ISSN: 2708-3659

AL-Rayan Journal

of Humanities & Applied Sciences

Scientific Biannual & Refereed

جامعة الريان
AL-RAYAN UNIVERSITY

Volume 6
Issue 2
Serial No.11

December 2023

Comparison of the predictive accuracy of hybrid models of NNAR, TBATS, ETS-EANN, and ARIMA for COVID-19 incidences in Saudi Arabia

Ahmad Ghaleb Al-Rabbasi [(1)]; Faisal A.M Ali [(1)];
Fahmy Al–Salwi [(1)]; Ahmed Abdel -Razzaq Alansari [(1)]

[(1)] Department of Data Science and Information Technology, University of Taiz, College of Administrative Sciences, Taiz, Yemen.

Abstract:

COVID-19 is a strain of the coronavirus that causes sickness. It first showed up in December 2019 and caused a world pandemic that is still going on. There is a vital need to monitor the spread of COVID-19 and to be able to predict where the pandemic will go so as to control its spread in a more effective way. COVID-19 data is an example of time-series data that can be used in different ways to make predictions. Even though there are many time-series forecasting models, it is hard to make broad general statements about which ones are better. This paper gives an empirical review of several time-series models for predicting COVID-19 cases, in Saudi Arabia, where Arab countries like Arab Gulf countries, lack studies in diagnosing this aspect. In particular, this research contributes to the literature on Time-Series Data forecasting in a variety of ways through outcomes of a variety of forecasting strategies using series forecasting techniques. These include hybrid combinations of an exponential smoothing state-space model ETS, EANN, autoregressive -integrated moving-average (ARIMA), a neural network autoregressive model (NNAR), and TBATS. Thus, they are eleven top combination architectures of the: (E-N, E-N-F, N-F, N-T, A-N, A-N-F, N-T-F, A-E-N, E-N-T, A-N-T, and A-N-E-F). The models were made with daily COVID-19 data from Saudi Arabia from Jan 2022 to May 30, 2022, which was available to the public. The best forecasting model was selected based on the high accuracy measure of root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE). The testing results show that based on the current data findings of this investigation, the E-N, E-N-F, and N-F models had the lowest forecasting error.

Keywords: Hybrid models forecasting; ARIMA; ETS; TBATS NNAR F, COVID-19

1. Introduction

Coronavirus illness (COVID-19) is a worldwide outbreak that has influenced every country on the planet. COVID-19 is a novel coronavirus that transmits fast from person to person, resulting in a big epidemic and major tragedy. COVID-19 is a zoonotic coronavirus (SARS-CoV) that reasons the disease (severe-acute-respiratory-syndrome). The virus is likely to have started in the Chinese city of Wuhan, with the first fatal cases being reported in late 2019. This virus is very deadly, particularly to the elderly and those suffering from chronic conditions[1]. The structure of the disease is highly fluid, and it spreads swiftly. Unfortunately, there had been 123,010 verified deaths and approximately two million confirmed cases worldwide as of April 15, 2020.

The paucity of historical data to help scientists analyze and predict the disease's impact is a big concern. Public health planning requires COVID-19 prediction. Accurately calculating active cases is one method to do this. Time-series data are quantitative data measured at similar time intervals (e.g., day, hour, or per minute). Stock markets and scientific, medicinal, and natural discoveries produce time-series results [2]. Time-series models in statistics and machine learning can be used to estimate the spread of illnesses by extrapolating from known incidences[3]. Malaria [4], influenza [5] [6], TB [7], and other infectious illnesses[8] were just a few of the diseases whose future dynamics were predicted using these methods. In the United States [9], Italy [9], India [10], and a few other places [11], time-series models have recently been employed to foretell the behavior of COVID-19. Several time-series models are employed in this study to project the number of new infections in Saudi Arabia. Daily COVID-19 infection rates were modeled using data that was made public from January 2022 through May 30 of the same year.

Due to disparities between countries' epidemiological surveillance and detection capacities, the number of confirmed cases fluctuates. Thus, modeling daily approved cases and estimating future potential new cases is critical for health system demand management. Mathematical and statistical modeling methods are used to estimate short- and long-term epidemic materials and resources. Furthermore, such estimations can guide the magnitude and type of treatments required to end the outbreak [12].

Many studies utilized various models to predict the occurrence and prevalence of COVID-19. There are many studies conducted in developed countries. For instance, USA, UK, Italy, Spain, France, China, and India [13]–[15]. While Arab countries, like Arab Gulf countries, lack studies diagnosing this aspect [16]. In addition, the spread of COVID-19 will have an impact on Economic Growth in Saudi Arabia, low global demand for crude oil, and the cost and burden of the precautionary and preventive measures to limit the spread of COVID-19. COVID-19 has pushed scientists and researchers all over the world to concentrate on it. Studies on its forecasting are very rare in the short-term literature [17]. Thus, in this research, statistical models that are utilized to forecast can play an important role in informing the disease's future trend. This research applied five hybrid models: NNAR, EANN, ETS, ARFIMA, and TBATS, hybrid combinations of these models. This research found the best model forecasting for the number of daily COVID-19 cases in Saudi Arabia for the next 43 days. The absolute number of notes was the only data available in this investigation. Other factors were not evaluated since they were unavailable. The remainder of this paper is organized as follows. In the second section, this study reviewed the relevant literature. In section 3, this study outlined the empirical strategy and presented the data used in the analysis. The key findings and policy implications are presented in section 4. Finally, in section 5, this study offered some concluding comments.

2. Related literature

Several linear classical approaches to time series forecasting have been expanded in the literature[12]. ARIMA is a popular linear time series model. The Box-Jenkins method and statistical properties make the ARIMA model popular [18] model-building process. Although the ARIMA model offers the benefits of great forecasting for a short time period and ease of implementation, it does have one fundamental flaw: the model's presumed linear form. ARIMA models can also use a variety of exponential techniques smoothing (ETS)[19].

One of the most prominent nonlinear intelligence models is neural network auto regression (NNAR) [20]. The shape of the model is no longer necessary in artificial neural networks. Furthermore, the model is built in an adaptive manner based on the qualities of the data [19]. These models have been used in a variety of studies to construct hybrid models due to their nonlinear modeling capabilities; exponential techniques smoothing (ETS) is statistical foundations for time series forecasting. As a result, a variety of models are based on it. Knowing the future values of a time series can be useful in a variety of situations because it allows decisions to be taken ahead of time and are accurately modified to get better results. Many aspects, such as cost and consumption savings, emissions reduction, and logistics optimization, can be considerably improved if the estimates are precise and reliable. Bates and Granger,[21] Newbold and Granger [22], and Winkler and Markakis [23] were among the first to attempt to construct parallel hybrid models for forecasting. Clemen [24] conducted a thorough evaluation of preliminary studies that employed hybrid parallel models and classified them according to their contribution to various fields. In their famous work, Bates and Granger [24] devised the parallel hybrid structure as the most advanced combination strategy, by linearly merging numerous predictions, they produced a parallel hybrid model that reduces error variance more than any of the previous models. When raw data is supplied to many models concurrently, the parallel hybrid structure method for integrating prediction models may be conceptualized. In the second step, a weighting algorithm calculates each prediction's component’s weight. In recent years, hybrid models have been presented in an attempt to achieve more accurate forecasting predictions. Pai & Hong [25] suggested a hybrid model for stock price prediction that utilizes the advantages of both the auto regressive –integrated-moving-average (ARIMA), and support vector machines (SVMs). Chen [26] developed a seasonal time series forecasting model that included a seasonal-(SARIMA) model with SVMs. Khashei [27] suggested a new hybrid model based on the fundamental concepts of synthetic neural networks and fuzzy regression deliver more accurate predicting results. To anticipate traffic flow, Zeng [28] suggested a hybrid ARIMA/ANN model. Aladag [29] built a hybrid model that combined Elman's frequent-neural-networks with ARIMA models. Pham [2] proposed a hybrid model that combined generalized-auto-regressive-covariance (GARC) and ARIMA models to forecast machine health. For forecasting the Tehran stock index, [30] presented a hybrid model incorporating fuzzy systems and ANNs. Kristjanpoller [31] proposed a hybrid model for gold price forecasting that combined GARCH and ANN models. For time series forecasting, Xiong [32] suggested a hybrid modeling approach that combines multi-output-support-vector-regression (MSVR)and interval Holt's exponential smoothing method. Many hybrid models combining (ARIMA & ANN) models have been described in the literature of time series prediction in recent years, in addition to utilizing the distinct advantages of both models in linear and nonlinear analysis respectively and addressing their limitations. To estimate short-term wind power, a novel least-square-support-vector-machine (LSSVM) and auto-regressive fractionally-integrated-moving-average (ARFIMA) model were proposed by Yuan [33], combined, created and employed a hybrid model based on ARIMA and radial-basis-function-neural-networks (RBFNs) to anticipate power prices. In addition, since the beginning of 2020, a growing corpus of literature has used multiple approaches to forecast the spread of the COVID-2019 pandemic [34]–[37]. ARIMA models [38], ETS models [39], [40], neural network (NN) models [41], TBATS models [42], models that have been derived from the (susceptible–infected–removed: SIR) and hybrid models were the most commonly employed [43], [44]. Ala'raj [45] used a lusty hybrid model with ARIMA residual corrections based on a modified (susceptible-infected-recovered-dead-SIRD) model. Using a US COVID-19 dataset, they gave long-term predictions for persons who were infected, recovered, died., and their model had a surprising capacity to generate correct predictions [46]. They used a no seasonal ARIMA model to forecast confirmed cases afar 15-April, 2020, for Spain, Italy, and France. The predictions had minimal (MAPE) and appropriate to short-term for investigation epidemiological of COVID-19 trends. Hasan [43] created model that integrates (ensemble-empirical-mode-decomposition: EEMD) and neural networks. The investigation revealed that the ANN-EEMD strategy outperformed established statistical approaches such as moving average and regression analysis to estimate real-time worldwide COVID-19 cases beyond 18-May 2020.

