International Journal of Innovation Sciences and Research

Vol.10, No, 05, pp.1643-1650, May, 2021

Available online at http://www.ijisr.com

RESEARCH ARTICLE

APPLYING MACHINE LEARNING IN HEALTHCARE TO FIGHT THE COVID-19: A CASE STUDY OF FORECASTING THE OUTBREAK IN SAUDI ARABIA

Ali M. Alzahrani[1,2], Tabrez Uz Zaman[1], Shroug S. Alshehri[3] and Mohamed O. Khozium[1,4*]

[1]Department of Health Information Technology and Management, Faculty of Public Health and Health Informatics, Umm Al-Qura University, Makkah, Saudi Arabia

[2]Department of Health Services Management, Faculty of Public Health and Health Informatics, Umm Al-Qura University, Makkah, Saudi Arabia; and King Abdullah Medical City, Makkah, Saudi Arabia

[3]College of Computer and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia

Received 17th March, 2021; Accepted 10th April, 2021; Published 30th May, 2021

ABSTRACT

During this evolved worldwide health crisis, the medical industry is searching for new technologies to monitor and control the Coronavirus pandemic's spread. Artificial intelligence is a technology that can easily trace the reach of this virus, classify high-risk patients, and maintain this infection in real-time. It can also foretell mortality risk by analyzing the previously obtained data of the patients. It can help oppose this virus by population screening, medical help, notification, and infection control suggestions. Machine Learning (ML) techniques help improve detection, tracing, prognosis, and prevention. This paper will explore how the healthcare sector benefits from utilizing ML models to fight against the COVID-19. It introduces the techniques that enhanced health care delivery efficiency, accessibility, and cost-effectiveness during the pandemic. This work presents a case study of forecasting the COVID-19 numbers in Saudi Arabia using the ARIMA model to provide further insights and practically support health care management.

Key words: ARIMA Model, COVID-19, Forecasting, Infections, Mortalities; Vaccinations.

INTRODUCTION

Healthcare is advanced and expensive nowadays. After the 1950s, healthcare was transformed from a fragmented distribution of healthcare services to an organized healthcare delivery effort. The beginning of the 1970s had an increasing interest in primary healthcare, medicine, and awareness about public health activities. The focus was to provide people healthcare through consultations in various fields of medicine. There were improvements like various diagnosis and treatment methods through increased research and development. Health promotion was gradually sought after in different parts of the world (1). The beginning of 2020 brought the novel Coronavirus (COVID-19) without any documented history of causing disease in humans (2), (3). Though Middle East Respiratory Syndrome (MERS) allowed different countries to learn some lessons, COVID-19 was a pandemic, and no country was prepared to deal with it. The healthcare systems in different countries were poorly affected, and hospitals were full of COVID-19 patients with no admissions for chronic diseases. Adding to the trouble for different countries was the unavailability of specific treatment or vaccines to manage and prevent the disease's spread (4), (5). World Health Organization responded to this situation by guiding and supporting different countries' healthcare systems (6), (7). Social distancing and other community mitigation methods proved further to prevent the virus's spread (8).

Corresponding author: *Mohamed O. Khozium[1,3*]

[1]Department of Health Information Technology and Management, Faculty of Public Health and Health Informatics, Umm Al-Qura University, Makkah, Saudi Arabia.

[3]College of Computer and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia.

Artificial Intelligent (AI) approaches are becoming more widely used for healthcare. The efficacy of this method is proven during the beginning of the pandemic. Artificial intelligence describes machines that mimic cognitive functions, such as learning, analyzing, drawing conclusions, and problem-solving (9). Machine learning (ML) is considered a type of artificial Intelligent where machines can learn information without human intervention. The machine is provided with data to learn by revealing patterns and underlying algorithms (10). There are many notable instances of ML concepts being applied in healthcare. Most healthcare applications are diagnosis, clinical treatment suggestions, drug discovery process, and outbreak prediction (11)–(14). The prediction has tremendously helped understand the COVID-19 trends to alarm stakeholders and decision-makers. This work's primary purpose is to explore the ARIMA model and forecast the number of infections, mortalities, and vaccinations in the coming sections. This paper practically benefits everyone involved in the administration of public health and health care management, such as hospitals, medical centers, and health departments. This paper consists of five sections. The first section is this introduction. Next, the second section presents an overview of recent reviews and surveys on the same topic. The third section provides a case study to predict the number of cases and the vaccinations in Saudi Arabia. The fourth section is the discussion. Finally, the last section is the conclusion.

