadu, 600119, India
** corresponding author: (p.naveen@mmu.edu.my; ORCiD: 0000-0003-4601-9770)*

Abstract - The COVID-19 pandemic has placed enormous strain on providing health care services internationally while reinforcing the argument for the need to strengthen forecasting techniques. Existing forecasting methods have drawbacks, especially in determining the long-term consequences of the pandemic and understanding its broad reach across various locations and populations. This project proposes an evaluation of machine learning (ML) models with the aim of improving predictions, particularly the accuracy in long-term forecasting, of subsequent trends of the COVID-19 pandemic. A systematic review highlights previous forecasting attempts as a reference for the approach. This project emphasizes extensive data collection, model formulation and testing to develop a strong prediction framework. The models considered for evaluation are Support Vector Regression (SVR), seasonal autoregressive integrated moving average (SARIMA), and artificial neural networks (ANN), which have overcome some of the deficiencies of epidemiological forecasting methods to date. The aim is to provide public health representatives with more rigorous forecasts, which could enhance planning and response measures and protect health and safety. Our findings show that the ANN model is superior, with high accuracy and comprehensive performance, confirming its broader use in various predictive applications. The Root Mean Square Error (RMSE) of prediction error was also relatively modest (R-square values were nearly 1).

Keywords— COVID-19, Machine Learning, Support Vector Regression, Seasonal Autoregressive Integrated Moving Average, Artificial Neural Networks

Received:12 November 2024; Accepted: 28 February 2025; Published: 16 June 2025

1. INTRODUCTION

The ongoing crisis caused by the COVID-19 pandemic has affected the worldwide economy as well as the health care system across various countries. Starting in 2020, the outbreak of the virus in different parts of the world and its consequences on the regions have indicated an acute need for innovative strategies to determine how the breathing monster moves and how to tame it. To better prepare the public health system, let alone avert the risks posed by the virus, having an accurate projection of the virus outbreaks is essential.

Historical data and several mathematical models have always been the basis for forecasting infectious diseases, with Epidemiological models leading the way [1], [2], [3]. Such models are predominantly based on past experiences, which create an outlook for new outbreaks. The models incorporate statistical and mathematical techniques to model the disease transmission process based on various ratios like infection rate, recovery rate and total death accounts. It is, however, worth noting that COVID-19 has unique features that defy most, if not all, of the assumptions of these

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

traditional models, like the degree to which COVID-19 varies in versatility, degree of transmittance, and the marketing and social factors. Such an unprecedented circumstance calls for an apparent change in the existing forecasting tools utilized to reflect the current circumstances surrounding the pandemic.

The weaknesses of the public health institutions have come to the front due to the COVID-19 pandemic. This scenario has reinforced the need for better computational techniques to advance forecasting capabilities over extended periods. The existing models for forecasting the disease outbreak have often not been accurate enough, particularly at critical times. Hence, this study employs machine learning techniques to predict COVID-19 more accurately to provide public health system managers with better resources for planning and intervention. To improve the predictability of the findings, models including ANN, SARIMA and SVR are developed in this research. By focusing on these perspectives, the study of the evolution of the pandemic will be improved, and the possibility of making public health management decisions will be of high quality.

2. LITERATURE REVIEW

COVID-19 is a new virus, a member of the Severe Acute Respiratory Syndrome (SARS) family called SARS-CoV-2. This virus has emerged as a global health crisis, affecting millions of people across the globe and posing a significant challenge to public health systems [4]. The nature of the pandemic is compounded by the speed at which it has spread and the differential extent of its effects in different parts of the globe. This situation has created an urgent demand for new approaches for comprehending and forecasting the development of the disease, which is necessary for efficient reaction and management mechanisms.

ML models have proved proficient in evaluating complex patterns over time and space in infectious diseases such as vector-borne, influenza and foodborne diseases [5]. Different countries have made use of various strategies to forecast the probability of the occurrence of outbreaks in the future. These include agent-based, time-series, compartment, and nonparametric weather forecasting models. Most of these models are built on the experience in terms of how such events and their duration are likely to unfold in the future [6]. They have been found helpful in preparing health service providers to prevent hand, foot and mouth disease (HFMD), dengue and malaria. Predictive models determine features and patterns that can be used to estimate the magnitude of potential health problems that may arise in the future. Such predictions assist in improving public health planning as they allow public health authorities to devise measures that avert worsening situations and formulate responses based on existing health problems. Among recommended ML algorithms used to predict the spread of specific diseases are the ones based on random forests for H1N1 influenza virus, swine flu, oyster virus, west Nile virus, and dengue [7], [8]. Other techniques include CART and LogitBoost for predicting dengue, genetic programming for predicting oyster norovirus and Bayesian networks for predicting Aedes mosquitoes and dengue. To improve Susceptible–Infected–Recovered (SIR) models, many researchers have started using ML as an additional computational tool.

