Artificial neural networks for prediction of COVID-19 in India by using backpropagation

Balakrishnama Manohar1 | Raja Das2

1Research Scholar, School of Advanced Sciences, Vellore Institute of Technology, Vellore, India
2Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, India

Correspondence
Raja Das, Assistant Professor, School of Advanced Sciences, Vellore Institute of Technology, Vellore 632014, India.
Email: rajadas@vit.ac.in

Abstract
The COVID-19 pandemic has affected thousands of people around the world. In this study, we used artificial neural network (ANN) models to forecast the COVID-19 outbreak for policymakers based on 1st January to 31st October 2021 of positive cases in India. In the confirmed cases of COVID-19 in India, it’s critical to use an estimating model with a high degree of accuracy to get a clear understanding of the situation. Two explicit mathematical prediction models were used in this work to anticipate the COVID-19 epidemic in India. A Boltzmann Function-based model and Beesham’s prediction model are among these methods and also estimated using the advanced ANN-BP models. The COVID-19 information was partitioned into two sections: training and testing. The former was utilized for training the ANN-BP models, and the latter was used to test them. The information examination uncovers critical day-by-day affirmed case changes, yet additionally unmistakable scopes of absolute affirmed cases revealed across the time span considered. The ANN-BP model that takes into consideration the preceding 14-days outperforms the others based on the archived results. In forecasting the COVID-19 pandemic, this comparison provides the maximum incubation period, in India. Mean square error, and mean absolute percent error have been treated as the forecast model performs more accurately and gets good results. In view of the findings, the ANN-BP model that considers the past 14-days for the forecast is proposed to predict everyday affirmed cases, especially in India that have encountered the main pinnacle of the COVID-19 outbreak. This work has not just demonstrated the relevance of the ANN-BP techniques for the expectation of the COVID-19 outbreak but also showed that considering the incubation time of COVID-19 in forecast models might produce more accurate assessments.

KEYWORDS
artificial intelligence, artificial neural network, backpropagation, Beesham’s prediction model, Boltzmann function-based model, COVID-19 outbreak, estimate model

1 INTRODUCTION

A viral disease caused by a novel coronavirus (nCov) or 2019-nCov was first identified in Wuhan, China, at the end of December 2019 (Barda et al., 2020; Fanelli et al., 2020; Kavadi et al., 2020; Nishiura et al., 2020). The person-to-person transfer of coronavirus is one of the...
epidemic’s most difficult hurdles. Infected instances of the coronavirus (COVID-19) are increasing at an exponential rate all over the world. COVID-19 is a dangerous disease caused by the SARS-COV (severe acute respiratory syndrome corona virus) family, and it has become the world’s worst health catastrophe of the 21st century. The World Health Organization (WHO) proclaimed it a global pandemic on March 11, 2020, just a few months after the first case was detected in Wuhan, China (Ibrahim et al., 2020). Patients infected with COVID-19 suffered with common symptoms, including cough, fever and respiratory problems. In the worst-case scenario, it could lead to major health problems such as kidney failure and pneumonia, which could lead to patient death. COVID-19 appears to be spread by coughs, sneezes and human-to-human transmission (Epidemiologi, 2019). Between January 20, 2020 and March 02, 2022, approximately 444 million people were infected globally, resulting in more than 5.99 million deaths cases; in India, nearly 43 million people were infected positively, resulting in 0.515 million deaths cases (WHO Coronavirus (COVID-19) Dashboard With Vaccination Data, 2022).

In this regard, The COVID-19 outbreak has been predicted using mathematical, dynamical and statistical methods. The SEIR (susceptible exposed infectious recovered) model (Rocklov et al., 2020), the logistic growth model (Roosa et al., 2020), and the adaptive neuro fuzzy inference system (ANFIS) model (Al-Qaness et al., 2020) are among models that have been suggested for this purpose. For example, (Al-Qaness et al., 2020) updated the ANFIS a model for predicting COVID-19’s spread using pollination and salp swarm algorithms. For the purpose of calculating the total number of confirmed cases in China, (Fu et al., 2020) based on the Boltzmann function technique. Niazkar and Niazkar are found in seven countries: Chinese, Korea, Japanese, Italia, Singaporean, Iranian and the U.S.A (Niazkar, & H. N.-E. J. of G. Medicine, and U, 2020) utilized multigen hereditary programming, an AI model, to build numerical models that include outstanding capacity for anticipating the COVID-19 pandemic. Prediction models based on countries were suggested. They advised that each COVID-19 outbreak be researched separately in each country. Furthermore, (Hu et al., 2020) proposed utilizing a dramatic capacity to conjecture the COVID-19 outbreak direction. We create district-based prediction models using Niazkar and Niazkar (2020). Anticipating the number of Coronavirus Patients in Malang City using the back-propagation neural network with the Fletcher reeves Method is recommended by Syaiful Anam and Mohammed Hakim Akbar Assidiq Maulana (Anam et al., 2021). To our knowledge, only few artificial neural network (ANN) have been used to anticipate the COVID-19 Outbreak. For the forecast of recuperated and mortality cases, Al-Najjar and Al-Rousan (Al-Najjar & N. A.-R.-E. R. for M. and U, 2020; F. M.-E.-P. K. S. Bali and U, 2017) utilized ANN. Wang et al. (2021) (Wang et al., 2022) proposed Deep Rank-Based Average Pooling Network for Covid-19 Recognition. Wang et al. (2021) (Wang et al., 2021) proposed DSSAE; Deep Stacked Sparse Autoencoder Analytical Model for COVID-19 Diagnosis by Fractional Fourier Entropy.