RELATED WORK

This section provides a comprehensive literature review on healthcare, machine learning, and the different methods used to monitor, prevent, and forecast the current pandemic covid-19.

initially, it delivers the perceptions about different survey researches in the correlated fields, thereby revealing this paper's contribution. subsequently, an exploration of some related works was conducted, which has widely influenced this research. The paper contributes to the present and innovative machine learning field by summarizing and classifying the state-of-art techniques used for different domains, particularly for disease detection, monitoring, forecasting, and healthcare prevention (14). In machine learning, computers possess the capacity to learn and do not require explicit programming, which is possible because they utilize complex algorithms and other techniques to distinguish data patterns, thereby making predictions (15). There is a steep rise in various epidemiological, clinical, and other medical data that helps machine learning techniques. Medical practice has greatly benefitted from this decision-making, innovative optimizations, conducting efficient clinical trials and research including HIV (Human Immunodeficiency Virus) patients, developing better tools for physicians, healthcare providers, and benefitting other non-health sectors. (15), (16). During COVID-19, one of the researchers (17) had emphasized the potential applications involving digital technologies such as analysis of big data, internet of things (IoT), artificial intelligence, deep learning techniques, and block-chain technology. These applications are used to make strategies for controlling this epidemic through forecasting techniques and identifying the impact of COVID-19 on the health care sector (17). Machine learning methods can quickly identify and determine complex dataset patterns and thereby forecast future outcomes in terms of diagnosis and prognosis of diseases, which is beneficial in healthcare (18). The advent of electronic health records (EHRs) has provided the opportunity to accommodate the influx of enormous medical data generated during diagnosis and research (19). The different machine learning models can routinely extract essential features from these enormous patient records datasets stored in the EHRs, thereby helping in disease detection, monitoring the data, and predicting probable diseases.

It was found that these machine-learning algorithms played a crucial role in analyzing the epidemic data and forecasting some epidemic patterns so that future action and improved decision-making could be done to stop the spread of the virus (20), (21). Different machine-learning models have long experimented in several application domains to counter the adverse factors for a threat but after proper identification and prioritization of these adverse factors (22). Several ML prediction methods popularly used nowadays can forecast the upcoming number of COVID-19 affected patients, which was a potential risk to humanity. These four standard forecasting models include the linear regression (LR) model, support vector machine (SVM) model, a minor absolute shrinkage and selection operator (LASSO) model, and exponential smoothing (ES) model particularly. The three main areas where these predictions focus are the number of freshly infected cases, the number of deaths, and the recovery rate in the coming ten days (22)–(25). Simultaneously, machine learning techniques also have some challenges, including a proper choice of parameters and an appropriate prediction model. Some researchers have used datasets for predictions also (26)–(30). In the research work (31), they used a model based on the Logistic, Weibull, and the Hill equations to detect China and Italy's infection rates. It needs to be mentioned that relative humidity, wind speed, and temperature comprised the three main environmental factors considered for this study (31). Benvenuto et al., suggested an ARIMA model for the prediction of COVID-19 spread (32). This research forecasted the different parameters for the upcoming two days based on the incidence and prevalence of COVID-19. The ARIMA forecast graph signifying the epidemic prevalence and incidence has also been discussed (32). Until all the countries roll out the COVID-19 vaccines and all people achieve robust immunity, these forecasting methods will support all the governments to design and support effective social distancing measures to combat the virus's spread in the future (33).

CASE STUDY: Forecasting the COVID-19 outbreak in Saudi Arabia

The focus in this section is regarding the process of applying the ARIMA model to forecast the number of cases and vaccinations in Saudi Arabia.

METHODOLOGY

The ARIMA model is one of the commonly used models to predict the dynamics of COVID-19 (25), (34). ARIMA is a statistical analysis model that uses time-series data to understand better the data set or forecast future trends (35). In Pakistan, Yousaf et al. have employed the ARIMA model to forecast the number of confirmed cases of COVID-19, the number of recovered cases, and the number of deaths. The results were alarming and showed an increase in the number of confirmed cases, recoveries, and deaths. Therefore, the authors urge the government to develop new strategies to control the pandemic (36). Ceylan has employed the ARIMA model with different parameters to predict the epidemiological trends of COVID-19 prevalence in Italy, Spain, and France. The study demonstrated that the ARIMA model was suitable to predict the prevalence of COVID-19 in the future (37). Also, Sahai et al. have employed the ARIMA model to forecast the incidence and spread of COVID-19 in the five most badly hit countries (India, Russia, Brazil, Spain, and the US). The findings revealed that the model had a good forecast, and governments can use it to manage and ramp up their healthcare preparedness for the pandemic (38). The ARIMA is a mass model of two different models: The Auto-Regressive (AR) model and the Moving Average (MA) model. This model results over Akaike information criterion (AIC) statistics and coverage of regression analysis. The autoregressive models try to forecast a series based solely on the series's past values, called lags. A model has a linear function of p lags plus error –AR model of order p (AR_p). This recursion in time goes back to the beginning of the series, which is why they are called long memory models. The model expression is shown in equation (1).

$$Y_t = \omega + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t \qquad (1)$$

Y_t is the target value of interest, and Y_{t-2} is the lagged target. e_t is the error. ω is the intercept, and $\emptyset$ is the coefficient. In the moving average models, instead of past values, the forecast is based solely on the series's past errors, called error lags. If a model depends on q previous error lags, it is an MA model of order q (MA_q). They are considered short memory models because these errors do not last long into the future. Equation (2) shows the model expression.