Several ML techniques have been used to predict COVID-19 case numbers. Several studies have pointed out the substantial role ML could have in the containment of the pandemic and its evolution in various nations. The model [9] introduced an artificial neural network model that performs predictions of COVID-19 future trends in incidence and prevalence and mentions the applicability of shifting autoregressive integrated moving average (ARIMA) models. Another study [10], used data from Brazil to assess the effectiveness of different forecasting techniques in a comparative time series analysis, using Stacking-ensemble learning, Ridge Regression, Cubist Regression, random forests, ARIMA, and SVR for shorter term predictions. In the study by Zrieq [11] made predictions regarding the COVID-19 trends for the highest 15 infected nations as of 2020. Indeed, many of these models have even so raised important issues.

The ANN approach has received growing attention for the last two quite active decades due to its merits in pattern classification, mainly because it has nonlinear, nonparametric features and adapts learning processes. Quite a few authors have compared ANNs with forecasting techniques [12]. ANN has proved useful in predicting the epidemic burdens of diseases, including but not limited to the dynamics of COVID-19. For example, in Malaysia, ANN models were reported to predict the number of new infections, active cases, recovered cases, and vaccination coverage. These models are often proven effective using measures such as Root Mean Square Error (RMSE) and Pearson Correlation Coefficient. SARIMA models, on the other hand, can incorporate seasonal factors and are common in modelling and predicting time series in the context of epidemiology. This ability is significant for COVID-19 forecasting because the respective trends may have seasonal characteristics. In modelling SARIMA, autoregressive and moving average parameters are combined to model several complexities in the dynamics of virus transmission. As [13] explains, SARIMA is fit for near predictions where a 7-day round pattern has been noted.

According to [14], the SVR is a ML tool based on a statistical theory. It improves regression results for nonlinear, high-dimensional cases and with a small amount of data. The principle idea of SVR consists of three steps: first, the regression function is recast in the form of an optimization problem by integrating penalty factor terms into the problem of performing linear regression within the feature space that has been transformed using a nonlinear kernel function; second, the data is projected into a high dimensional feature space; third, nonlinear regression is performed in this space. The effectiveness of SVR with sequential moderating relations in predicting the number of COVID-19 cases in Zimbabwe was confirmed by [15]; which provided better results than the combined prediction and other predictions. To conclude, the literature review affirms the importance of ML models in fighting the COVID-19 pandemic and their potential to understand the factors guiding the dynamics of the outbreak. Nevertheless, the forecasts achieved by such models are not accurate over a significant period, which is a major setback because these types of long-term strategies are crucial for health systems. In addition, methods such as SARIMA and SVR are limited in terms of the number of regions and time frames that pandemic dynamics can cover. Thus, this study seeks to alleviate these shortcomings by developing ML models that can carry out long-term forecasts of the COVID-19 pandemic spread more accurately, integrate a variety of datasets and use real-time data to model human interaction behaviour to enhance the general applicability of the models. Table 1 outlines different ML models used in COVID-19 prediction studies. It consolidates information along two dimensions: different ML models listed in columns and distinct research articles in rows, stated by authors and their years of publication.

Table 1. Overview of Machine Learning Models for COVID-19 Prediction

References	ANN	ARIMA	SVR	SVM	DT	KNN	LR	RR	CR	SARIMA
[5]	✓									
[6]	✓		✓							
[7]	✓			✓	✓	✓				
[8]	✓					✓	✓			
[9]	✓									
[10]	✓	✓	✓					✓	✓	
[15]			✓							
[16]			✓							✓
[17]										✓
[18]										✓
[19]				✓	✓	✓	✓	✓		
[20]		✓								✓

3. RESEARCH METHODOLOGY

Figure 1 presents a detailed workflow in this research, covering multiple phases from data collection to model assessment. Ensuring the accuracy and reliability of the produced models requires careful consideration of each stage.