Ribeiro, [47] used different techniques and models to offer short-term estimates of new cases of COVID-19 in Brazil, including ARIMA, random forest RF, and support-vector-regression (SVR). The reliability of the modules was assessed using the (MAPE and MAE) standard. The investigation revealed that SVR is the best, while other models did well in terms of forecasting. To predict new cases of COVID-19 reported cases in India Sardar [42]. The researchers employed TBATS and ARIMA, their hybrid was used in five different states (Delhi, Maharashtra, Gujarat, Tamil Nadu, and Punjab) from 30-Apr, 2020 to 31-May, 2020. The hybrid approach exhibited the best forecasting abilities. Talkhi [48] researchers employed different single and hybrid models to forecast the number of COVID-19 cases in Iran. From 15-August, 2020, to 14-September 2020. In terms of forecasting confirmed cases, the MLP and Holt-Winter HW models were the most accurate.

Table 1 Summarizes the literature on hybrid models including ARIMA TBATS, EANN, EST, and ARNN models.

Table 1 Shows 20 studies from around the world that used hybrid statistical ARIMA, EANN TBATS, ETS, and ANN models.

Authors	Data	Method	Country/region
Alaraj et al.[45]	Recovered, dead and confirmed	ARIMA SEIRD	US
Chakraborty et al [49]	Proved	ARIMA , WBF	Canada-India and
Joseph et al.[39]	proved	ETS, ARIMA , INGARCH and	10 Countries
Fantazzini,[50]	Proved	HVAR, SIR, VAR ARIMA, ARIMAX, and ETS	158 countries
Melin et al [41]	Confirmed	ME-NNAR	Mexico
Alsunaidi et al.[51]	Recovered Active and confirmed	SARIMA,ARIMA, KF and HW	Pakistan
Cao et al. [40]	Proved	SEIQDR ETS, ARIMA and	China
Hasan, N.[40]	Proved	EEMD- ANN	(aggregate)
Kırba et,al.[52]	Proved	ARIMA, LSTM and NNNAR.	Eight Countries in
Yonar, et,al [53]	Proved	LES & ARIMA	7 Countries
Dhamodharavadhani [54].	Deceased	GRNN, NNNAR ,RBFNN, and PNN	India
Papastefanopoulos,[55]	Confirmed	ARIMA, N-Beats, DeepAR, and FB, HWAAS	10 countries
Bhandary et al.,[56])	Recovered, & deceased Confirmed	ARIMA, SutteARIMA, ETS, & SIR	Nepal
Perone,[57]	Recovered, & deceased Confirmed	ARIMA, NNAR , ETS, TBATS and hybrid models	Italy
Salgotra & Gandomi[58])	Recovered, & deceased Confirmed	genetic programming,NNAR	Australia
Mohan, et.al,[59])	Recovered, & deceased Confirmed	ARIMA, NLP,	India
Jin, Dong, Yu, & Luo,[60])	Recovered, & deceased Confirmed	ARIMA ,ANFIS, N-BEATS, and WT-RVFL, models	UK, India, and the US
ArunKumar, et. al, [61])	confirmed, recovered, and deaths	(RNN, GRU and LSTM, ARIMA and SARIMA) to	Comber countries
Kalantari,[62])	confirmed, recovered, and deaths	ARIMA, NNAR.	10 countries
Atchadé, et al,[63])	confirmed, recovered, and deaths	ARIMA,ETS	Comber countries

3. Materials and Methods

The purpose of this study is to compare the hybrid models to predict future COVID-19 conduct. The COVID-19 dataset was used, which included the total number of reported, new cases of COVID-19 in Saudi Arabia. The data could be found at (https://github.com/owid/covid-19-data). These stats were reported on a daily basis on the website from Jan 2022 to May 30, 2022. The R software was used for all data analysis. The aim of the data analysis according to the hybrid models that were identified in this study was to find a model for predicting the number of daily new COVID-19 cases in the future. In this part, the models which were used are briefly described as follows:

3.1 ARIMA model:

Box & Cox, (1964) proposed the Box-Jenkins technique, ARIMA models. This method includes non- stable time series that are rendered stable through differencing. One of the most widely utilized and well-known ways for predicting time series is the ARIMA [55]. In ARIMA models, time series linear correlation is investigated. Auto-regressive (AR) and moving-average (MA) models, white noise process are all included in these models. The unit root tests could be utilized to determine the sequence of differencing, whereas the MLE and AICc methods can be used to determine the AR and MA processes' optimal parameters [55].

The "auto.arima" function, created by Hyndman (2008)[64] and included in the package "forecast," is used to recognize ARIMA models (in R-environment). To determine the best ARIMA models, this function uses the number of (p) parameters of the (AR), the number of (q) parameters average (MA) and the order I of differencing (I). It includes unit root tests, as well as other tests. The bias-corrected Akaike's information criterion (AICc) was also reduced, and maximum likelihood estimation was improved (MLE).

Finally, the are four typical prediction accuracy measures are used to assess the over-all goodness of fit: (MAE), (MAPE), and (RMSE). The following is the ARIMA model's approximated basic equation.

$$\mathrm{y}_t = \phi_1\, \mathrm{y}_{t-1} + \phi_2\, \mathrm{y}_{t-2} + \ldots\ldots.\, \phi_p + \mathrm{y}_{t\text{-}p} + \mathrm{y}_1\varepsilon_{t-1} + \ldots\ldots\, y_q\varepsilon_{t-q} + \varepsilon_t \qquad (1)$$

Where (p, q *and* εt) refers to (AR), (MA), and the residuals of errors at time t. respectively. The (ϕ) represents the coefficient of each parameter p, and (y_t) indicates the expected values.

3.2 ETS model

The "ETS" function in the package "forecast" in R environment, developed by Hyndman et al., (2008) [64].It is used for identifying ETS models. In the ETS basic models, the two main equations are the smoothing and forecast . The observation/measurement equation is obtained by inserting the last two equations into an invention state space model [1]. The first equation can be used to describe observable data, whereas the second equation can be used to describe unobserved state behavior. The states are the level, trend, and seasonality of the data. We utilize the AICc metric in particular to select the optimal ETS model. The RMSE

MAPE and MAE, measures are used to determine the quality of fit. The following are the estimated equations for the basic ETS-(A,N,N) model with additional error [65].

$$\hat{Y}_{(t+1/t)} = l_t \qquad (2)$$

$$l_t = l_{(t-1)} + \alpha \times [y_t - l_{t-1}] \qquad (3)$$

Where (l_t) is the current estimate for the level, $[y_t - l_{t-1}]$, it represents each time a prediction that is one step ahead. $t+1$ as a result of a weighted average of all the data collected, $0 \le \alpha \le 1$ is the smoothing parameter that regulates the rate of weight drop, and $y_t - l_{t-1}$ is the error at time (*t*). As a result, each predicted observation is the product of the previous level plus an error. There is a specific probability distribution for each form of error, whether additive or multiplicative. It is expected that additive model errors, such as this one, follows a normal distribution. As a result, the equations (2) and (3) can be rearranged as follows:

Equation of observation: $y_t = l_{t-1} + \varepsilon_t$ (4)

Transition equation: $l_t = l_{t-1} + \alpha\varepsilon_t$ (5)

State space theories underpin exponential smoothing. Approaches are represented by equations (4) and (5).