$$Y_t = \omega + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \qquad (2)$$

Table 1. Test results for the infection and mortality cases

Data	Model	Mean	RMSE	MAPE	RScore	Mean absolute error	Median absolute error	MSE	MSLE
New cases	ARIMA(0,1,0)	613.5	281.30	31.14	-1.11	224.34	181.34	79131.80	0.23126
Total cases	ARIMA(0,2,0)	390631.0	2804.50	0.50	0.73	1971.24	1181.22	7865179.0	0.00005
New deaths	ARIMA(2,0,0)	6.93	2.47	26.62	-1.77	2.03	2.03	6.09	0.10609
Total deaths	ARIMA(1,1,1)	6671.0	19.96	0.21	0.89	14.18	8.93	398.40	0.00001

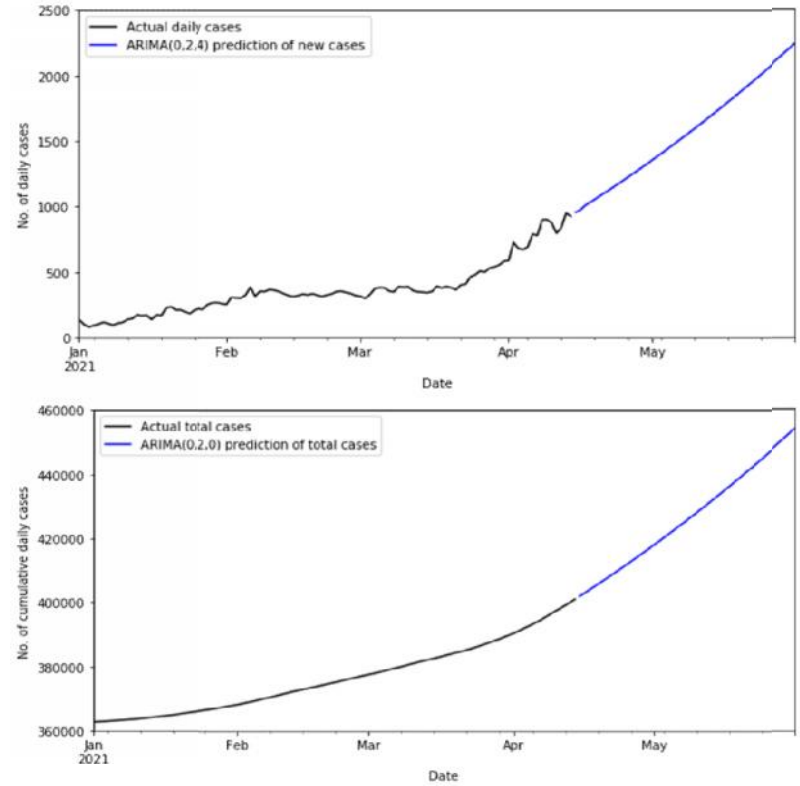

Figure 1. ARIMA prediction results of new cases and total cases in Saudi Arabia

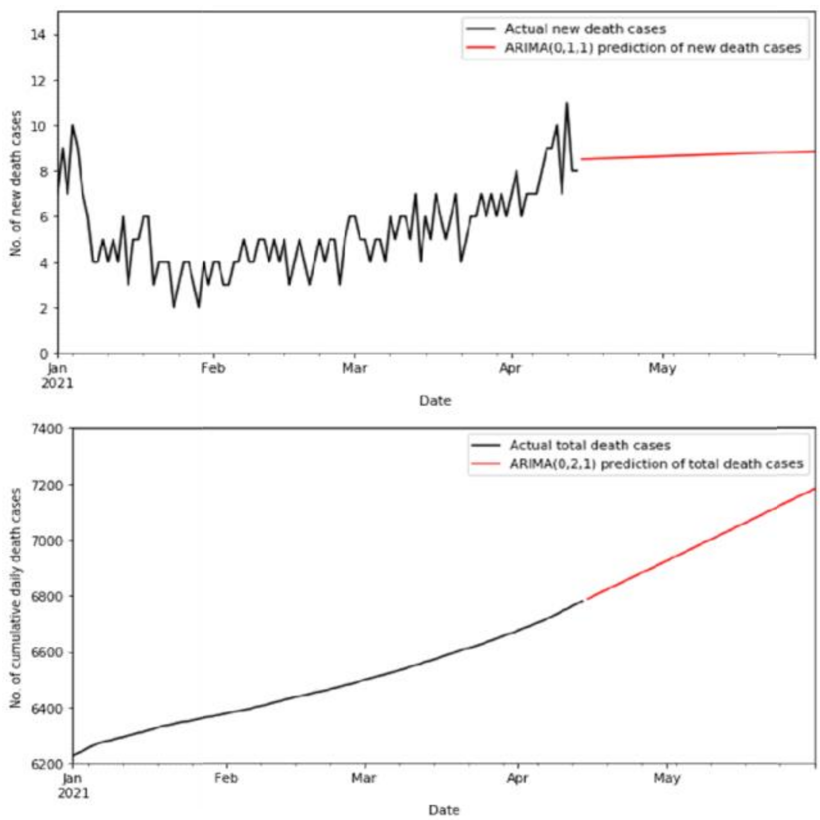

Figure 1. ARIMA prediction results of new death cases and total death cases in Saudi Arabia