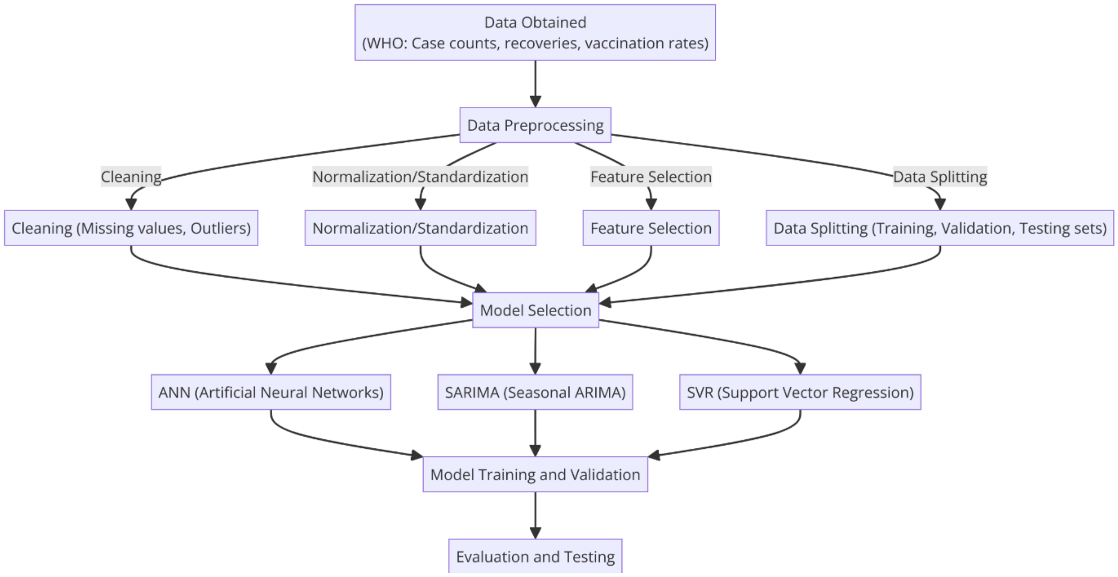

Figure 1. Research Workflow

3.1 Dataset

The dataset discussed in this research paper is obtained from the WHO [21]. It is an extensive dataset regarding the statistics involved in the global perspective of COVID-19. It is designed for empirical investigations, which collect data from multiple countries at different time periods and thus show the evolution of the pandemic over time. It consists of 51,360 rows and 8 columns, representing the pandemic's effect across various regions. The WHO enforces strict data collection and reporting protocols, ensuring reliability and timeliness. This dataset comprises the qualitative variables, including the names of countries and WHO regions, and the quantitative variables of case and death rates. These types of datasets make a variety of data analyses on the spatial and temporal range of the impact of COVID-19 possible. Table 2. shows the dataset columns and their description.

Table 2. Dataset Description

Columns	Description
Date_reported	Specific date on which the reported cases were stated.
Country_code	Standardized codes with each of them identifying a specific country.
Country	List of countries in the form of their international recognized names.
WHO_region	Categorizes countries according to their respective WHO regions.
New_cases	Captures the number of new cases reported on a given day.
Cumulative_cases	Total number of diagnosed positive cases over a period.
New_deaths	Number of new deaths caused by COVID-19 on a specific date.
Cumulative_deaths	Captures the total number of death cases caused by COVID-19 on a particular day.

3.2 Data Preprocessing

Preprocessing is also essential in preparing datasets for analysis, as it allows the data to be well-structured, homogeneous and suitable for a model. The COVID-19 dataset is facing some challenges that need to be resolved, such as cleaning missing values, dealing with outliers, normalizing, selecting relevant features, and partitioning the data, which are essential to ensure the quality and suitability of the subsequent analyses.

The dataset contains missing values and thus needs to be cleaned. For instance, 214 entries are missing for Country_code, 3,852 for WHO_region, 14,845 for New_cases, and 27,139 for New_deaths. Filling in such gaps is fundamental to maintaining the quality of the analysis. Based on the consideration and the importance of the findings, we used strategies such as imputation or removal to address the missing data effectively.

Examining the summary statistics reveals specific extreme data points in the dataset. For instance, a scenario with negative cases in New_cases (being -65,079 at its minimum) and New_deaths (having a minimum of -3432) is said to be unrealistic. Also, the significant variability indicated by the standard deviations and the range between the 75th percentile and the maximum values for both New_cases and New_deaths supports this observation. It becomes vital to determine the outliers and deal with them appropriately so as not to distort the results to ensure the correctness of the analysis. We took steps to preserve the accuracy of the data by correcting these numbers and capping or removing extreme outliers. We used the z-score and interquartile range (IQR) techniques to address extreme values.

Because of the significant variation exhibited by New_cases, Cumulative_cases, New_deaths, and Cumulative_deaths values, normalization or standardization is an essential data preprocessing step. The values in these columns range from zero for cases to more than 40 million, while death cases are more than 1 million. It is difficult to model these features without normalizing or standardizing the data set because the difference in scale may be too significant and give undue weight to the training phase of the model for some features. In our case, we used Min-Max scaling to normalize the data so that the values will range from 0 to 1. By employing these methods, we can argue that all the variables are equally efficient and are non-redundant in each analysis, thereby increasing the strength of our models.