3.3 Neural Networks of Auto Regression Model (NNAR)

NNAR mechanism learning challenges employ statistical models. The (NNAR) is a type of neural network that is also a parametric non-linear model that is used to solve forecasting difficulties. The " NNAR " function in the package "caret" (R environment), written by Hyndman, et al 2008 [64], and it is used to identify NNAR models. We may write NNAR models as *(p,k)* for non-seasonal data, where *p* represents the total number of non-seasonal delays that were used as inputs and (*k*) signifies the number of hidden layer nodes/neurons. The NNAR *(p,k)* The procedure is analogous to the AR process, but using nonlinear functions. The *AICc* measure is used to calculate the optimum number of non-seasonal delays, and calculating $(p+p+1)/2$ yields the ideal number of neurons, where *(P)* is the seasonal AR order (if any) and (*P*) is the non-seasonal AR order. Finally, utilizing the RMSE, MAPE, and MAE measures, the quality of fit is examined. The NNAR equation in its most basic form is as follows[65]:

$$y_t = f(y_{t-1}) + \varepsilon_t \qquad (6)$$

Where y_t pointing to the foresee values , $y_{t-1} = (y_{t-1}, y_{t-1}, ...y_{t-n})$ is a vector that contains the observed data's lagged values, f is an N-layer

AL-Rayan Journal of Humanities & Applied Sciences - ISSN: 2708-3659
Volume (6) ، Issue (2) December. 2023، Serial No: (11)

single-layer neural network (n) a single layer of hidden neurons and ε_t is the blunder in time (t).

3.4 TBATS model

Transform, Box-Cox, ARMA Trend errors, and Seasonal components are the five elements that make up the acronym BATS. $\omega, \phi, p, q, m_1.,......m_n$ is included to present the Box-Cox, damping, ARMA (q ,p), and Seasonal periods $t_1.,......t_n$ [46]. This model is an extension of classic seasonal models [46], which have many seasonal periods. As described by De-Livera, Hyndman, and Snyder [66] , the "tbats ()" was used by models in particular. The package "predict" includes this method. (in the R environment). Using the measure Akaike's information criterion (*AIC*), the optimum parameter for the Box–Cox transformation, ARMA *(q, p)* and the rank dampening parameter are selected by 10 Fourier terms. The TBATS model's basic equation was written as follows [66]:

$$y_t^{\omega} = l_{t-1} + \phi b_{t-1} + \sum_{I=1}^{T} s_{t-mi}^{i} + d_t. \qquad (7)$$

Where y_t^{ω} denotes the parameter for Box–Cox transform (ω) implemented to the (y_t) at time (t), ϕ is the tendency that has slowed, and (l_t) is level of the area. A long-term trend is denoted by b, whereas a seasonal pattern is denoted by T, and the ($_i th$) component is denoted by s_t^i, m_i indicates the periods seasonal, and d_t denotes an ARMA *(p,q)* residuals procedure.

3.5 Hybrid model

The "hybridModel" action in the package "forecastHybrid" (in R environment) advanced by Shaub and Ellis is used to identify hybrid models. To combine the single time series forecasting approaches, we employed the following procedure: i) First, we use the Box-Cox (1964) power alteration to make the normality supposition more plausible; ii) Next, we use cross validation-errors ("cv.errors") to give greater weight to models that produce the best forecast and perform comparably better; and iii) finally, MAPE,., MAE, and RMSE are utilized to validate the forecast accuracy measures.

3.6 Evaluation of the metrics:

During the rounds of training and testing, we used four performance metrics to assess the goodness of fit of the methods used in this study: (MAE), (RMSE) and (MAPE) [67], [68]. The following are the definitions of these metrics:

$$MAE = 1/n \sum_{i=1}^{n} |y_i - \hat{y}_i| \qquad (8)$$

$$MAPE = 1/n\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}\times 100\% \quad (9)$$

$$RMSE = \sqrt{1/n\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (10)$$

Where (y) represents the anticipated values and (yi) represents the actual values, (n) represents the total number of observations. Examples of such metrics are mean absolute error (MAE) and root mean squared error (RMSE), both of its dependent measures of different kinds of error. MAE is easy to minimize because lowering it results in median projections, whereas minimizing RMSE leads to mean predictions. In fact, whereas the first metric relies on absolute mistakes, the second relies on squared errors. Because it is based on percentage errors, MAPE is not scale-dependent like MAE and RMSE [67], [68].

As a result, it has the advantage of being a unitless metric. It does, however, necessitate certain key discernment as consequence, in skewed forecasts since it returns infinite, and it penalizes negative mistakes more severely than positive errors. Finally, De-Livera, Hyndman, and Snyder [66] provide a scale-free error metric that is likely the greatest adaptable and trustworthy forecasting precision measurement. It outperforms MAPE in that it does not produce endless and may measure prediction accuracy across time series. Because each model has distinct advantages and weaknesses, this study took the prudent approach of examining the output of all of them.

4. Results and data collecting

The aim of this study was to make a comparison of hybrid models to predict future COVID-19 conduct. The data utilized in this paper includes daily COVID-19 data from Saudi Arabia from Jan 2022 to May 30, 2022. The R software was used for all data analysis. The aim of the data analysis according to the hybrid models that were identified in this study was to find a model for predicting the number of daily new COVID-19 cases in the future. Figures 1 and 2 show the daily trend of new COVID-19 cases in Saudi Arabia from 1 January 2022 to May 30, 2022. The performance of the approaches was assessed using a testing and training dataset. The first 70% of data is utilized for training, and the remaining 30% is used to test the models. After that, the models' forecasting quality is assessed using three metrics: MAE, MAPE, and RMSE.

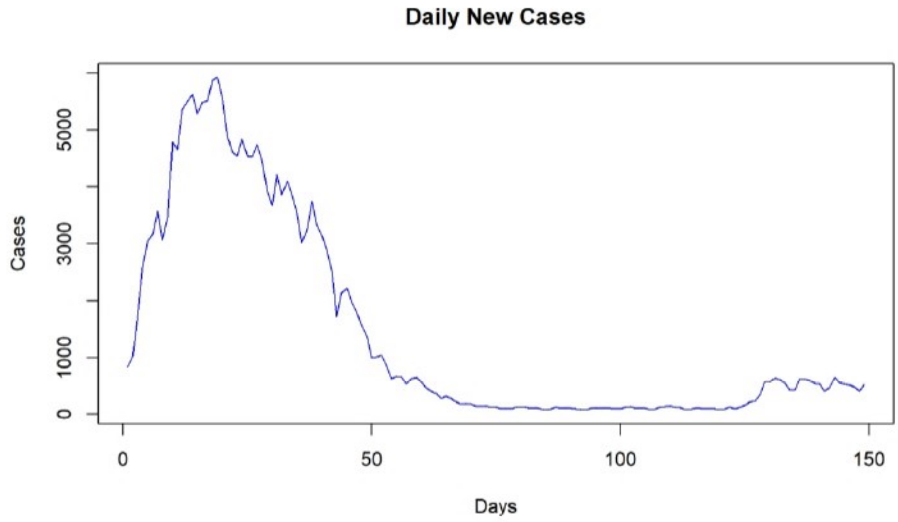


Figure 1. New COVID-19 cases in Saudi, from Jan 01,2022 to May 30, 2022.

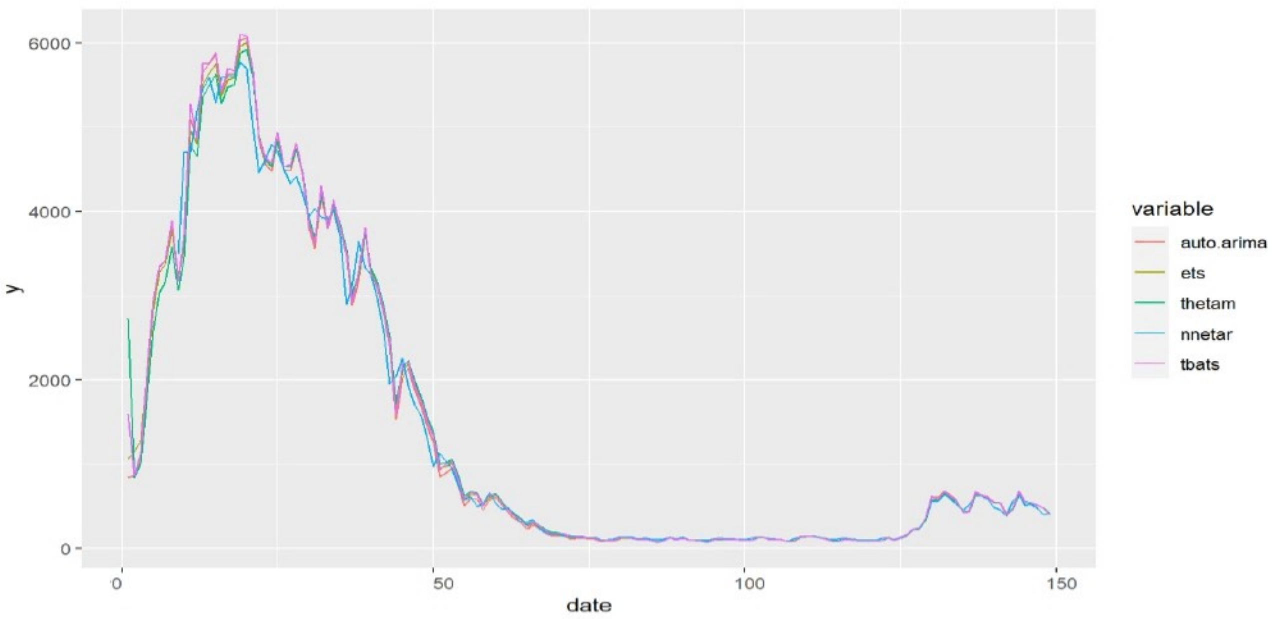


Fig. 2 New COVID-19 Cases in Saudi, from Jan 01,2022 to May 30, 2022.