Table 1. Test results for vaccination

Data	Model	Mean	RMSE	MAPE	R Score	Mean absolute error	Median absolute error	MSE	MSLE
Newvaccinations	(2,1,1)	141646.4	49030.61	30.87	-0.05	37949.30	29797.82	$2.4040x10^{9}$	0.12
Total vaccinations	(1,1,1)	4489830.2	1999708.8	35.52	-1.56	1747605	1624207	$3.9988x10^{12}$	0.25

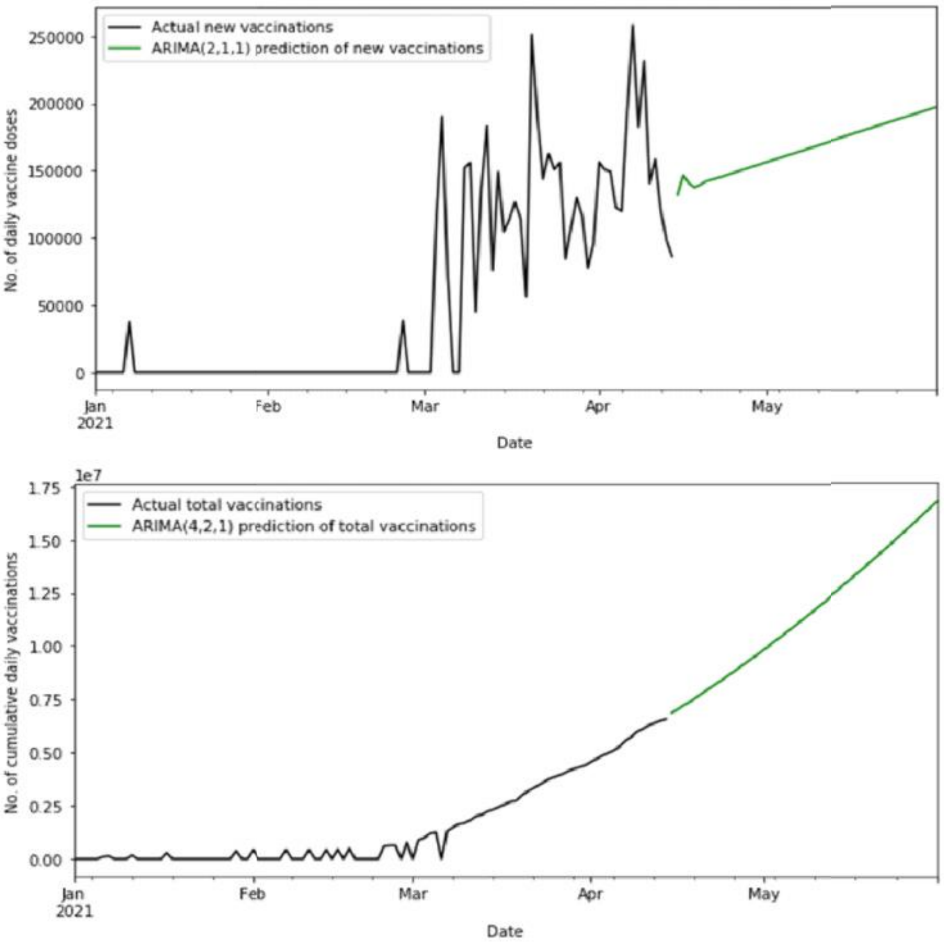

Figure 2. ARIMA prediction results of new and total vaccinations in Saudi Arabia

Table 2. Forecasting results (from April 15, 2021, to May 31, 2021)