The most important task of the predictive modelling process is identifying the key features that would be crucial to the prediction outcome. In this dataset, New_cases, Cumulative_cases, New_deaths, and Cumulative_deaths variables are the key predictors of the outcome of this analysis, which was the impact of COVID-19. However, applying feature selection techniques will refine our analyses as we can define features with the strongest prediction power. This may require determining whether inputs such as Country_code and WHO_region should be considered relevant and valuable in prediction. We used feature selection methods to retain the most salient features, including feature priority and correlation matrices obtained from the machine learning models.

The dataset's extensive size makes it possible to create a training-validation-test split. This division is crucial in testing a model's performance evaluation metrics as they claim to generalize to new data. Adopting this plan would help us lower overfitting while improving the quality of our results. The training, validation, and testing datasets were split with a ratio of 70%, 15%, and 15%, respectively.

3.3 Model Selection, Training and Validation

In this study, we implemented three ML models - ANN, SARIMA and SVR- to predict COVID-19 trends. Their notable specific advantages informed the selection of these models in addressing several challenges posed by the pandemic. ANN is a more robust model of the three models because it assists with prediction. Its unique capacity to capture complex frameworks allows it to be used in a dynamic situation as that is the case with the spread of COVID-19. SARIMA was selected because of its seasonality inclusion, which is critical in explaining oscillations in the COVID-19 series. This model is helpful in short- and medium-term forecasting in countries that exhibit such trends because it accounts for seasonality. SVR featured in the analysis as it has proved to be successful in regression problems in high dimensional spaces.

3.4 Prediction and Evaluation

The predictive accuracy of the unseen data will be computed after the models have been trained and validated. In making this step, the performance of the models in real world scenarios will be evaluated. The models were assessed by looking into the accuracy of ANN, SARIMA, and SVR. The evaluation was conducted using the following metrics:

- R-squared (R^2): It indicates the proportion of variance in the dependent variable explained by the independent variables. It also correlates to the model's goodness. The closer the value of R^2 is to 1, the better the model captures COVID-19 trends.
- Mean Squared Error (MSE): The mean of squared errors is the mean squared deviations between predicted and actual values. Small MSE values indicate high precision.

- Root Mean Squared Error (RMSE): It is a valuable measure when assessing any model to provide estimates of actual outcomes. It indicates the residual error through square root of mean square error. Small RMSE values show less deviation from predicted values, which helps determine model performance.

4. MODELS IMPLEMENTATION AND VALIDATION

In this section, the implementation and validation of predictive models, such as ANN and SARIMA, play a critical role in forecasting COVID-19 case trends. These models help public health authorities and policymakers make informed decisions about resource allocation, containment strategies, and early warning systems. This study uses both ANN and SARIMA models to predict various types of COVID-19 cases—new, active, and recovered—based on historical data. The ANN models are evaluated based on key metrics such as R^2, MSE, and RMSE, while the SARIMA model leverages time series data to capture seasonal trends. This paper outlines these models' configurations, performance, and outcomes, providing valuable insights for COVID-19 trend prediction and healthcare planning.

4.1 ANN

In terms of New Cases, the ANN model presented an R squared value of 0.89 which suggests that the ANN can model 89% of the variability in new COVID-19 case data. The overall Mean Squared Error (MSE) equals to 14509697.730, while the RMSE equals to 28284 which points to moderate error in predicting new case trends. For Recovered Cases, the computed R squared was estimated to be 0.91, which performs better than the previous model. The MSE also falls to 12008600.075 while the RMSE falls further to 27386, showing a fair degree of precision in recovery trend forecasting. Regarding the Active Cases, the ANN model attained an R squared of 0.93, MSE of 12008600.075 and RMSE of 30822. This accuracy, although not at par with that of recovered cases, recovery forecasting still suggests strong performance in active case forecasting as well. Table 3, Table 4 and Table 5 show the overall performance, R^2 and MSE values respectively for different cases.

Table 3. Overall Performance of ANN

	ANN Configuration	R^2	MSE	RMSE
New Case	3-16-1	0.9865	1009625.187	1004.80
Recovered case	3-14-1	0.9876	1198000.675	1094.53
Active case	3-20-1	0.9962	42363918.89	6508.75

Table 4. R^2 Value for ANN

	ANN Configuration	R^2	MSE	RMSE
New Case	3-16-1	0.9832	0.9861	0.9911
Recovered case	3-14-1	0.9898	0.9816	0.9852
Active case	3-20-1	0.9964	0.9956	0.9985

Table 5. MSE Values for ANN

	ANN Configuration	R^2	MSE	RMSE
New Case	3-16-1	8298940.59	1591418.269	1263871.848
Recovered case	3-14-1	1012386.40	1893696.83	1365710.246
Active case	3-20-1	39233782.47	47519059.13	51768961.619

4.2 SARIMA

Regional and country separation was vital when applying SARIMA, as it allowed us to target the seasonal trend with smaller data. To this end, the selected dataset comprised the daily updates of Malaysia's new, active, and recovered COVID-19 cases. Concerning Malaysian COVID-19 case trends, different modifications of the SARIMA models, including forecasting new cases, recovered cases, and active cases, were executed, and performance metrics were computed.