In the training phase, we trained the five models by training data for numbers of daily new COVID-19 cases, separately. Table 2 shows the best-chosen parameters for the single models[1], while Table 3 shows the forecasting measurement scales for single and hybrid models[2].

[1] (Appendix A) show the values of parameter for the ARIMA, EANN = F, TBATS and ETS models.

[2] Because the time series lasted less than a year and comprised daily data, the frequency was adjusted to 7, which allowed for weekly seasonality, as proposed by Hyndman [64].

Table 2 The single models structure for new COVID-19 Cases in Saudi

Models	**AICc**	**Structure**
ARIMA	2046.27	non-seasonal (1,1,1)
ETS	2099.156	(M,Ad,N)
NNAR	---	(8,4,1)
TBATS	AIC: 2093.39	(0.074, {0,0}, 0.932, -)
EST-A-N-N=F	2443.394	(A,N,N)

To fit the ARIMA statistical model for the data, the auto-arima function was utilized. As a result, the best ARIMA model was proposed. Non-seasonal (1,1,1), for numbers of daily new COVID-19 cases. The next model is ETS with fits the best proposed for it was (M, Ad, N), and then the EST-A-N-N model with fits the best proposed for it was (A, N, N). Then, the NNAR model fits (8,4,1). In the next model TBATS (0.074, {0,0}, 0.932, -), hybrid models were created by combining the best single models having equal weights and proved to be suitable than error-based weighting. Each of the models in the Hybrid models is given a weight. There are four ways that can be used in order to do so. "equal", "*cv.errors*" (i.e. *Cross Validated errors*), and "*insample.errors*". the researchers applied two techniques to construct these models: "equal" and "*cv.errors*". In the testing stage, the training model to predict the duration of the test data was used and was compared to the testing data. In the testing and training phases, the performance measures MAE, RMSE, and MAPE were calculated for all of the models. Table 3 summarizes the top model's findings. In addition, these findings were presented graphically in Fig. 3 using bar graphs.

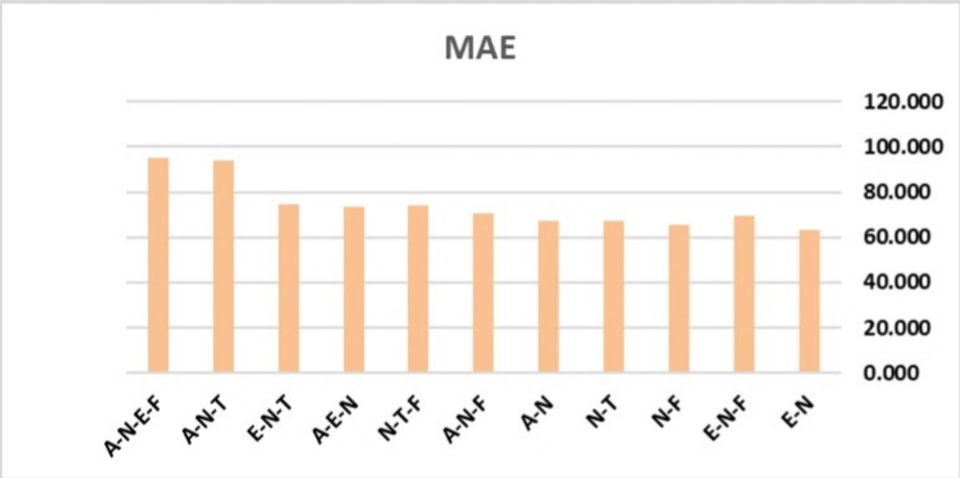


Fig 3(A) Models performance ranked by the MAE metric for new COVID-19 Cases in the test phase.

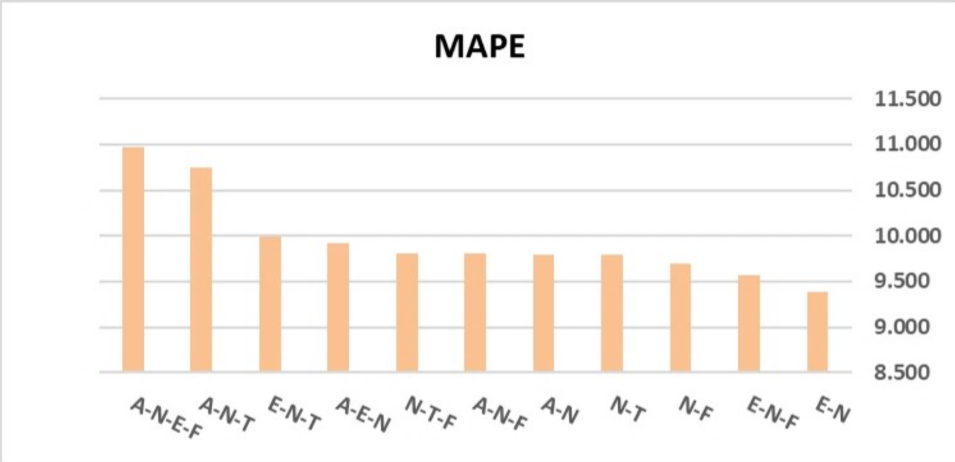


Fig 3(B) MAPE-ranked models metric for new COVID-19 Cases in the test phase.

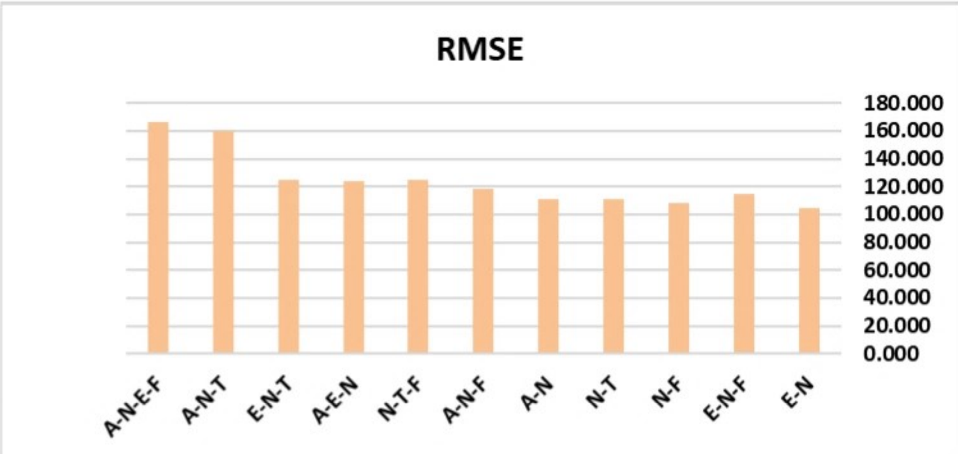


Fig (3.C): RMPE-ranked models metric for new COVID-19 Cases in the test phase.

Table 3. Measures of forecast accuracy for hybrid models for new COVID-19 Cases.

	MAE	MAPE	RMSE	ACF1	Theil's U
E-N	63.374	9.381	104.319	-0.050	0.706
E-N-F	69.602	9.567	114.376	0.001	0.720
N-F	65.676	9.693	108.292	-0.004	0.727
NT	67.119	9.792	110.535	-0.040	0.726
A-N	67.119	9.792	110.535	-0.040	0.726
A-N-F	70.401	9.804	118.031	-0.029	0.731
N-T-F	74.343	9.811	124.763	-0.064	-0.064
A-E-N	73.328	9.919	123.799	-0.058	0.733
E-N-T	74.498	9.988	124.955	-0.060	0.735
A-N-T	93.711	10.742	159.87910	-0.09837	0.77036
A-N-E-F	95.348	10.965	166.04220	-0.03341	0.79465

At lag 1, the autocorrelation function (ACF) revealed that current values were unrelated to prior ones as shown in Table 3. The correlation Coeff between one point in the time series and the next ranged from -0.06 to 0.79 for the wave of daily COVID-19 cases. The highest value (0.79) was obtained for the A-N-E-F model. Since MAPE has always meant less than ten, according to Lewis'[69] interpretation, all of the proposed forecasting models outperformed the forecasts from the other models (no-change) "nave" techniques, i.e. the forecasts with no modifications for casual factors Hyndman,[64]. This enables the adoption of increasingly complex and sophisticated model's hybrid combinations of (ARIMA, ETS, TBATS, NNAR, EST-A-N-N) to be justified. In table 4 and Fig 3 (A, B, C) the researchers compared the top hybrid models which were represented in eleven hybrid models out of the models that were studied by considering MAE, MAPE, Theil's U, and RMSE minimization. For the wave of daily COVID-19 cases in Saudi Arabia, hybrid E-N, is better than the respective other models in MAPE the accuracy measures, and the E-N-T is better than the respective other models in RMSE and AME the accuracy measures. For further clarification, the researchers have compared all models with each other according to the measurement criteria as shown in tables 4,5,6, and calculated the percentage of all hybrid models' MAPE, RMSE, and MAE minimization efficiency improvements. Specifically, hybrid E-N and F-T-N outperform all other models.