Further section of this study is organized as follows: Section 2 presents the two explicit mathematical prediction models that is a Boltzmann Function based model, Beesham’s prediction model and the ANN-BP models-based prediction, Section 3 presents results of the mathematical prediction models and that are applied for analysing the fact and trend of COVID-19 pandemic, Section 4 presents the discussion of ANN models of this study, finally, Section 5 presents the conclusion. The present work methodology and analysis shows in Figure 1”.

2 | MATERIALS AND METHODS

The goal of this study, based on the comparison, the ANN-BP model gives better results than a mathematical model (boltzmann function based model, beesham’s prediction model) that estimates the COVID-19 outbreak in India, which is some in the range of 2-days and 14-days model (O. O.-E. M. Journal and U, 2020). The complete affirmed instances of COVID-19 out of six areas in India were considered in this review. These districts were chosen from among the various districts affected with COVID-19 in India and culture areas to exhibit variety. Furthermore, their numbers of affirmed cases differ by an order of magnitude, providing the proposed models to be tested in districts with both high and low quantities of affirmed cases. One more justification behind picking these regions is that a couple of them have been contaminated with SARS-COV-2 for a more timeframe than numerous different regions. The relating information was isolated into two areas: Train information and Test information. The originally was utilized to train the ANN models, while the second was taken advantage of for examination purpose. Subsequently, the test information is anticipated aggregate affirmed cases were contrasted with those of noticed cases.

2.1 | Mathematical models

2.1.1 | Boltzmann function-based prediction model

The Boltzmann function utilized to develop a forecast model for forecasting the COVID-19 outburst (Fu et al., 2020). The derived relationship, which is presented in below Equation, is related to the sigmoid function except for a linear transform (Fu et al., 2020):

\[ P(t) = C_2 + \frac{C_1 - C_2}{1 + e^{(t-b)/\alpha}} \]
where $C_1$, $C_2$, $t_0$ and $\Delta t$ are constant coefficients. More specifically, $C_1$ denotes infections where could spread SARS-COV-2 to healthy individuals, while $C_2$ represents an estimate of COVID-19 instances that have been confirmed. Above equation includes the following four constant coefficients need to be calibrate using the train data of each district.

2.1.2 | Beesham's prediction model

Since the infectious rate of the COVID-19 outburst was rising rapidly, several models with exponential function have been suggested for forecasting the COVID-19 outburst in the literature (A. B.-A. P. J. of T. Medicine and U, 2020). Likewise, Beesham's mathematical model, which was suggested for anticipating the positive cases of the COVID-19 in South Africa (A. B.-A. P. J. of T. Medicine and U, 2020), contains an exponential function. This prediction model is shown in below Equation (A. B.-A. P. J. of T. Medicine and U, 2020):

$$p(t) = l^m e^{nt}$$

where the coefficients $l$, $m$ and $n$ are constants, for the train data, these coefficients can be computed using a standard regression analysis or parameter estimation procedure.

2.2 | Artificial neural network

Artificial neural network is a well-known AI technique that is based on biological structure human neurons. It has been effectively applied to a variety of challenges in many fields (Dharani & Kirupa Krishnan, 2021; Niazkar, 2019). It is, in essence, a useful tool for determining a connection
between the data input and output. ANNs is commonly used in the training process and has three layers: (1) an input, (2) a hidden and (3) an output layer. The first and final ones, respectively, include neurons related with the input and output vectors (Niazkar et al., 2019a). The nodes in the input layer must be associated to the nodes in the hidden layer, so each hidden node must be associated to the nodes in the output layer. These nodes are responsible for transforming the input data to the desired output data. A weighted summation of the input data is also transferred using a transfer function. In general, neurons in every layer of an ANN ought to connect with the ones with inside the subsequent layer and former layers, however, inter-layered connections aren’t allowed. The data continues to float through the network until a particular relationship is found. At last, the better the ANN is prepared, the more exact the outcomes (Niazkar, 2019).

2.2.1 | Backpropagation algorithm

We use the ANN backpropagation technique to predict COVID-19 cases in this study. The backpropagation algorithm has become the most used learning algorithm for ANN design. The learning algorithm is divided into two parts (Rumelhart, et al., 1986). Network settings are chosen in the first step, referred to as the forward phase, and the computer file is then carried forward via the network, layer-by-layer. The output error computation completes the forward phase (Brause, 2001; Haykin, 2004).