Table 6 shows the descriptive statistics of new, recovered and active cases of COVID-19 in Malaysia. The data were obtained daily between January 25, 2020, and January 24, 2022. Table 7 provides a detailed breakdown of the distribution of COVID-19 cases predicted by the SARIMA model, including the 25th percentile, median (50th percentile), 75th percentile, and maximum values for new, recovered, and active cases in Malaysia. This percentile distribution offers valuable insights into the variability and spread of case numbers, which can inform healthcare policy adjustments and resource planning.

Table 6. SARIMA - Descriptive Statistics

Case Type	New Cases	Recovered Cases	Active Cases
Count	358	358	358
Mean	7,321.79	7,243.02	88,142.25
Std Dev	6,042.34	5,992.63	74,107.56

Table 7. SARIMA - Distribution of Cases

Case Type	New Cases	Recovered	Active Cases
25th Pct	3,140.50	3,196.25	39,808.25
50th Pct	5,231.00	5,090.00	8,808.75
75th Pct	8,982.00	8,808.75	107,790.0
Max	24,599	24,855	268,552

The following summary presents forecasts of active COVID-19 cases based on the SARIMA endeavours model. They are essential for planning healthcare responses and resource needs. Figure 2 presents active cases for a predictive six-month period and indicates whether the increase will continue or subside. This forecast is intended to further and enhance strategic planning and activity forwarding. Figure 3 focuses on a shorter-term projection that anticipates the necessity for course changes aimed at immediate aspects of control, for example, quarantine requirements and provision of suitable medical support resources. Figure 4 case is concerned with the very next forecast period. It is essential for the first months of the outbreak, including emergency proclamation on public health emergencies and reallocating funds for the first time. Table 8 shows the overall performance of SARIMA.

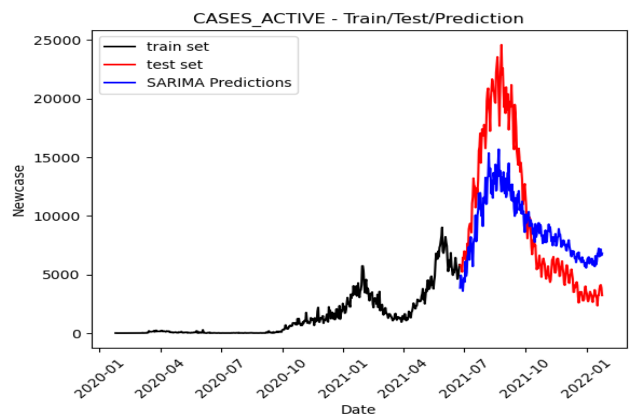

Figure 2. Prediction of Active COVID-19 Cases for the Period of 6 Months

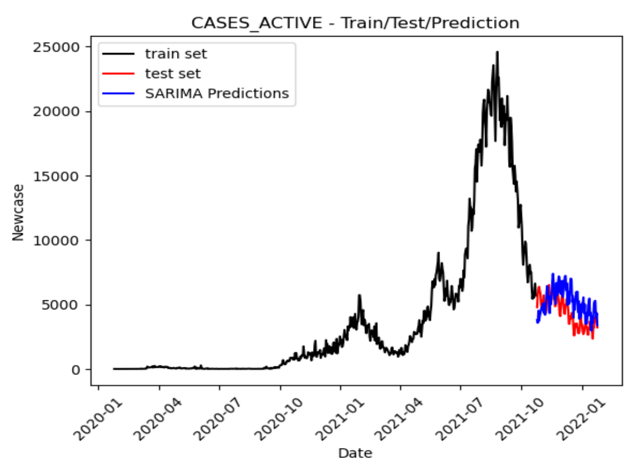

Figure 3. Prediction of Active COVID-19 Cases for the period of 3 Months

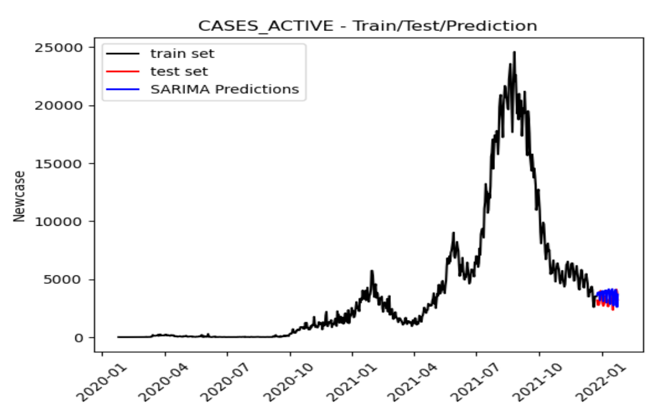

Figure 4. Prediction of Active COVID-19 Cases for the period of 1 Month

Table 8. SARIMA Overall Performance

	Configuration	R^2	MSE	RMSE
New Cases	(1,1,2)(1,1,1)12	0.95	0.92	0.94
Recovered Cases	(1,1,2)(1,1,1)12	0.96	0.93	0.95
Active Cases	(1,1,2)(1,1,1)12	0.94	0.91	0.93