Table 4. The efficiency of hybrid models is compared by MAPE.

	E-N	N-F	N-T	A-N	E-N-F	A-N-F	N-T-F	A-E-N	E-N-T	A-N-T	A-N-E-F
E-N	0										
N-F	-3.22%	0									
N-T	-4.20%	-1.01%	0								
A-N	-4.20%	-1.01%	0.00%	0							
E-N-F	-1.94%	1.32%	2.35%	2.35%	0						
A-N-F	-4.31%	-1.13%	-0.12%	-0.12%	-2.42%	0					
N-T-F	-4.38%	-1.20%	-0.19%	-0.19%	-2.49%	-0.07%	0				
A-E-N	-5.42%	-2.28%	-1.28%	-1.28%	-3.55%	-1.16%	-1.09%	0			
E-N-T	-6.08%	-2.95%	-1.96%	-1.96%	-4.22%	-1.84%	-1.77%	-0.69%	0		
A-N-T	-12.67%	-9.77%	-8.84%	-8.84%	-10.94%	-8.73%	-8.67%	-7.66%	-7.02%	0	
A-N-E-F	-14.45%	-11.60%	-10.70%	-10.70%	-12.75%	10.59%	-10.52%	-9.54%	-8.91%	-2.03%	0

Notes: Negative figures represent the percentage increases in efficiency from employing hybrid models.

Table 5. The efficiency of hybrid models is compared by MAPE

	E-N	N-F	N-T	A-N	E-N-F	A-N-F	N-T-F	A-E-N	E-N-T	A-N-T	A-N-E-F
E-N	0										
N-F	-3.67%	0									
N-T	-5.62%	-2.03%	0								
A-N	-5.62%	-2.03%	0.00%	0							
E-N-F	-8.79%	-5.32%	-3.36%	-3.36%	0						
A-N-F	-11.62%	-8.25%	-6.35%	-6.35%	-3.10%	0					
N-T-F	-16.39%	-13.20%	-11.40%	-11.40%	-8.33%	-5.40%	0				
A-E-N	-15.74%	-12.53%	-10.71%	-10.71%	-7.61%	-4.66%	0.78%	0			
E-N-T	-16.51%	-13.34%	-11.54%	-11.54%	-8.47%	-5.54%	-0.15%	-0.93%	0		
A-N-T	-34.75%	-32.27%	-30.86%	-30.86%	-28.46%	-26.17%	-21.96%	-22.57%	-21.84%	0	
A-N-E-F	-37.17%	-34.78%	-33.43%	-33.43%	-31.12%	-28.92%	-24.86%	-25.44%	-24.75%	-3.71%	0

Notes: Negative figures represent the percentage increases in efficiency from employing hybrid models.

Table 6 The efficiency of hybrid-models is compared by MAE.

	E-N	N-F	N-T	A-N	E-N-F	A-N-F	N-T-F	A-E-N	E-N-T	A-N-T	A-N-E-F
E-N	0										
N-F	-3.51%	0									
N-T	-5.58%	-2.15%	0								
A-N	-5.58%	-2.15%	0	0							
E-N-F	-8.95%	-5.64%	-3.57%	-3.57%	0						
A-N-F	-9.98%	-6.71%	-4.66%	-4.66%	-1.13%	0					
N-T-F	-14.75%	-11.66%	-9.72%	-9.72%	-6.38%	-5.30%	0				
A-E-N	-13.57%	-10.44%	-8.47%	-8.47%	-5.08%	-3.99%	1.38%	0			
E-N-T	-14.93%	-11.84%	-9.90%	-9.90%	-6.57%	-5.50%	-0.21%	-1.57%	0		
A-N-T	-32.37%	-29.92%	-28.38%	-28.38%	-25.73%	-24.87%	-20.67%	-21.75%	-20.50%	0	
A-N-E-F	-33.53%	-31.12%	-29.61%	-29.61%	-27.00%	-26.16%	-22.03%	-23.09%	-21.87%	-1.72%	0

Notes: Negative figures represent the percentage increases in efficiency from employing hybrid models.

In figures 4 to 9, the researchers fit all the Eleven hybrid models for both the time series. The bright dark grey region displays the forecast interval at 70%, while the grey area indicates it at 90%. From May 30, 2022, to July 15, 2022, the best hybrid models predict a large drop in new COVID-19 cases. However, the number of new COVID-

19 cases will stabilize between the end of June 2022 and the beginning of July 2022. This is confirmed by the all forecasting hybrid methods, E-N, N-F, and E-N-F show that: i) after 10 days (Jul10), the number of the COVID-19 infections per day in Saudi Arabia will be 448, 413, or 386, respectively; ii) after 20 days (Jul 30), they will be 398, 214, or 254; and iii) after 43 days (July 13), they will be 406, 105, or 194. In particular, E-N, N-F, and E-N-F models exhibited nearly identical accuracy.

Fig 4: F-N and N-T hybrid models forecasting for daily new COVID-19 cases in Saudi Arabia:

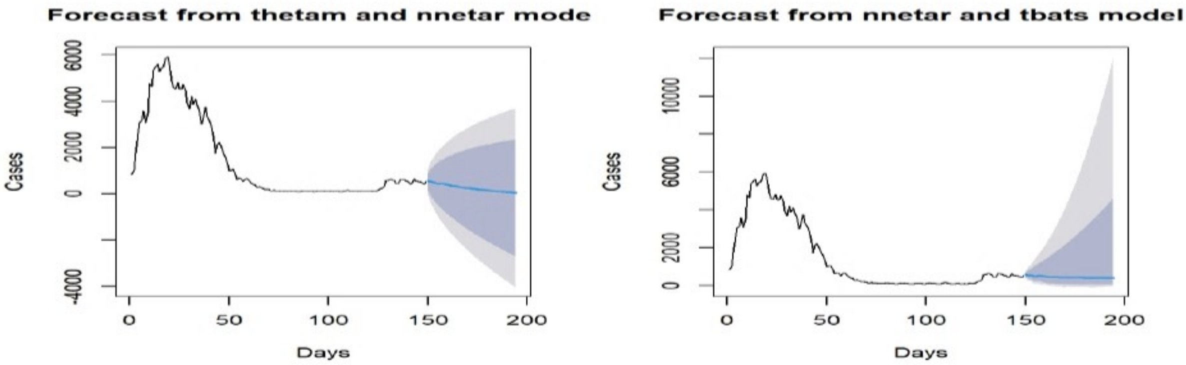


Fig 5: A-N and E-N hybrid models forecasting future for daily new COVID-19 cases in Saudi Arabia:

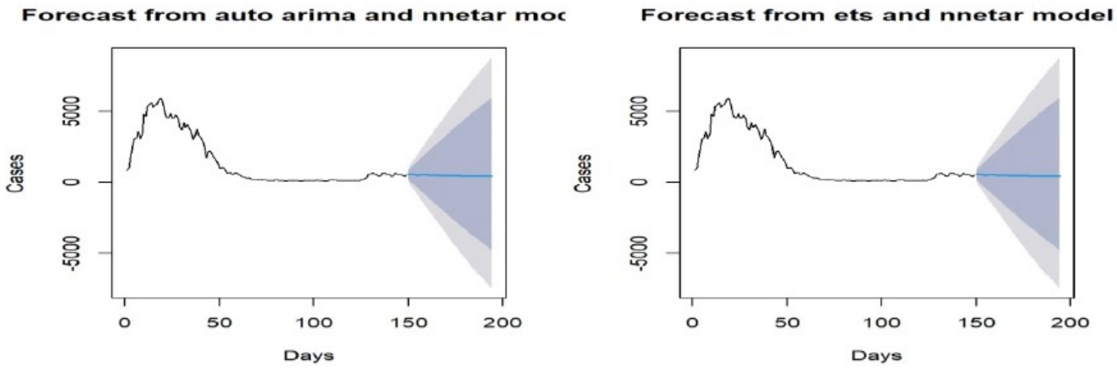


Fig 6: A-N-F and E-N-F hybrid models forecasting for daily new COVID-19 cases in Saudi Arabia:

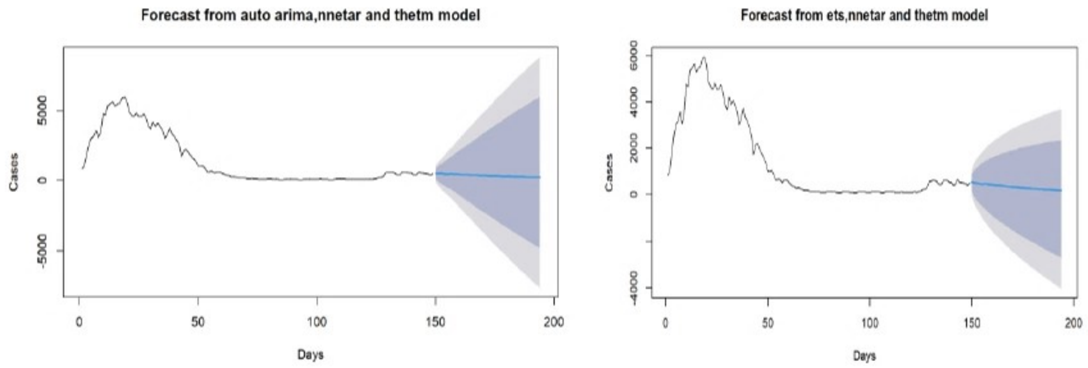


Fig 7: N-T-F and A-E-N hybrid models future for daily new COVID-19 cases in Saudi Arabia:

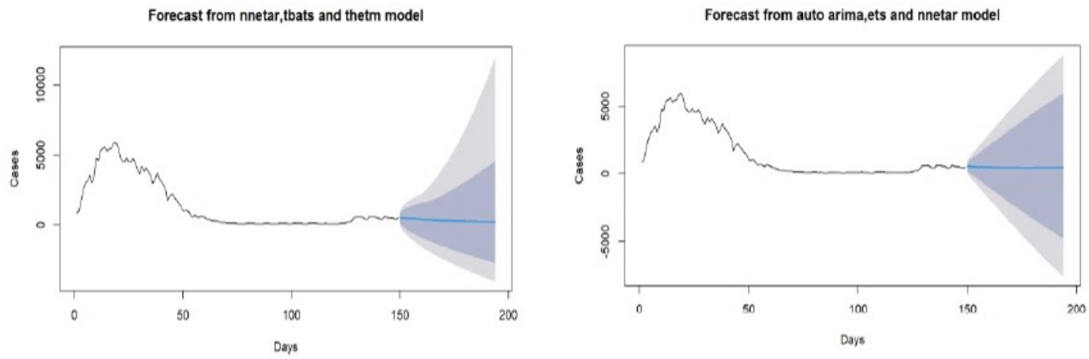


Fig 8: A-N-T and E-N-T hybrid models forecasting future for daily new COVID-19 cases in Saudi Arabia:

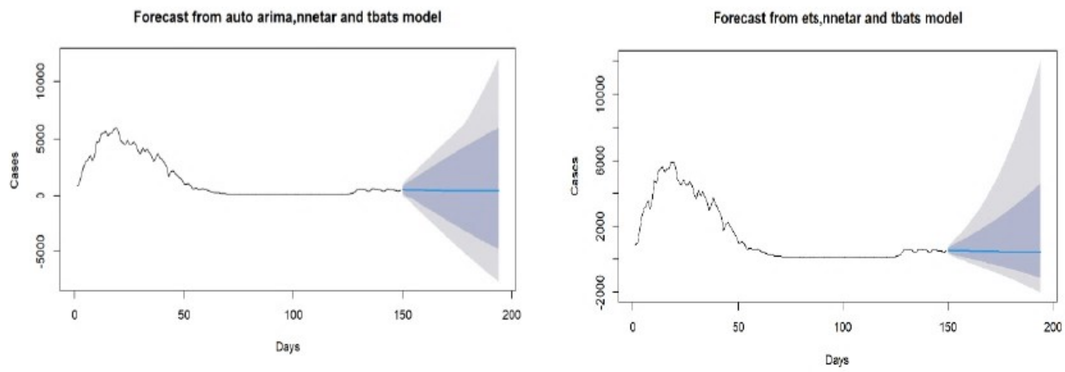
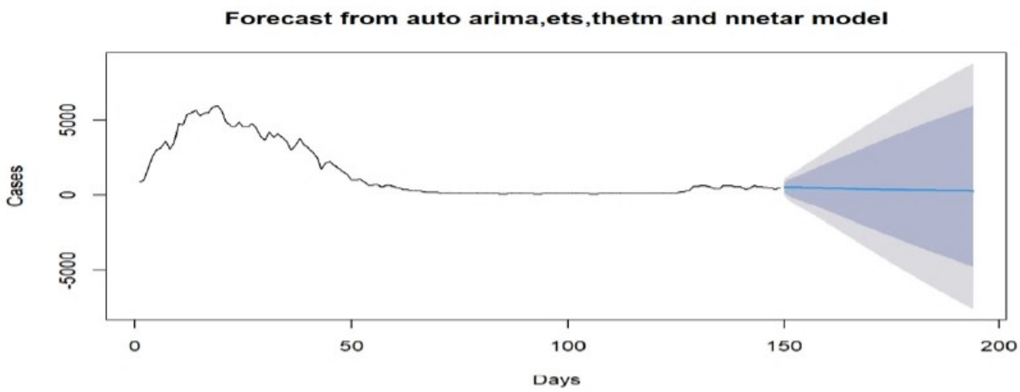


Fig 9: A-E-F-N hybrid models forecasting future for daily new COVID-19 cases in Saudi Arabia:



5. Conclusion:

Combining several models has been shown in the literature to be an effective method for generating superior outcomes in comparison to individual models and improving forecasting accuracy. In this study, eleven combinations of ETS, (EST, ANN), ARIMA, TBATS, and NNAR hybrid models which are the most essential and extensively used linear and nonlinear predicting models, are provided in order to increase the forecasting accuracy of models in the daily new COVID-19 cases.

Thus, by comparing the results of these hybrid models with each other, the models E-N, E-N-F, and N-F are in first place in preference compared to other models. While A-N-T and A-N-E-F are occupying the last places in the list of preferences by comparing the studied models. Also, the usage of existing aforementioned hybrid models' time series forecasting approaches, hybrid models by ETS, EST-A-N-N, and Neural networks appear to improve the chances of forecasting a number of combinations of linear and nonlinear epidemic patterns better than others. Finally, the mixing of several models is an excellent technique to produce more accurate COVID-19 pandemic time series forecasting results. The results show that E-N, E-N-F, and N-F fared were the best, while other models did well in terms of forecasting except for the two models A-N-T and A-N-E-F.

References:

[1] C. Xu *et al.*, "Estimation of reproduction numbers of COVID-19 in typical countries and epidemic trends under different prevention and control scenarios," *Front. Med.*, vol. 14, no. 5, pp. 613–622, Oct. 2020, doi: 10.1007/s11684-020-0787-4.

[2] H. T. Pham, V. T. Tran, and B. Yang, "A hybrid of nonlinear autoregressive model with exogenous input and autoregressive moving average model for long-term machine state forecasting," *Expert Systems With Applications*, vol. 37, no. 4. pp. 3310–3317, 2010, [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2009.10.020.

[3] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks," *BMC Bioinformatics*, vol. 15, no. 1, pp. 1–9, 2014.

[4] M. Hanf, A. Adenis, M. Nacher, and B. Carme, "The role of El Niño southern oscillation (ENSO) on variations of monthly Plasmodium falciparum malaria cases at the cayenne general hospital, 1996-2009, French Guiana," *Malar. J.*, vol. 10, pp. 1–4, 2011.

[5] A. Dominguez, P. Muñoz, A. Martínez, and A. Orcau, "Monitoring mortality as an indicator of influenza in Catalonia, Spain.," *J. Epidemiol. Community Heal.*, vol. 50, no. 3, pp. 293–298, 1996.

[6] T. A. Reichert, L. Simonsen, A. Sharma, S. A. Pardo, D. S. Fedson, and M. A. Miller, "Influenza and the winter increase in mortality in the United States, 1959–1999," *Am. J. Epidemiol.*, vol. 160, no. 5, pp. 492–502, 2004.

[7] J. Yi, C. T. Du, R. H. Wang, and L. Liu, "Applications of multiple seasonal autoregressive integrated moving average (ARIMA) model on predictive incidence of tuberculosis," *Zhonghua Yu Fang Yi Xue Za Zhi*, vol. 41, no. 2, pp. 118–121, 2007.

[8] Q. Liu, X. Liu, B. Jiang, and W. Yang, "Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model," *BMC Infect. Dis.*, vol. 11, no. 1, pp. 1–7, 2011.