The backpropagation algorithm is a method for conducting ANN layer training in a methodical way. The Backpropagation algorithm is a popular method for resolving difficult problems. It’s been utilized in a variety of applications, including rainfall forecasting and compound function forecasting (Ratnawati et al., 2019). There are multiple layers in the backpropagation algorithm, which include input, hidden and output layers. A hidden layer is made up of m units as well as a bias. The output bias of \( v_{oj} \) and \( w_{0k} \) behaves as a weight, with the output bias always being one (Infotama, 2016).

The feed-forward step from pattern input training, backpropagation of related errors, and weight update are the three steps of the Backpropagation training process. The output of the pattern will be determined by calculating each input unit in the hidden layers during the advanced stage. The network’s output will be compared to the target during the training phase, and the error will be determined. The optimization process is then performed in order to acquire the components that spread the inaccuracy. The weight between the input and output layers is updated using this factor (Fausett, 2006).

The back-propagation learning method has the advantages of accuracy and ease in calculations. However, the algorithm’s convergence limit and poor training speed are disadvantages, especially in the case of difficult tasks that necessitate a large network (Haykin, 2004). The sigmoid function and other activation functions may help the back-propagation learning technique work well. Rearranging the training samples after each epoch’s presentation, or following an easy-to-learn example with a challenging one, might also assist. It is beneficial to preprocess the input data in order to eliminate the mean and de-correlate the data. By giving a learning rate parameter to neurons in the last layers that is less than those in the front end, it may be possible to achieve greater performance by providing for the neurons in the different levels to learn at basically the same pace (Brause, 2001; Haykin, 2004).

2.2.2 | Backpropagation algorithm

1. Initialize weights.
2. While stopping condition is false.

*Feed forward*

1. Each input unit \( X_i, i = 1, 2, \ldots, n \) receives the input signal, \( x_i \) and broadcasts it to the next layer.
2. The output of every hidden unit \( (z_j, j = 1, 2, 3, \ldots, m) \) is acquired by the use of Equation (1). The value is propagated returned to the subsequent layer.

\[
z_{inj} = v_{oj} + \sum_{i=1}^{n} x_i v_{ij},
\]
\[
z_j = f(z_{inj}).
\]

3. For each output \( (y_k, k = 1, 2, 3, \ldots, p) \) provides up the input that has been weighted the use of the subsequent equation.

\[
y_{ink} = w_{0k} + \sum_{j=1}^{m} z_j w_{jk}
\]
4. The output signal is acquired by calculating the activation function with the following equation.

\[ y_k = f(y_\cdot, in_k) \]  

(3)

**Backpropagation of error**

1. Each output unit \((y_k, k = 1, 2, 3, ..., p)\) obtains a sample related to the input sample, after which the error information is calculated using the following equation.

\[ \delta_k = (t_k - y_k) f'(y_\cdot, in_k) \]  

(4)

2. The error correction is calculated in order to update the weights afterward.

\[ \Delta w_{jk} = \alpha \delta_k x_j \]  

(5)

3. The bias correction is likewise calculated with Equation (6) after which passes \(\delta_k\) to the layer units afterward.

\[ \Delta w_{ok} = \alpha \delta_k \]  

(6)

4. The output weight is multiplied by the error information from the preceding layers units and the result is added as the input delta, which is determined using the following equation.

\[ \delta_\cdot in_j = \sum_{k=1}^{p} \delta_k w_{jk} \]  

(7)

5. The activation functions first derivative is computed. As a result, it’s multiplied by the error information.

\[ \delta_j = \delta_\cdot in_j f'(z_\cdot in_j) \]  

(8)

6. The error correction is then computed using (9) so that the weights can be updated afterward.

\[ \Delta v_{ij} = \alpha \delta_j x_i \]  

(9)

7. In addition, the bias correction is calculated from the equation below.

\[ \Delta v_{oj} = \alpha \delta_j \]  

(10)

**Update weights and biases**

1. The weights and biases of every output unit are updated by the use of the formula in Equation (11), to decrease the error:

\[ w_{jk}^{\text{new}} = w_{jk}^{\text{old}} + \Delta w_{jk} \]  

(11)

2. Therefore, the weights and biases of every hidden unit are updated through the usage of the formula in the following Equation (12):

\[ v_{ij}^{\text{new}} = v_{ij}^{\text{old}} + \Delta v_{ij} \]  

(12)

3. Check the stop condition.

To decide when the training process should be terminated and the size of the hidden layer, a statistical model known as cross-validation (Jha, n.d.) is used. The training data set is divided into two sections by the method: the estimate subset and the validation subset. The estimate subset is used to train a network, whereas the validation subset is used to assess the model’s effectiveness. Finally, the model is fine-tuned using the whole collection of training patterns before being evaluated on previously unseen data. The backpropagation algorithm is a technique that solves Predictive problems with good results. A feed-forward Backpropagation network algorithm was utilized to train ANN-technique using...
summarizes these strategies. In this Table, n.d. shows the data shows, the information of greater hyderabad municipal corporation (GHMC), Medchal site, we take 1st January to 31st

table, the first confirmed cases are recorded. Furthermore, the daily positive instances of the six major districts varies, with each district’s data typically beginning when the first confirmed cases are recorded. Furthermore, the daily positive instances of the six primary districts analysed show the most substantial changes. The wide range of data and significant variability necessitate the use of a sophisticated technique to forecast the COVID-19 epidemic.