4.3 SVR

The capability of the SVR model for forecasting COVID-19 data relevant to the pandemic is helpful for public health experts and policymakers in designing appropriate disease control strategies and setting up tracking and alert systems. The performance of the model was assessed in terms of R2, RMSE, and MSE, while for the dataset on COVID-19, separate SVR models were developed for active cases, new cases, recovered cases and vaccination data.

Table 9 presents the performance evaluation of SVR models in predicting different categories of COVID-19 cases, including new cases, recovered cases, and active cases. For new cases, the SVR model's R2 value is 0.89, which means that the model can explain 89% of the variance in the new case data. The MSE is 14,509,697.730, and the RMSE is 28,284, which suggests a fair amount of error in predicting new case trends. In the case of Recovered Cases, the MSE is 12,008,600.075, and the RMSE is 27,386, which explains an improvement in the model's R2, which was recorded as 0.91. The RMSE, on the other hand, shows fair performance in predicting recovery trends for that model growth. For Active Cases, focusing specifically on MSE and RMSE for new case number predictions says the SVR model achieved an MSE of 12,008,600.075 and an R2 value of 0.93, along with RSME for the cases, which was 30,822, which is considered ok. While this performance is slightly lower than recovered cases, it still reinforces their claim that active case numbers can be predicted more reliably.

Table 9. SVR Overall Performance

	R^2	MSE	RMSE
New Cases	0.89	14509697.730	28,284
Recovered Cases	0.91	12008600.075	27,386
Active Cases	0.93	12008600.075	30,822

4.4 Model Comparison

The comparison results are provided in Table 10, which shows the competitive standing of the models in all prediction categories: new cases, recovered cases and active cases, respectively. It was established that the ANN model performed better than SARIMA and SVR because it had the lowest MSE and RMSE values and the highest R^2. The model achieved the following: for example, in new cases, the listed MSE by ANN was 1,009,625.187, and 1,004.80 for RMSE, giving a resultant R^2 value of 0.9865. On the other hand, SARIMA had an MSE of 0.92 and RMSE of 0.94, whereas SVR MSE and RMSE attained values of 14,509,697.730 and 2,8284, respectively. Regarding recovered cases, the previous findings were reinforced by all models where ANN, in this case, had the best score with an MSE of 1,198,000.675 and R^2 of 0.9876. SARIMA's performance was good, with an MSE of 0.93, but overall measures were not as reasonable as ANN. Results obtained for SVR are equally poor, with an MSE of 12,008,600.075. As for the active cases, the previous observations were upheld, and ANN performed with the R^2 of 0.9962 and had lower RMSE than both SARIMA and SVR.

In conclusion, ANN achieved high accuracy, as demonstrated by R2 values close to unity and low values of the RMSE. Because of the capability to model complex structures in the data, it has a wide range of applicability in all types of predictive tasks. The ability of the ANN to assimilate historical data and evolve with the changing times emphasizes the need to use it in prediction models regarding the fluctuating trends of the COVID-19 pandemic. On the other hand, the SARIMA model performed fairly. It could recognize seasonal effects, but reliance on past data hampered accurate forecasts. In the meantime, the SARIMA model is suitable for making short-term forecasts and using any cyclic patterns that have been previously established but do not have the capacity to respond to the brisk temporal changes of pandemics. The SVR model also performed moderately, although its performance was sensitive to parameter tuning. How SVR configurations work is quite complex, even though it can model nonlinear relationships. Within this study, SVR worked well in certain situations; however, it was not always able to reach the accuracy of the required ANN model. In conclusion, although each of the three models has advantages, the ANN model stands out in terms of the ability to predict and process complex data. This finding highlights the need to adopt more sophisticated machine-learning techniques to improve the accuracy of the forecasts made in public health contexts.