Date	New cases	Total cases	New deaths	Total deaths	New vaccinations	Total vaccinations	Vaccination Percentage
2021-04-15	959	402,094	9	6,789	132,454	6,875,973	9.88%
2021-04-16	978	403,039	9	6,798	146,821	7,034,575	10.10%
2021-04-17	1,011	403,993	9	6,807	140,758	7,217,242	10.36%
2021-04-18	1,034	404,954	9	6,815	137,473	7,368,066	10.58%
2021-04-19	1,058	405,924	9	6,824	139,497	7,538,169	10.82%
2021-04-20	1,082	406,901	9	6,832	142,058	7,738,646	11.12%
2021-04-21	1,105	407,887	9	6,841	143,520	7,918,100	11.38%
2021-04-22	1,130	408,881	9	6,849	144,613	8,105,960	11.64%
2021-04-23	1,154	409,883	9	6,858	145,879	8,285,046	11.90%
2021-04-24	1,179	410,893	9	6,866	147,266	8,473,249	12.17%
2021-04-25	1,203	411,912	9	6,875	148,642	8,666,069	12.45%
2021-04-26	1,228	412,938	9	6,883	149,986	8,857,832	12.73%
2021-04-27	1,254	413,973	9	6,892	151,327	9,052,393	13.01%
…	…	…	…	…	…	…	…
2021-05-19	1,865	438,786	9	7,080	180,974	13,793,360	19.83%
2021-05-20	1,895	440,007	9	7,089	182,321	14,030,620	20.18%
2021-05-21	1,926	441,237	9	7,097	183,669	14,269,780	20.52%
2021-05-22	1,956	442,474	9	7,106	185,016	14,510,810	20.87%
2021-05-23	1,987	443,720	9	7,115	186,364	14,753,740	21.22%
2021-05-24	2,019	444,973	9	7,123	187,712	14,998,550	21.57%
2021-05-25	2,050	446,235	9	7,132	189,059	15,245,260	21.93%
2021-05-26	2,082	447,505	9	7,140	190,407	15,493,850	22.28%
2021-05-27	2,114	448,783	9	7,149	191,754	15,744,330	22.65%
2021-05-28	2,146	450,069	9	7,158	193,102	15,996,700	23.01%
2021-05-29	2,178	451,364	9	7,166	194,449	16,250,950	23.38%
2021-05-30	2,210	452,666	9	7,175	195,797	16,507,100	23.74%
2021-05-31	2,243	453,977	9	7,184	197,144	16,765,130	24.12%

Y_t is the target value of interest, and $e_{t-1}$is the error lag target. e_t is the error. ω is the intercept, and θ is the coefficient. For integration, stationarity is extremely important. Stationarity means that a time-series data's statistical properties are constant over time, without trend or seasonality.

Also, a shift in time does not cause a change in the distribution shape. If the dataset is not stationary, then some difference operation is applied. ARIMA models are written as ARIMA (p,d,q), which indicates the model's order. The order can be selected based on Akaike information criterion (AIC) statistics and regression analysis. The model formulation is in equation (3).

$$Y_t = \omega + \emptyset_1 Y_{t-1} + \cdots + \emptyset_p Y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t \qquad (3)$$

The forecasting process is executed using python and other essential tools to handle dataset, statistical analysis, model, and visualization.

Dataset: The complete open dataset is available and maintained by Our World In Data (39), (40). It was gathered and shared by the ministry of health in Saudi Arabia. The dataset covers 207 countries with 58 features. The data relating to Saudi Arabia is selected, and it has 408 records starting from March 2, 2020, to April 14, 2021. The data from this year is chosen, which means 104 records. After exploring and wrangling the data, the columns related to infection, mortality, and vaccination are selected. The date is assigned as the key index.

Model order selection: First, a stationarity check is done using a statistical test called the Augmented Dickey Fuller (ADF), focusing on the data's p-value to interpret the results. The null hypothesis proves that the data is not stationary, and the alternative hypothesis states that the data is stationary. The significance of the p-value will determine the stationarity. If the p-value $\geq$ 0.05, the null hypothesis is accepted, and vice versa. In the case of a non-stationarity, the difference operation is applied to minimize the significance of the p-values. After figuring the order of d, the estimation of the other two parameters, p, and q, is based on AIC. The lag order specification results are given with different p and q values to get to the model's best order. The best fit will have the smallest AIC value.

Performance: To evaluate the model performance, the following forecasting quality metrics are considered. The R squared, coefficient of determination; a percentage of variance explained by the model. The Mean Absolute Error is an interpretable metric because it has the same measurement unit as the initial series. The Median Absolute Error, an interpretable metric, particularly interesting because it is robust to outliers.

The Mean Squared Error (MSE), most commonly used, gives a higher penalty to big mistakes. The Root Mean Squared Error (RMSE) does not describe average error alone and avoids taking the absolute value, since it is undesirable in many mathematical calculations. The Mean Squared Logarithmic Error (MSLE) is practically the same as MSE, but it initially takes the series's logarithm. Thus, the attention given to small mistakes is usually used when data has exponential trends. Finally, the Mean Absolute Percentage Error (MAPE) is the same as MAE but percentage (41).

CASE 1: Infection and mortality cases in Saudi Arabia

Training and testing: The data used are the newly infected cases, the cumulative infected cases, the daily death cases, and the cumulative daily death cases. After data wrangling, the plot of the data is explored. The first observation from the data plot is the increasing trend, which means the data is not stationary. For the stationarity check, the p-value $\geq$ 0.05, hence, accepting the null hypothesis, which means non-stationary data. The p-values of the new cases, total cases, new deaths, and total deaths are 0.99, 1.0, 0.34, and 1.0, respectively.

Training the model will start by splitting the data into two sets. The training data includes the initial dataset except for the last 30 records, the test data. The model will predict the test data using the train data as the reference. The prediction data will be compared to the actual (test) data and check the model performance. For evaluation, we will consider the MAPE and the RMSE. The RMSE value should be relatively smaller than the mean value of the actual test data.

Forecasting: The initial data is fed to the model to predict the numbers. On May 31, 2021, the daily infection cases are expected to reach 2,243, and the cumulative infection cases to reach 453,977 cases. The numbers are reasonable from the testing results, and it has a slight, increasing trend that requires caution.