| Model          | Equations |
|----------------|-----------|
| First model    | $d_i = g_{1}(d_{i-1})$ |
| Second model   | $d_i = g_{2}(d_{i-1}, d_{i-2})$ |
| Third model    | $d_i = g_{3}(d_{i-1}, d_{i-2}, d_{i-3})$ |
| Fourth model   | $d_i = g_{4}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4})$ |
| Fifth model    | $d_i = g_{5}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5})$ |
| Sixth model    | $d_i = g_{6}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6})$ |
| Seventh model  | $d_i = g_{7}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7})$ |
| Eighth model   | $d_i = g_{8}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7}, d_{i-8})$ |
| Ninth model    | $d_i = g_{9}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7}, d_{i-8}, d_{i-9})$ |
| Tenth model    | $d_i = g_{10}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7}, d_{i-8}, d_{i-9}, d_{i-10})$ |
| Eleventh model | $d_i = g_{11}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7}, d_{i-8}, d_{i-9}, d_{i-10}, d_{i-11})$ |
| Twelfth model  | $d_i = g_{12}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7}, d_{i-8}, d_{i-9}, d_{i-10}, d_{i-11}, d_{i-12})$ |
| Thirteenth model | $d_i = g_{13}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7}, d_{i-8}, d_{i-9}, d_{i-10}, d_{i-11}, d_{i-12}, d_{i-13})$ |
| Fourteenth model | $d_i = g_{14}(d_{i-1}, d_{i-2}, d_{i-3}, d_{i-4}, d_{i-5}, d_{i-6}, d_{i-7}, d_{i-8}, d_{i-9}, d_{i-10}, d_{i-11}, d_{i-12}, d_{i-13}, d_{i-14})$ |

MATLAB (Dharani & Kirupa Krishnan, 2021; Niazkar et al., 2019a). In this study, 14 ANN methods for estimating everyday high-quality instances are proposed. Table 1 summarizes these strategies. In this Table 1, $g_i$ is the ith function ($i = 1, 2, 3, ..., 14$) that ANN can approximate and $d_i$ is the number of everyday positive instances at time $t$. For the sake of clarifying, we take randomly, a diagram view of the 14th ANN technique is shown in Figure 2.

According to the models described in Table 1, when $g_i$ is discovered, it may be used to anticipate the COVID-19 breakout in the future. In the input, hidden, and output layers of the ANNs used to detect $g_i$, there are 14, 10 and 1 neuron, respectively. Because the 14 day’s model introduced in Table 1 needs the 14 prior days’ daily positive examples, the first 14 days of data were omitted from the total data to allow training of all ANNs models. As a result, the COVID-19 Outbreak data was split into two sections: (1) Training data (75%) and (2) Testing data (25%). The first was used to train ANNs, and the latter was used to calculate how well ANN-technique models predicted the COVID-19 Outbreak. ANNs immediate results were round up to integer values using the round function contained in Excel (Niazkar & Afzali, 2016).

### 2.2.3 | Data collection

The positive cases COVID-19 have been collected from WHO circumstance information (WHO Coronavirus (COVID-19) Dashboard|WHO Coronavirus (COVID-19) Dashboard With Vaccination Data, n.d.) and the information advising the number regarding COVID-19 victims utilized for assessment was taken from the India COVID-19 site (WHO Coronavirus (COVID-19) Dashboard|WHO Coronavirus (COVID-19) Dashboard With Vaccination Data, n.d.), India. The information was likewise distributed on the COVID-19.telangana.gov.in site, we take 1st January to 31st October 2021 of the information was gathered from cases that had been publicized in the past. In this study, confirmed cases of six main districts in India were considered.

### 2.2.4 | Data analysis

Descriptive analysis of the information collected was directed utilizing Excel, which offers powerful offices for investigating information and executing mathematical strategies (Dharani & Kirupa Krishnan, 2021). The consequences of this investigation and the time of information utilized for each area are summed up in Table 2. In this Table 2 shows, the information of greater hyderabad municipal corporation (GHMC), Medchal Malkajigiri and Karimnager depending on minimum, maximum and SD, have a greater range than others. Furthermore, Figure 3 shows the data periods of the six major districts varies, with each district’s data typically beginning when the first confirmed cases are recorded. Furthermore, the daily positive instances of the six primary districts analysed show the most substantial changes. The wide range of data and significant variability necessitate the use of a sophisticated technique to forecast the COVID-19 epidemic.
2.2.5 | Evaluation criteria

We used two dimensionless measures were chosen from the writing to compute the presentation of ANN - strategy models in anticipating the COVID-19 pandemic (Dharani & Kirupa Krishnan, 2021; Niazkar & Afzali, 2019; Niazkar et al., 2019b): (1) Mean square error (MSE), (2) Mean absolute percent error (MAPE). These records are composed for positive instances in Equation (13) to Equation (14), respectively:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (Y_{i, \text{known}} - Y_{i, \text{estimated}})^2$$  \hspace{1cm} (13)