Table 10. Overall Model Comparison

Cases	Metrics	ANN	SARIMA	SVR
New Cases	MSE	1,009,625.187	0.92	14,509,697.730
	RMSE	1,004.80	0.94	2,8284
	R^2	0.9865	0.95	0.89
Recover Cases	MSE	1,198,000.675	0.93	12,008,600.075
	RMSE	1,094.53	0.95	27,386
	R^2	0.9876	0.96	0.91
Active Cases	MSE	42,363,918.89	0.91	12,008,600.075
	RMSE	6,508.75	0.93	30,822
	R^2	0.9962	0.94	0.93

5. CONCLUSION

The research reported in this paper aimed to compare the effectiveness of three models, ANN, SARIMA, and SVR, regarding the spread and impact of COVID-19. Each model has certain substantial aspects, making them ideal for dealing with different areas of the pandemic data analysis. For instance, ANNs are good at assessing high order and nonlinear features; they can obtain accurate values (high determinant coefficients, R^2 very close to 1) and low prediction errors (RMSE) which is evidence of its case type employing active, new, recovered, or vaccination data that various models can use. For example, SARIMA captures seasonal sinusoidal fluctuations, which is essential for short-term to medium-term monitoring in places with a cyclic tendency of infection. Yet, it does not perform well using non-periodic or global data sets. SVR can generate high-dimensional data, but it does not seem accurate as it suffers from a high RMSE default. Considering the above statements, each of these three models has its own merits; the ANN model appears to be the most advantageous in prediction and complex data handling. Future studies need to focus on further improving these models by integrating their components with accurate data and investigating mixed models in forecasting to reduce errors and increase forecasting reliability. The constantly changing state of machine-learning approaches is a good opportunity to enhance public health management, aiming to increase the level of preparedness and response to potential outbreaks of diseases.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Obai Ali Abdelrahman: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Naveen Palanichamy: Project Administration, Supervision, Writing – Review & Editing;
Su-Cheng Haw: Project Administration, Writing – Review & Editing;
Subhashini Gopal: Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] Q. Lin, L. W. Ang, Y. Shao, and S. Palaniappan, "Temporal Climatic Shifts in Henan Province: A 16-decades Perspective Through Regression, SARIMA, and NAR Modeling," *Journal of Informatics and Web Engineering*, vol. 3, no. 2, pp. 159–168, Jun. 2024. doi: 10.33093/jiwe.2024.3.2.12.

[2] M. Y. Xin, L. W. Ang, and S. Palaniappan, "A Data Augmented Method for Plant Disease Leaf Image Recognition based on Enhanced GAN Model Network," *Journal of Informatics and Web Engineering*, vol. 2, no. 1, pp. 1–12, Mar. 2023. doi: 10.33093/jiwe.2023.2.1.1.

[3] C. C. Chai, W. H. Khoh, Y. H. Pang, and H. Y. Yap, "A Lung Cancer Detection with Pre-Trained CNN Models," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 41–54, Feb. 2024. doi: 10.33093/jiwe.2024.3.1.3.

[4] D. B. Cohen, M. Luck, A. Hormozaki, and L. L. Saling, "Increased meaningful activity while social distancing dampens affectivity; mere busyness heightens it: Implications for well-being during COVID-19," *PLoS One*, vol. 15, no. 12, p. e0244631, Dec. 2020. doi: 10.1371/journal.pone.0244631.

[5] B. Manohar and R. Das, "Artificial Neural Networks for the Prediction of Monkeypox Outbreak," *Trop Med Infect Dis*, vol. 7, no. 12, p. 424, Dec. 2022. doi: 10.3390/tropicalmed7120424.

[6] D. T. Andariesta and M. Wasesa, "Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: a multisource Internet data approach," *Journal of Tourism Futures*, Jan. 2022. doi: 10.1108/JTF-10-2021-0239.

[7] N. A. Nayan *et al.*, "COVID-19 Prediction With Machine Learning Technique From Extracted Features of Photoplethysmogram Morphology," *Front Public Health*, vol. 10, Jul. 2022. doi: 10.3389/fpubh.2022.920849.

[8] A. Becerra-Sánchez *et al.*, "Mortality Analysis of Patients with COVID-19 in Mexico Based on Risk Factors Applying Machine Learning Techniques," *Diagnostics*, vol. 12, no. 6, p. 1396, Jun. 2022. doi: 10.3390/diagnostics12061396.

[9] P. Kumari and D. Toshniwal, "Real-time estimation of COVID-19 cases using machine learning and mathematical models - The case of India," in *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, IEEE, Nov. 2020, pp. 369–374. doi: 10.1109/ICIIS51140.2020.9342735.

[10] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos S. Coelho, "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," *Chaos Solitons Fractals*, vol. 135, p. 109853, Jun. 2020. doi: 10.1016/j.chaos.2020.109853.

[11] R. Zrieq *et al.*, "Predictability of COVID-19 Infections Based on Deep Learning and Historical Data," *Applied Sciences*, vol. 12, no. 16, p. 8029, Aug. 2022. doi: 10.3390/app12168029.