The model parameters are set to ARIMA (0,2,4) for daily cases and ARIMA(0,2,0) for the cumulative cases. Figure 1 depicts the forecasting result of the infection cases in Saudi Arabia. For the death cases, the number of daily deaths may reach 9 cases, and the total deaths will possibly reach 7,184 cases on May 31, 2021. The parameters are set to ARIMA (0,1,1) for new death cases and ARIMA(0,2,1) for the total death cases. The forecasting result of the death cases in Saudi Arabia is shown in Figure 2.

CASE 2: Vaccination in Saudi Arabia

Training and testing: The features to use from the dataset are new vaccinations and total vaccinations. Similarto the new vaccination data, the plot of the total vaccination data entails an evident non-stationarity. The p-values of the new vaccinations and total vaccinations are equal to 0.60 and1.0, respectively.

First, the data is divided into training and testing sets. Using the best model order for training, we use the training data to predict the test data. The estimated testing data are slightly different from the actual data due to the data availability and the vaccine data introduced later in the time series. Therefore, having excess zerosaffected the results since it is not enough to train the model. However, the statistics summary of the test data has enough variation to form a reasonable prediction.

Forecasting: For the daily vaccinations, the numbers might reach 197, 144 doses on May 31, 2021. The numbers for cumulative vaccinations may reach 16,765,130 doses. The model parameters are set to ARIMA (2,1,1) for daily vaccinations and ARIMA(4,2,1) for the cumulative vaccinations. The results are more reasonable considering the training phase due to the sufficient data from the last few months. Figure 3 illustrates the results of the forecasting for both features.

RESULTS AND DISCUSSION

From last year and until now, the situation of the COVID-19 epidemic is still alarmingly present in Saudi Arabia, where authorities have taken precautionary measures that included the temporary suspension of Umrah, travel restrictions, regional lockdowns, and curfews that eventually were lifted on June 21, 2020. The mask-wearing is still mandatory and the limitation on social gatherings. After October, the cases in the Kingdom had a noticeable drop in numbers. Earlier this year, there was a spike in the reported cases. The government had to enforce new rules and go on a lockdown for 30 days, to limit social gatherings to 20 people and close indoor entertainment centers, gyms, and restaurants. On April8, 2021, the number of new cases in Saudi Arabia has passed the 900 cases mark for the first time since August. The vaccine doses have been administered across the Kingdom and people registered to receive the vaccine to maintain disease spread. This instability in numbers proves the uniqueness and significance of the virus and how it is developing. As presented through this work, some statistical analysis and models, such as ARIMA, enable us to investigate, monitor, forecast, and analyze the virus's epidemiology. In the presented case study, the model performs to predict the COVID-19 numbers for the next two months. It explores two views; the numbers of infection and mortality and the numbers of vaccination.

From the latest data, Saudi Arabia has reported a daily of 929 and a total of 401,157 cases on April 14, 2021. For the death cases, a daily of 8 and a total of 6,781cases are reported. The actual data is given to the ARIMA model to predict the numbers of the next month. At this rate, the infection cases might reach 2,243new cases and 453,977cumulative cases by May 31, 2021. Also, the number of mortalities may reach 9 of daily deaths and 7,184of total deaths. The model performed well enough to get sensible results. However, the actual results can differ due to multiple external factors. These factors include vaccine availability, precautionary measures, social gatherings, and special events. For instance, Ramadan and Eid are approaching, which entails gathering at mosques and performing Umrah. It can impact the rate of cases and deaths for the prediction. The latest data indicate that the estimated number of Saudi Arabia's population exceeds 34813867 people. The dataset shows that Saudi Arabia has administered around 6,577,743 doses of vaccines so far. The ARIMA model expectation for the total vaccination numbers is exponentially increasing to 197,144 daily and 16,765,130 doses by May 31, 2021. It actively demonstrates that 24.12% of the entire population will get fully vaccinated with two doses for each person. At this rate, the country can reach safety when 70% of people are fully vaccinated, attaining herd immunity, which is expected by November, 2021. There are no sufficient records of vaccination data to train the model correctly, and the excess zeros in the dataset might affect the model performance and the results. There are some external factors to be put into consideration when looking at the correctness of the forecasted numbers. First, vaccine availability must continue at the same rate in the future, and the second dose must be provided as scheduled. Second, the people's culture and background can impact the numbers, especially if they refuse to get the vaccine. Third, the political and economic aspects of the situation play a role in the vaccination rate. Fourth, the virus mutation and whether the vaccine will be of the same efficacy Finally, the validity and the expiry of the vaccine.