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Y_{i, \text{known}} - Y_{i, \text{estimated}}}{Y_{i, \text{known}}} \right| \times 100$$  \hspace{1cm} (14)

Here $Y_{i, \text{known}}$ and $Y_{i, \text{estimated}}$ are the ith known and estimated values of everyday positive instances, respectively, $N$ is the total number of data and “$i$” is the counter. The above Equation (13) to Equation (14), when a predict model yields, it performs better (high or lower values) of MSE and MAPE.
3 | FINDINGS

Results of the boltzmann function-based prediction model and beesham’s prediction model: the calibrated coefficients were used to predict the positive cases of COVID-19 in India shown in Figure 4. The results predicted by the boltzmann function-based model and beesham’s prediction model are compared with the actual cases for the train data (1st January to 31st October 2021) in Figure 4. The coefficient values ($C_1$, $C_2$, $t_0$ and $\Delta t$) and $R^2$ for boltzmann function-based prediction model are shown in Table 3, and the coefficient values ($l$, $m$ and $n$) and $R^2$ for beesham’s prediction model are shown in the Table 4.

These graphs present a comparison of the mathematical models and ANN model on (Niazkar et al., 2020), where we can observe that ANN model outperform other two models and the boltzmann function-based prediction model, beesham’s prediction model. As the ANN model increases, the curve adapts better towards the actual data of each month for different districts shown in Figure 4.

3.1 | Results of the artificial neural network

The proposed models were utilized to forecast daily confirmed COVID-19 cases in India’s major six districts after ANN was trained. Figure 5 shows the MAPE and MSE values achieved by ANN-BP models in forecasting the COVID-19 epidemic in the six districts tested. As can be seen, the proposed ANN-BP technique perform differently. In addition, using solely positive instances from the past day to estimate daily cases reported (the first model) does not produce credible estimates for the six districts studied.

To deside the two best ANN-BP technique, a rank method with uniform density for each precesion index was used to rate each ANN-BP technique model, as recommended by the literature (Fausett, 2006). This system determines which models have the better and worst results. Each ANN-BP technique model is ranked first according to the MAPE and MSE values obtained using this system. After then, the algebraic total of the two rank numbers was generated and used as a new benchmark for rank the performance of each ANN-BP model in each of the six districts. The conclusive findings of using this rank system for Adilabad, GHMC, Karimnagar, Khammam, Medchal Malkajigiri, Nalgonda and overall are provided in Table 5. The lower a model’s rank, the more precise it performs.

The above Figure 6 Represents performances of the network ANN models. For these models lower MSE values are 1.08, 9.630435, 7.0336364, 2.173913, 4.054545 and 4.0545454 respectively.

Based on the grading outcome in Table 5, the best model was chosen, and the associated forecasts were compared with the recorded data in Figure 7 for the Test data in various districts. In every figure in Figure 7, the actual versus predicted data of daily instances reported are displayed. In further sense, each point in Figure 7 has an abscissa and an ordinate indicate the positive case that was seen and predicted for the same day. The predicted value is closer to the observed value the fairly close a point gets to the identity line. Additionally, the farther one point digresses from the line of uniformity, the fewer precise the quantity of affirmed instances is assessed. The subtleties of accomplished outcomes are exhibited on its independently for six significant areas in the below:

3.1.1 | Estimating the number of COVID-19 instances pandemic in Adilabad

The Performances of the ANN-BP techniques model for Adilabad from 16 August to 31 October 2021 are presented in Figure 5a. In comparison to other ANN-BP technique models, the 5th, 13th, 12th, 8th, 14th models obtain near positive instances to the practical instances of Adilabad, as presented in Figure 5a and Table 5. Despite the fact that all forecasting models have marginal flaws, the ANN-BP models discussed above expected the number of daily confirmed COVID-19 cases to be close to the observed ones. The daily confirmed cases temporal changes seen in Figure 3, show substantial changes, these exact estimates are promising. Furthermore, the time period for estimating confirmed cases in Adilabad
is following the COVID-19 outbreak's first peak. This helps the performances of the ANN-BP techniques model by allowing them to become familiar with the outbreak's trend in train data. As a result, the ANN-BP techniques models mentioned above are recommended for predicting positive instances when the initial peak of the COVID-19 pandemic is observed in train data.

**FIGURE 3** Comparision of daily positive cases and ANN-BP model predicted cases of COVID-19.
3.1.2 | Estimating the number of COVID-19 outbreak in GHMC

For the test data, the exhibition of the ANN-BP technique type in anticipating every day instances detailed in GHMC is displayed in Figure 5b. In comparison to other models, the 5th, 14th, 9th, 7th and 8th ANN-BP models approximated the positive instances with a predictably high
As shown in Table 5 and Figure 5b, according to Figure 3, highest number of every day positive instances in GHMC is the same in the train and test data. This primarily gives ANN enough train statistics so that various ANN-BP technique types can anticipate accurate estimates. As a result, Figure 5b narrate the acceptable behaviour of multiple ANN types in forecasting confirmed instances from 16 August to 31 October 2021 in a schematic manner.