[9] E. Gecili, A. Ziady, and R. D. Szczesniak, "Forecasting COVID-19 confirmed cases, deaths and recoveries: revisiting established time series modeling through novel applications for the USA and Italy," *PLoS One*, vol. 16, no. 1, p. e0244173, 2021.

[10] H. Tandon, P. Ranjan, T. Chakraborty, and V. Suhag, "Coronavirus (COVID-19): ARIMA-based Time-series Analysis to Forecast near Future and the Effect of School Reopening in India," *J. Health Manag.*, vol. 24, no. 3, pp. 373–388, 2022.

[11] M. Aslam, "Using the kalman filter with Arima for the COVID-19 pandemic dataset of Pakistan," *Data Br.*, vol. 31, p. 105854, 2020.

[12] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, "Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis," *Int. J. Infect. Dis.*, vol. 93, pp. 201–204, 2023.

[13] R. C. Das, "Forecasting incidences of COVID-19 using Box-Jenkins method for the period July 12-Septembert 11, 2020: A study on highly affected countries," *Chaos, Solitons and Fractals*, vol. 140. 2020, doi: 10.1016/j.chaos.2020.110248.

[14] X. Liu, Z. Ahmad, S. K. Khosa, M. Yusuf, O. A. Alamri, and W. Emam, "A New Flexible Statistical Model: Simulating and Modeling the Survival Times of COVID-19 Patients in China," *Complexity*, vol. 2021, pp. 1–16, Aug. 2021, doi: 10.1155/2021/6915742.

[15] M. Qamruzzaman, W. Jianguo, S. Jahan, and Z. Yingjun, "Financial innovation, human capital development, and economic growth of selected South Asian countries: An application of ARDL approach," *International Journal of Finance and Economics*, vol. 26, no. 3. pp. 4032–4053, 2021, doi: 10.1002/ijfe.2003.

[16] I. Ahmed, A. M. Medabesh, K. H. Mehrez, N. Shili, C. Socci, and S. Deriu, “Socioeconomic spillovers of COVID-19: case of Saudi Arabia,” *Kybernetes*, 2021.

[17] 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, 2020, doi: 10.1016/j.dib.2020.105340.

[18] G. E. Box and D. R. Cox, “An analysis of transformations revisited, rebutted,” *Journal of the American Statistical Association*, vol. 77, no. 377. pp. 209–210, 1964, doi: 10.1080/01621459.1982.10477788.

[19] Rajni and T. Kumar, “Forecasting of COVID-19 Outbreak in India: A Time Series Analysis,” in *Advances in Transdisciplinary Engineering*, 2023, vol. 32, pp. 698–707, doi: 10.3233/ATDE221334.

[20] N. Banholzer, T. Mellan, H. J. T. Unwin, S. Feuerriegel, S. Mishra, and S. Bhatt, “A comparison of short-term probabilistic forecasts for the incidence of COVID-19 using mechanistic and statistical time series models 1 Introduction,” pp. 1–21, 2023.

[21] C. W. J. Granger, “Investigating Causal Relations by Economic Models and Cross spectral Methods,” *Econometrica*, vol. 37, no. 3. pp. 424–438, 1969.

[22] P. Newbold and C. W. J. Granger, “Experience with forecasting univariate time series and the combination of forecasts,” *J. R. Stat. Soc. Ser. A*, vol. 137, no. 2, pp. 131–146, 1974.

[23] W. and Markakis, “Combining Forecasts,” *Public Hist.*, vol. 3, no. 3, pp. 139–139, 1983, doi: 10.2307/3377747.

[24] R. T. Clemen, “Combining forecasts: A review and annotated bibliography,” *Int. J. Forecast.*, vol. 5, no. 4, pp. 559–583, 1989, doi: 10.1016/0169-2070(89)90012-5.

[25] P. F. Pai and W. C. Hong, “Support vector machines with simulated annealing algorithms in electricity load forecasting,” *Energy Conversion and Management*, vol. 46, no. 17. pp. 2669–2688, 2005, doi: 10.1016/j.enconman.2005.02.004.

[26] K. Y. Chen and C. H. Wang, “A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan,” *Expert Syst. Appl.*, vol. 32, no. 1, pp. 254–264, 2007, doi: 10.1016/j.eswa.2005.11.027.

[27] M. Khashei, S. Reza Hejazi, and M. Bijari, “A new hybrid artificial neural networks and fuzzy regression model for time series forecasting,” *Fuzzy Sets and Systems*, vol. 159, no. 7. pp. 769–786, 2008, doi: 10.1016/j.fss.2007.10.011.

[28] D. Zeng, J. Xu, J. Gu, L. Liu, and G. Xu, “Short term traffic flow prediction using hybrid ARIMA and ANN models,” *Proceedings - 2008 Workshop on Power Electronics and Intelligent Transportation System, PEITS 2008*. pp. 621–625, 2008, doi: 10.1109/PEITS.2008.135.

[29] C. H. Aladag, E. Egrioglu, and C. Kadilar, “Forecasting nonlinear time series with a hybrid methodology,” *Applied Mathematics Letters*, vol. 22, no. 9. pp. 1467–1470, 2009, doi: 10.1016/j.aml.2009.02.006.

[30] R. Ebrahimpour, H. Nikoo, S. Masoudnia, M. R. Yousefi, and M. S. Ghaemi, “Mixture of mlp-experts for trend forecasting of time series: A case study of the tehran stock exchange,” *Int. J. Forecast.*, vol. 27, no. 3, pp. 804–816, 2011, doi: 10.1016/j.ijforecast.2010.02.015.

[31] W. Kristjanpoller and M. C. Minutolo, “Gold price volatility: A forecasting approach using the Artificial Neural Network-GARCH model,” *Expert Syst. Appl.*, vol. 42, no. 20, pp. 7245–7251, 2015, doi: 10.1016/j.eswa.2015.04.058.

[32] T. Xiong, C. Li, and Y. Bao, “Interval-valued time series forecasting using a novel hybrid HoltI and MSVR model,” *Economic Modelling*, vol. 60. pp. 11–23, 2017, doi: 10.1016/j.econmod.2016.08.019.

[33] C. Huang *et al.*, “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,” *Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[34] R. Bhardwaj, "A Predictive Model for the Evolution of COVID-19," *Transactions of the Indian National Academy of Engineering*, vol. 5, no. 2. pp. 133–140, 2020, doi: 10.1007/s41403-020-00130-w.

[35] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos, Solitons and Fractals*, vol. 134, p. 109761, May 2020, doi: 10.1016/j.chaos.2020.109761.

[36] G. Giordano *et al.*, "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy," *Nature Medicine*, vol. 26, no. 6. pp. 855–860, 2020, doi: 10.1038/s41591-020-0883-7.

[37] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," *Lancet*, vol. 395, no. 10225, pp. 689–697, Feb. 2020, doi: 10.1016/S0140-6736(20)30260-9.

[38] S. I. Alzahrani, 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, Jul. 2020, doi: 10.1016/j.jiph.2020.06.001.

[39] O. Joseph, B.-A. Senyefia, N. Christiana Cynthia, O.-A. Eunice, and B. Ernest Yeboah, "Covid-19 Projections: Single Forecast Model Against Multi-Model Ensemble," *Int. J. Syst. Sci. Appl. Math.*, vol. 5, no. 2, p. 20, 2020, doi: 10.11648/j.ijssam.20200502.12.

[40] J. Cao, X. Jiang, and B. Zhao, "Epidemic Prediction of COVID-19," *Adv Biotech & Micro*, vol. 15, no. 4. p. 555916, 2020, [Online]. Available: https://juniperpublishers.com/aibm/AIBM.MS.ID.555916.php.

[41] P. Melin, J. C. Monica, D. Sanchez, and O. Castillo, "Multiple ensemble neural network models with fuzzy response aggregation for predicting covid-19 time series: The case of mexico," *Healthc.*, vol. 8, no. 2, 2020, doi: 10.3390/healthcare8020181.

[42] T. Sardar, S. S. Nadim, S. Rana, and J. Chattopadhyay, "Assessment of lockdown effect in some states and overall India: A predictive mathematical study on COVID-19 outbreak," *Chaos, Solitons and Fractals*, vol. 139, p. 110078, Oct. 2020, doi: 10.1016/j.chaos.2020.110078.

[43] N. Hasan, "A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model," *Internet of Things (Netherlands)*, vol. 11, 2020, doi: 10.1016/j.iot.2020.100228.

[44] G. Perone, "Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy," *Eur. J. Heal. Econ.*, Aug. 2021, doi: 10.1007/s10198-021-01347-4.

[45] M. Ala'raj, M. Majdalawieh, and N. Nizamuddin, "Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections," *Infect. Dis. Model.*, vol. 6, pp. 98–111, 2021, doi: 10.1016/j.idm.2020.11.007.