CONCLUSION

This work explores various machine learning uses in healthcare practices during the COVID-19 outbreak. The paper demonstrates a thorough literature review on healthcare, machine learning, and methods used for monitoring, preventing, and managing the current pandemic. One of the most common statistical models for time series forecasting is the ARIMA model. This paper provides a case study to forecast the numbers of infections, mortalities, and vaccinations in Saudi Arabia using the ARIMA model. To sum up everything, the reported cases and death cases have a slightly increasing trend. The vaccination numbers are increasing exponentially. At this rate, it is expected that over 70% of Saudi Arabia will be fully vaccinated by the end of this year. This paper intended to investigate the COVID-19 situation, forecast the numbers, and provide insights to help alarm stakeholders and healthcare decision-makers. The paper discusses some major external factors that might affect the trend's rate.

REFERENCES

(1) Hippisley-Cox, J. J. Fenty, and M. Heaps, 2007. “Trends in Consultation Rates in General Practice 1995 to 2006: Analysis of the QRESEARCH database,” *QRESEARCH Inf. Cent. Heal. Soc. care*, Jan.

(2) Weiss S. R.and J. L. Leibowitz, 2011. “Coronavirus pathogenesis,” *Adv. Virus Res.*, vol. 81, pp. 85–164, doi: 10.1016/B978-0-12-385885-6.00009-2.

(3) WHO Team, “Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19),” *World health orgnization (WHO)*, 2020. https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19) (accessed Mar. 31, 2021).

(4) Weiland N. and Cochrane, E. 2021. “Government Eyes War Powers to Speed Medical Manufacturing Ahead of Virus,” *New York times*, 2020. https://www.nytimes.com/2020/02/28/us/politics/trump-coronavirus.html (accessed Mar. 31).

(5) Huang C. *et al.*, 2020. “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,” *Lancet*, vol. 395, Jan. doi: 10.1016/S0140-6736(20)30183-5.

(6) WHO Team, “Situation Report - 44 Coronavirus disease 2019 (COVID-19),” *World health orgnization (WHO)*, 2020. https://www.who.int/publications/m/item/situation-report---44 (accessed Mar. 31, 2021).

(7) Organization, W. H. 2020. “Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: interim guidance, 28 January,” World Health Organization, Geneva PP - Geneva. (Online). Available: https://apps.who.int/iris/handle/10665/330893.

(8) Tabrez Uz Zaman and A. M. Adetunji Hamed Ademola, Mohammed Shamim, Salih El Fadil Mohammed, Pasha Tariq Sultan, “A comprehensive review of COVID-19 pandemic and community mitigation strategies,” *Med Heal. Sci*, vol. 8, no. 2, 2020.

(9) S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Upper Saddle River, New Jersey: Prentice Hall, 2003.

(10) Foresee medical team, "Benefits of Machine Learning in Healthcare," *Foresee Medical*, 2020. https://www.foreseemed.com/blog/machine-learning-in-healthcare (accessed Mar. 31, 2021).

(11) Rayan, Z., Alfonse, M. and Salem, A. B. M. 2018. "Machine Learning Approaches in Smart Health," *Procedia Comput. Sci.*, vol. 154, no. 1985, pp. 361–368, doi: 10.1016/j.procs.2019.06.052.

(12) Vaishya, R. M. Javaid, I. H. Khan, and Haleem, A. 2020. "Artificial Intelligence (AI) applications for COVID-19 pandemic," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 4, pp. 337–339, doi: 10.1016/j.dsx.2020.04.012.

(13) Doupe, P. Faghmous, J. and S. Basu, 2019. "Machine Learning for Health Services Researchers," *Value Heal.*, vol. 22, no. 7, pp. 808–815, doi: 10.1016/j.jval. 2019.02.012.

(14) Garg A. and V. Mago, 2021. "Role of machine learning in medical research: A survey," *Comput. Sci. Rev.*, vol. 40, p. 100370, 2021, doi: 10.1016/j.cosrev.100370.

(15) K. R. Bisaso, G. T. Anguzu, S. A. Karungi, A. Kiragga, and Castelnuovo, B. 2017. "A survey of machine learning applications in HIV clinical research and care," *Comput. Biol. Med.*, vol. 91, no. February, pp. 366–371, 2017, doi: 10.1016/j.compbiomed..11.001.

(16) Agrawal, A. J. Gans, and A. Goldfarb, 2016. "The Simple Economics of Machine Intelligence," *Harvard Business Review*, https://hbr.org/2016/11/the-simple-economics-of-machine-intelligence (accessed Mar. 30, 2021).

(17) Ting, D. S. W. L. Carin, V. Dzau, and Wong, T. Y. 2020. "Digital technology and COVID-19," *Nat. Med.*, vol. 26, no. 4, pp. 459–461, 2020, doi: 10.1038/s41591-020-0824-5.

(18) Chen, M. S. Mao, and Y. Liu, 2014. "Big Data: A Survey," *Mob. Networks Appl.*, vol. 19, no. 2, pp. 171–209, doi: 10.1007/s11036-013-0489-0.

(19) Chen, M. Y. Hao, K. Hwang, L. Wang, and L. Wang, 2017. "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," *IEEE Access*, vol. 5, pp. 8869–8879, doi: 10.1109/ACCESS.2017.2694446.