### 3.1.3 Estimating the number of COVID-19 pandemic in Karimnagar

Karimnagar was also one of the districts to be badly hit by COVID-19. From 16 August to 31 October 2021 are presentation of the ANN-BP technique models in anticipating affirmed instances of COVID-19 is displayed in Figure 5c. The precision obtained may be due to the time period of Karimnagar COVID-19 data used in this work; this spans the district’s first epidemic peak’s rising and decreasing limbs, as seen in Figure 3. Because, they achieve the lowest MSE and MAPE amid the ANN- technique models displayed in Table 5 and Figure 5c, the 11th, 10th, 12th, 13th and 9th models apper the first five best models.

### 3.1.4 Estimating the number of COVID-19 pandemic in Khammam

Khammam was also one of the districts to be badly hit by COVID-19. The presentation of the ANN techniques in anticipating affirmed instances of COVID-19 is displayed in Figure 5d. The precision obtained maybe due to the time period of Khammam COVID-19 data used in this study, this spans the district’s first epidemic peak’s rising and decreasing limbs, as seen in Figure 3 because they achieve the lowest MSE and MAPE among the ANN techniques displayed in the Table 5 and Figure 5d, the 14th, 5th, 7th, 9th, 11th models fared as well as the top five models.

### 3.1.5 Estimating the number of COVID-19 instances outbreak in Medchal

Based on Table 5 and Figure 5e, the 14th, 5th, 7th, 9th and 12th ANN models have the best performance in anticipating the COVID-19 pandemic in Medchal. The data period chosen for Medchal displays that the overall trend of the observations is monotonically decreasing, as seen in Figure 3 this means that the train data’s maximum was higher than the test data’s maximum. In comparison to other districts, this could lead to lower MAPE and higher MSE values.
Comparison of ANN-BP technique using MSE and MAPE.

(a) Adilabad
(b) GHMC
(c) Karimnagar
(d) Khammam
(e) Medchal
(f) Nalgonda

Figure 5
3.1.6 | Estimating the number of COVID-19 outbreak in Nalgonda

Figure 5f shows the criteria values for COVID-19 forecasted cases from 16 August to 31 October 2021. The 6th, 10th, 7th, 13th and 9th models outperform the others models, as seen in Table 5 and Figure 5f. Finally, Figure 5f shows that many ANN-BP based models can accurately forecast the COVID-19 outbreaks, mainly when a high number of cases must be calculated.

4 | DISCUSSION

As formerly stated, in this work, ANN-BP model was applied to forecast daily instances reported of COVID-19. First and foremost, when used only for forecasting within range of Train data, ANN often produces the best results. Predictions of upcoming confirmed cases, in particular, invariably include period values outside of the variety of the Train data. To be greater specific, this type of ANN-BP model application may be appropriate for short duration forecasting, such as more than a few days ahead. Short term forecasts, on the other hand, do not give healthcare decision-makers a full view of the COVID-19 pandemic. Furthermore, ANN might be utilized as a period series information expectation instrument, albeit such applications require the presence of a huge data set with the goal for ANN to catch potential examples. Nonetheless, no such data set exists for the COVID-19 scourge during this time-frame. These imperfections might restrict the relevance of ANN for COVID-19 pandemic expectation, notwithstanding the way that this work proposed 14 novel ANN-BP based models for this reason.

Following the application 14-days Artificial Neural Network models to anticipate the COVID-19 pandemic in six regions, the upsides of MSE and MAPE is utilized to work out the base and limit of two measurements for every model in significant six locales in India. Figure 5 portrays that the best normal MAPE and MSE esteems are accomplished for GHMC, though the most noteworthy normal MAPE and MSE esteems were acquired for Karimnagar and Medchal Malkajigiri, respectively. The outcomes likewise show that GHMC has most elevated MSE. This clearly shows that using only one metric to assess the presented of forecast results may be insufficient, but using multiple metrics may provide a more accurate view of the outcomes accuracy. As said by Figure 5, the whole performance of ANN-BP models for GHMC, Karimnagar, Medchal Malkajigiri which incorporate both the increasing and declining appendages of the main pinnacle, performs good Adilabad, Khammam, Nalgonda which only have a rising appendage as shown in Figure 3. This is due to the fact that the data used to train ANN-BP technique models contains the most positive cases, allowing them to make more accurate predictions.

Table 5 shows that, the 14-days model took top position in the position system, while the 7th and 5th, 9th days models were ranked second and third, respectively. The 14-day's model estimates confirmed cases in each day based on records from the previous 14-days, and the results show that this model can make the best forecasts. This is likewise the most extreme assessed hatching period for (O. O.-E. M. Journal and U, 2020) SAR-COV-2, which shows the execution of the incubation time frame into the forecast models might achieve the good assessments. At last, the fourteenth model is additionally recommended to anticipate day by day affirmed instances of COVID-19, especially for regions that have encountered the principal pinnacle of the pandemic.
The use of ANN-BP techniques to anticipate the COVID-19 virus revealed that best evaluation results were not obtained via the same technique in each of the districts studied. According to Table 5, ANN-BP models perform differently in different districts, resulting in varied ranking numbers. This could be owing to the fact that each district's COVID-19 outbreak record is somewhat unique, as shown in Figure 3 and different

**FIGURE 6** Performance of network different ANN-BP based model.
outburst drifts definitely request distinctive forecast models. In order to find the finest ANN-BP model for a specific district, an analysis similar to the one undertaken in this study is recommended though the fourteenth model is demonstrated to be sufficient without such investigation. The requirement for a district-wise prediction model is supported by earlier research (Rocklöv et al., 2020)–(Anam et al., 2021).