[12] M. S. Ghanim, D. Muley, and M. Kharbeche, "ANN-Based traffic volume prediction models in response to COVID-19 imposed measures," *Sustain Cities Soc*, vol. 81, p. 103830, Jun. 2022. doi: 10.1016/j.scs.2022.103830.

[13] C. V. Tan *et al.*, "Forecasting COVID-19 Case Trends Using SARIMA Models during the Third Wave of COVID-19 in Malaysia," *Int J Environ Res Public Health*, vol. 19, no. 3, p. 1504, Jan. 2022. doi: 10.3390/ijerph19031504.

[14] Z. Fu and Z. Wang, "Prediction of Financial Economic Time Series Based on Group Intelligence Algorithm Based on Machine Learning," *Turkish Journal of Field Crops*, vol. 26, no. 2, pp. 492–502, 2021.

[15] C. Shoko and C. Sigauke, "Short-term forecasting of COVID-19 using support vector regression: An application using Zimbabwean data," *Am J Infect Control*, vol. 51, no. 10, pp. 1095–1107, Oct. 2023. doi: 10.1016/j.ajic.2023.03.010.

[16] A. Yaqin, M. Rahardi, F. F. Abdulloh, Kusnawi, S. Budiprayitno, and S. Fatonah, "The Prediction of COVID-19 Pandemic Situation in Indonesia Using SVR and SIR Algorithm," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, Dec. 2022, pp. 570–573. doi: 10.1109/ICITISEE57756.2022.10057813.

[17] G. Perone, "Using the SARIMA Model to Forecast the Fourth Global Wave of Cumulative Deaths from COVID-19: Evidence from 12 Hard-Hit Big Countries," *Econometrics*, vol. 10, no. 2, p. 18, Apr. 2022. doi: 10.3390/econometrics10020018.

[18] K. Duangchaemkarn, W. Boonchieng, P. Wiwatanadate, and V. Chouvatut, "SARIMA Model Forecasting Performance of the COVID-19 Daily Statistics in Thailand during the Omicron Variant Epidemic," *Healthcare*, vol. 10, no. 7, p. 1310, Jul. 2022. doi: 10.3390/healthcare10071310.

[19] S. T. Lim, J. Y. Yuan, K. W. Khaw, and X. Chew, "Predicting Travel Insurance Purchases in an Insurance Firm through Machine Learning Methods after COVID-19," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 43–58, Sep. 2023. doi: 10.33093/jiwe.2023.2.2.4.

[20] A. Andueza, M. Á. Del Arco-Osuna, B. Fornés, R. González-Crespo, and J.-M. Martín-Álvarez, "Using the Statistical Machine Learning Models ARIMA and SARIMA to Measure the Impact of Covid-19 on Official Provincial Sales of Cigarettes in Spain," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 1, p. 73, 2023. doi: 10.9781/ijimai.2023.02.010.

[21] "COVID-19 data | WHO COVID-19 dashboard," Datadot. Accessed: Feb. 24, 2024. [Online]. Available: https://data.who.int/dashboards/covid19/data

BIOGRAPHIES OF AUTHORS

	**Obai Ali Abderlahman** is an ambitious final-year undergraduate student pursuing a Bachelor of Computer Science (Honors) in Data Science at Multimedia University in Malaysia. Throughout his academic journey, he has demonstrated exceptional problem-solving skills. He is passionate about building a career in data science or machine learning, where he aims to develop innovative solutions that address real-world challenges and create a meaningful impact. He can be contacted at email: obaialid45@gmail.com.
	**Palanichamy Naveen** is a Senior Lecturer at the Faculty of Computing and Informatics, Multimedia University. She joined the faculty after completing her PhD at Curtin University, Malaysia. She earned her Bachelor of Engineering (CSE) and Master of Engineering (CSE) from Anna University, India. Her research interests include smart grids, cloud computing, machine learning, deep learning, generative AI and recommender systems. She is actively involved in several research projects funded by Multimedia University. She can be contacted at p.naveen@mmu.edu.my.
	**Su-Cheng Haw** is Professor at Faculty of Computing and Informatics, Multimedia University, where she leads several funded projects on the XML databases. Her research interests include XML databases, query optimization, data modelling, semantic web, and recommender system. She can be contacted at email: sucheng@mmu.edu.my.
	**Subhashini Gopal** earned her Bachelor and Master of Engineering in Computer Science at Anna University, where she is now a PhD candidate. As an Assistant Professor at St. Joseph's Institute of Technology in Chennai, India, she teaches and mentors students. Her research focuses on Deep Learning, Machine Learning, Natural Language Processing, and Cloud Computing. She actively contributes to the academic community by publishing her findings and presenting at conferences and journals related to these fields. She can be contacted at email: gopal.subha191@gmail.com.