(20) Makridakis, S. E. Spiliotis, and V. Assimakopoulos, 2018. "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS One*, vol. 13, no. 3, p. e0194889, Mar. 2018, (Online). Available: https://doi.org/10.1371/journal.pone.0194889.

(21) Bontempi, G. S. Ben Taieb, and Y.A. 2013. Le Borgne, "Machine Learning Strategies for Time Series Forecasting," in *Lecture Notes in Business Information Processing*, vol. 138.

(22) F. Rustam *et al.*, 2020. "COVID-19 Future Forecasting Using Supervised Machine Learning Models," *IEEE Access*, vol. 8, pp. 101489–101499, doi: 10.1109/ ACCESS.2020.2997311.

(23) Mojjada, R. K. A. Yadav, A. V. Prabhu, and Y. Natarajan, 2020. "Machine learning models for covid-19 future forecasting," *Mater. Today Proc.*, pp. 1–5, doi: 10.1016/j.matpr.2020.10.962.

(24) Mahalle, P. A. B. Kalamkar, N. Dey, J. Chaki, A. ella Hassanien, and G. R. Shinde, "Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art," 2020, doi: 10.36227/techrxiv.12101547.v1.

(25) Alzahrani, S. I. I. A. Aljamaan, and E. A. Al-Fakih, "Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions," *J. Infect. Public Health*, vol. 13, no. 7, pp. 914–919, 2020, doi: https://doi.org/10.1016/j.jiph.2020.06.001.

(26) Fong, S. J. G. Li, N. Dey, R. Gonzalez-Crespo, and E. Herrera-Viedma, "Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 6, no. 1, p. 132, 2020, doi: 10.9781/ijimai.2020.02.002.

(27) Fong, S. J. G. Li, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction," *Appl. Soft Comput.*, vol. 93, p. 106282, 2020, doi: https://doi.org/10.1016/j.asoc.2020.106282.

(28) Batista, M. *Estimation of the final size of the second phase of the coronavirus epidemic by the logistic model*. 2020.

(29) Hu, Z. Q. Ge, S. Li, L. Jin, and M. Xiong, "Evaluating the effect of public health intervention on the global-wide spread trajectory of Covid-19," *medRxiv*, p. 2020.03.11.20033639, Jan. 2020, doi: 10.1101/2020.03.11.20033639.

(30) L. Jia, K. Li, Y. Jiang, X. Guo, and T. zhao, "Prediction and analysis of Coronavirus Disease 2019." 2020.

(31) K. P. S. S. Hembram and J. Kumar, "Epidemiological study of novel coronavirus (COVID-19): macroscopic and microscopic analysis," *Int. J. Community Med. Public Heal.*, vol. 8, no. 3, Feb. 2021, (Online). Available: https://www.ijcmph.com/index.php/ijcmph/article/view/7632.

(32) D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the ARIMA model on the COVID-2019 epidemic dataset," *Data Br.*, vol. 29, p. 105340, 2020, doi: https://doi.org/10.1016/j.dib.2020.105340.

(33) Y. Zhang, B. Jiang, J. Yuan, and Y. Tao, "The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study," *medRxiv*, p. 2020.03.04.20031187, Jan. 2020, doi: 10.1101/2020.03.04.20031187.

(34) P. Kumar *et al.*, "Forecasting the dynamics of COVID-19 Pandemic in Top 15 countries in April 2020: ARIMA Model with Machine Learning Approach," *medRxiv*, p. 2020.03.30.20046227, Jan. 2020, doi: 10.1101/2020.03.30.20046227.

(35) J. Chen, "Autoregressive Integrated Moving Average (ARIMA)," *investopedia*, 2019. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp (accessed Mar. 31, 2021).

(36) M. Yousaf, S. Zahir, M. Riaz, S. M. Hussain, and K. Shah, "Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan," *Chaos, Solitons & Fractals*, vol. 138, p. 109926, 2020, doi: https://doi.org/10.1016/j.chaos.2020.109926.

(37) Z. Ceylan, "Estimation of COVID-19 prevalence in Italy, Spain, and France," *Sci. Total Environ.*, vol. 729, p. 138817, 2020, doi: https://doi.org/10.1016/j.scitotenv.2020.138817.

(38) A. K. Sahai, N. Rath, V. Sood, and M. P. Singh, "ARIMA modelling & forecasting of COVID-19 in top five affected countries," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, pp. 1419–1427, 2020, doi: https://doi.org/10.1016/j.dsx.2020.07.042.

(39) Owidbot, “Data on COVID-19 (coronavirus) by Our World in Data,” *github*, 2020. https://github.com/owid/covid-19-data/tree/master/public/data/ (accessed Mar. 20, 2021).
(40) Max Roser Hannah Ritchie E. O.-O. and J. Hasell, “Coronavirus Pandemic (COVID-19),” *Our World Data*, 2020.
(41) Advaitsave, “Introduction-to-Time-Series-forecasting-Python,” *github*, 2018. https://github.com/ advaitsave/ Introduction-to-Time-Series-forecasting-Python (accessed Mar. 26, 2021).

****