**FIGURE 7** Daily confirmed cases of observed versus predicted of COVID-19 for tested data in India different districts.
COVID-19 confirmed cases can be used to estimate not just the number of patients but also the number of facilities needed. Clearly, the more exact the affirmed cases are assessed, the better the future standpoint might be provided. Regardless of their conspicuous advantages, expectation models will undoubtedly have limits. Because these estimates are based on information provided, they undercount positive instances, either because asymptomatic individuals are ignored or because case finds are scarce in some districts (Anam et al., 2021). Besides, developing an ANN-BP technique forecast model that consider the former 14-days decreases the quantity of 14-day records in the ANN Train information. On the other hand, the longer a district has been exposed to COVID-19, the more data it has to work with. This clearly gives more information to AI models like ANN to prepare on, maybe bringing about more precise forecast results (Anam et al., 2021).

Comparison between Mathematical and ANN-BP models using the deviation from the target value (Figure 4) indicated that the ANN-BP model for six districts delivered the most accurate results. Extrapolation for long-term prediction of up to 75 days (16 August to 31 October 2021) using the ANN models were tested. The actual prediction of mathematical and ANN models for the six districts was reported which showed the progression of the outbreak.

This paper evaluated the applicability of two machine learning models, mathematical models (Boltzmann Function-based model and Beesham’s prediction model) and ANN models, for predicting the COVID-19 outbreak. The models showed promising results in terms of predicting the time series without the assumptions that mathematical models require. ANN models, as an alternative to mathematical models, showed potential in predicting COVID-19, as they did for modelling other outbreaks. Considering the availability of only a small amount of training data, it is expected that ANN will be developed further as the basis for, or a component of, future outbreak prediction models.

5 | CONCLUSIONS

Due to the high level of uncertainty and lack of crucial data, standard mathematical models have shown low accuracy for long-term prediction. This paper presents a comparative analysis of mathematical models and ANN-BP model to predict the COVID-19 outbreak. The results of ANN-BP model reported a high generalization ability for long-term prediction. With respect to the results reported in this paper and due to the highly complex nature of the COVID-19 outbreak and differences from district-to-district, this study suggests ANN as an effective tool to model the time series of outbreak. In this work, to construct the most precise ANN-BP model for estimating the COVID-19 epidemic and we consider 1 to 14 days ANN-BP models were presented and analysed. The results show that including COVID-19’s incubation period into ANN-BP prediction models resulted in more accurate estimations. The proposed ANN-BP model is recommended to assess rough quantities of day by day affirmed cases dependent on 14-day chronicled records, especially in the six areas that have encountered the primary pinnacle of the COVID-19 pandemic. This 14 days ANN-BP model was applicable to remaining all districts all areas in India. We should note that this paper provides an initial benchmarking to demonstrate the potential of ANN for future research.

ACKNOWLEDGEMENTS

This article is decided to dedicate to all medical providers, according to the authors. Who devotedly sacrifice their lives around the world during the COVID-19 epidemic we’d also want to thank the MHFW and Telangana state government for making the internet data available to researchers.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in #TelanganaFightsCorona Government of Telangana at https://covid19.telangana.gov.in/.

ORCID

Raja Das https://orcid.org/0000-0001-9945-5969

REFERENCES

A. B.-A. P. J. of T. Medicine and U. (2020). Mathematical modelling of COVID-19 in South Africa. Asian Pacific Journal of Tropical Medicine, 13(5), 235. https://www.apjtm.org/text.asp?2020/13/5/235/283519
Al-Najjar, H. N. A.-R.-E. R. for M. and U. 2020, “A classifier prediction model to predict the status of coronavirus COVID-19 patients in South Korea, 9518320322.
Al-Qaness, M. A. A., Ewees, A. A., Fan, H., Abd, M., & Aziz, E. (2020). Optimization method for forecasting confirmed cases of COVID-19 in China. Journal of Clinical Medicine, 9(3), 674. https://doi.org/10.3390/jcm9030674
Anam, S., Maulana, M. H. A. A., Hidayat, N., Yanti, I., Fitriah, Z., & Mahanani, D. M. (2021). Predicting the number of COVID-19 sufferers in Malang City using the backpropagation neural network with the fletcher–reeves method. Applied Computational Intelligence and Soft Computing. https://doi.org/10.1155/2021/6658552
Barda, N., Riesel, D., Akiv, A., Levy, J., & U. F.-N., and U. (2020). Developing a COVID-19 mortality risk prediction model when individual-level data are not available. Nature communications, 11(1), 1–9. https://doi.org/10.1038/s41467-020-18297-9
Brause, R. W., 2001, Medical analysis and diagnosis by neural Networks,” Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), Vol. 2199, pp. 1–13. https://doi.org/10.1007/3-540-45497-7_1

D. P.-J. B. Epidemiologi and U. (2019). Evaluation of formal risk assessment implementation of Middle East respiratory syndrome coronavirus in 2018. *Journal Berkala Epidemiologi*, 7, 197–206.