[46] Z. Ceylan, "Estimation of COVID-19 prevalence in Italy, Spain, and France," *Science of the Total Environment*, vol. 729. 2020, doi: 10.1016/j.scitotenv.2020.138817.

[47] 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 and Fractals*, vol. 135, 2020, doi: 10.1016/j.chaos.2020.109853.

[48] N. Talkhi, N. A. Fatemi, Z. Ataei, and M. J. Nooghabi, "Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods," *Biomed. Signal Process. Control*, vol. 66, p. 102494, 2021.

[49] T. Chakraborty and I. Ghosh, "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis," *Chaos, Solitons & Fractals*, vol. 135, p. 109850, 2020.

[50] D. Fantazzini, "Short-term Forecasting of the COVID-19 Pandemic using Google Trends Data: Evidence from 158 Countries," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3671005.

[51] S. J. Alsunaidi *et al.*, "Applications of big data analytics to control covid- 19 pandemic," *Sensors*, vol. 21, no. 7. 2021, doi: 10.3390/s21072282.

[52] İ. Kırbaş, A. Sözen, A. D. Tuncer, and F. Ş. Kazancıoğlu, "Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches," *Chaos, Solitons & Fractals*, vol. 138, p. 110015, 2020.

[53] H. Yonar, A. Yonar, M. A. Tekindal, and M. Tekindal, "Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the Curve Estimation Models, the Box-Jenkins and Exponential Smoothing Methods," *EJMO*, vol. 4, no. 2, pp. 160–165, 2020.

[54] S. Dhamodharavadhani and R. Rathipriya, "Covid-19 mortality rate prediction for india using statistical neural networks and gaussian process regression model," *African Health Sciences*, vol. 21, no. 1. pp. 194–206, 2021, doi: 10.4314/ahs.v21i1.26.

[55] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, "COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population," *Appl. Sci.*, vol. 10, no. 11, p. 3880, Jun. 2020, doi: 10.3390/app10113880.

[56] S. Bhandary *et al.*, "Trend analysis, modelling and impact assessment of COVID-19 in Nepal," *medRxiv*, 2020.

[57] G. Perone, "Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy," *Eur. J. Heal. Econ.*, pp. 1–24, 2021.

[58] R. Salgotra and A. H. Gandomi, "Time series analysis of the COVID-19 pandemic in Australia using genetic programming," in *Data Science for COVID-19 Volume 1: Computational Perspectives*, 2021, pp. 399–411.

[59] S. Mohan, A. K. Solanki, H. K. Taluja, Anuradha, and A. Singh, "Predicting the impact of the third wave of COVID-19 in India using hybrid statistical machine learning models: A time series forecasting and sentiment analysis approach," *Comput. Biol. Med.*, vol. 144, 2022, doi: 10.1016/j.compbiomed.2022.105354.

[60] W. Jin, S. Dong, C. Yu, and Q. Luo, "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning," *Comput. Biol. Med.*, vol. 146, 2022, doi: 10.1016/j.compbiomed.2022.105560.

[61] K. E. ArunKumar, D. V. Kalaga, C. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends," *Alexandria Eng. J.*, vol. 61, no. 10, pp. 7585–7603, 2022, doi: 10.1016/j.aej.2022.01.011.

[62] M. Kalantari, "Forecasting COVID-19 pandemic using optimal singular spectrum analysis," *Chaos, Solitons & Fractals*, vol. 142, p. 110547, Jan. 2021, doi: 10.1016/j.chaos.2020.110547.

[63] M. N. Atchadé, Y. M. Sokadjo, A. D. Moussa, S. V. Kurisheva, and M. V. Bochenina, "Cross-Validation Comparison of COVID-19 Forecast Models," *SN Comput. Sci.*, vol. 2, no. 4, p. 296, Jul. 2021, doi: 10.1007/s42979-021-00699-1.

[64] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2008.

[65] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[66] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing," *J. Am. Stat. Assoc.*, vol. 106, no. 496, pp. 1513–1527, Dec. 2011, doi: 10.1198/jasa.2011.tm09771.

[67] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecast.*, vol. 32, no. 3, pp. 669–679, 2016.

[68] G. Paul and L. Richard, "On the asymmetry of the symmetric MAPE," *Int. J. Forecast.*, pp. 405–408, 1999.

[69] C. D. Lewis, *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann, 1982.

Appendix A.

Table A1: The parameter values of the ARIMA, NNAR, TBATS EANN, ETS hybrid models

ARIMA (1,1,1)

Coefficients:

ar1ma1

0.9568 -0.8390

s.e. 0.0423 0.0568

sigma^2 = 57369: log likelihood = -1020.05

AIC=2046.11 AICc=2046.27

BIC=2055.1

(EANN) = F

forecast: ets (y = y, model = "EANN", opt.crit = "mse")

Smoothing parameters:

alpha = 0.9999

Initial states:

l = 2727.7663

sigma: 293.9481

AIC AICc BIC

2443.229 2443.394 2452.240

ETS

Smoothing parameters:

alpha = 0.9999

beta = 0.0357

phi = 0.9071

Initial states:

l = 722.7654

b = 377.0652

sigma: 0.1736

AIC AICc BIC

2098.564 2099.156 2116.588

NNAR

Average of 20 networks, each of which is a 8-4-1 network with 41 weights

options were - linear output units

sigma^2 estimated as 7930

TBATS

Parameters

Lambda: 0.074246

Alpha: 1.129206

Beta: 0.04241233

Damping Parameter: 931892

Seed States:

[1,] 9.5977738

[2,] 0.2293691

attr(,"lambda")

[1] 0.07424625

Sigma: 0.2672315

AIC: 2093.39

مقارنة الدقة التنبؤية للنماذج الهجينة من NNAR وTBATS وETS-EANN و ARIMA لانتشار حالات COVID -19 في المملكة العربية السعودية

أ.د أحمد غالب الرباصي، د. فيصل علي محمد علي، د. فهمي الصلوي، د. أحمد عبد الرزاق الأنصاري

الملخص:

في ديسمبر 2019، ظهر مرض فيروس كورونا (COVID-19) وتسبب في جائحة عالمية لا تزال خطورتها مستمرة. هناك حاجة ملحة لمراقبة انتشار COVID-19 والتنبؤ به للسيطرة على هذا الانتشار بشكل أكثر فعالية. تعد نماذج السلاسل الزمنية مهمة في التنبؤ بتأثير تفشي COVID-19 واتخاذ التدابير اللازمة للاستجابة لهذه الأزمة. على الرغم من وجود عدد من نماذج التنبؤ بالسلاسل الزمنية، فإنه من الصعب إصدار بيانات عامة واسعة عن أيها الأفضل. تقدم هذه الورقة مراجعة تجريبية لعدد من نماذج السلاسل الزمنية للتنبؤ بحالات انتشار COVID -19 في المملكة العربية السعودية، إذ تفتقر الدول العربية إلى دراسات في تشخيص هذا الجانب، وعلى وجه الخصوص يسهم هذا البحث في الأدبيات المتعلقة بالتنبؤ ببيانات السلاسل الزمنية بطرق متعددة، تتضح من خلال مقارنة نتائج مجموعة متنوعة من أساليب التنبؤ باستخدام تقنيات التنبؤ المتسلسل. وتشمل هذه المجموعات نماذج هجينة من النماذج الأُسِيَّة ETS و TBATS ونموذج ARIMA المتوسط المتحرك الانحدار الذاتي ونموذج NNAR الانحدار الذاتي للشبكة العصبية، وبالتالي تم صياغة أحد عشر نموذجًا هجينًا للجمع بين النماذج المتمثلة في (E-N ، و E-N-F، و N-F، و N-T، و A-N، و A-N-F، و N-T-F ، و A-E-N، و E-N-T ، و A-N-T ، وA-N-E-F). كما تم جمع البيانات اليومية عن انتشار حالات COVID-19 في المملكة العربية السعودية من يناير 2022 إلى 30 مايو 2022 من موقع منظمة الصحة العالمية. وتم اختيار أفضل نموذج تنبؤ بناءً على معايير المقارنة المتمثلة في جذر متوسط الخطأ التربيعي (RMSE) ومتوسط النسبة المئوية للخطأ المطلق (MAPE) ، ومتوسط الخطأ المطلق (MAE). تظهر نتائج هذه الدراسة أن أفضل النتائج هي عن النماذج المتمثلة في E-N و E-N-F و N-F إذ كان لها أدنى خطأ في التنبؤ.

الكلمات المفتاحية: تنبؤات، النماذج الهجينة COVID-19, ARIMA; ETS; TBATS NNAR F

ISSN: 2708-3659

جامعة الريان
AL-RAYAN UNIVERSITY

مجلة الريان

للعلوم الإنسانية والتطبيقية

علمية محكّمة - نصف سنوية

المجلد السادس
العدد الثاني
الرقم التسلسلي 11

ديسمبر 2023