Dharani, N. and Kirupa Krishnan, G. 2021, ANN based COVID-19 prediction and symptoms relevance survey and analysis, Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021, pp. 1805–1808. https://doi.org/10.1109/ICCMC51019.2021.9418448.

F. M.-E.-P. K. S. B. Bali and U. 2017, “Komparasi Algoritma Conjugate gradient Dan gradient descent Pada MLPNN Untuk Tingkat Pengetahuan Ibu (Studi kasus Pemberian ASI Ekslusif),” *Knisstikom-baliadic*, 2017.

Fanelli, D., Chaos, F. P., & Fractals, and U. 2020, “Analysis and forecast of COVID-19 spreading in China, Italy and France, Elsevier, Vol. 134, pp. 109761. https://doi.org/10.1016/j.chaos.2020.109761

Fausett, L., 2006, Fundamentals of neural networks: Architectures, algorithms and applications.

Fu, X., Ying, Q., Zeng, T., Long, T., & Y. W.-J. of Infection, and U. (2020). Simulating and forecasting the cumulative confirmed cases of SARS-CoV-2 in China by Boltzmann function-based regression analyses. *Journal of Infection*, 8(5), 578–606.

Haykin, S., 2004, “A comprehensive foundation,” *IEEE*, 2, 41.

Hu, Z., Ge, Q., Li, S., & Xiong, M. (2020). Artificial intelligence forecasting of Covid-19 in China. *International Journal of Educational Excellence*, 6(1), 71–94. https://doi.org/10.18562/ijee.054

Ibrahim, I., Abdelmalek, D., Elshahat, M., A. E.-J. of Infection, and U. 2020, “COVID-19 spike-host cell receptor GRP78 binding site prediction, Elsevier, 80 (5), 554–562. https://doi.org/10.1016/j.jinf.2020.02.026

A. S.-J. M. Infotama and U. (2016). Jaringan Syaraf Tiruan Untuk Memprediksi Laju Pertumbuhan Penduduk Menggunakan Metode Backpropagation (Studis Kasus Di Kota Bengkulu). *Jurnal Media Infotama.*, 12(1). https://doi.org/10.3767/jmi.v12i1.273

Jha, G. (2007). “Artificial neural networks and its applications.” IARI.

Kavadi, D., Patan, R., Chaos, M. R., & S., and U. (2020). Partial derivative nonlinear global pandemic machine learning prediction of covid 19. Chaos, Solitons & Fractals, 139, 110056.

Niazkar, H. R., & Niazkar, M. (2020). Application of artificial neural networks to predict the COVID-19 outbreak. *Global Health Research and Policy*, 5(1), Dec. https://doi.org/10.1186/s41525-020-00175-Y

Niazkar, M. (2019). Revisiting the estimation of Colebrook friction factor: A comparison between artificial intelligence models and C-W based explicit equations. *KSE Journal of Civil Engineering*, 23(10), 4311–4326. https://doi.org/10.1007/s12205-019-2217-1

Niazkar, M., & Afzali, S. H. (2016). Streamline performance of excel in stepwise implementation of numerical solutions. *Computer Applications in Engineering Education*, 24(4), 555–566. https://doi.org/10.1002/CAE.21731

Niazkar, M., & Afzali, S. H. (2019). Developing a new accuracy-improved model for estimating scour depth around Piers using a hybrid method. *Iranian Journal of Science and Technology - Transactions of Civil Engineering*, 43(2), 179–189. https://doi.org/10.1007/s40996-018-0129-9

Niazkar, M., & H. N.-E. J. of G. Medicine, and U. (2020). Covid-19 outbreak: Application of multi-gene genetic programming to country-based prediction models. *Electronic Journal of General Medicine*, 17(5). https://doi.org/10.29333/ejmg.8232

Niazkar, M., Talebbeydokhti, N., & Afzali, S. H. (2019a). Novel grain and form roughness estimator scheme incorporating artificial intelligence models. *Water Resources Management*, 33(2), 757–773. https://doi.org/10.1007/s11269-018-2141-Z

Niazkar, M., Talebbeydokhti, N., & Afzali, S. H. (2019b). One dimensional hydraulic flow routing incorporating a variable grain roughness coefficient. *Water Resources Management*, 33(13), 4599–4620. https://doi.org/10.1007/s11269-019-02384-8

Niazkar, M., Türkkan, G. E., Niazkar, H. R., & Türkkan, Y. A. (2020). Assessment of three mathematical prediction models for forecasting the covid-19 outbreak in Iran and Turkey. *Computational and Mathematical Methods in Medicine*, 2020, 1–13. https://doi.org/10.1155/2020/7056285

Nishiura, H., Kobayashi, T., Yang, Y., K.-H. of clinical, and U. 2020, The rate of underascertainment of novel coronavirus (2019-nCoV) infection: Estimation using Japanese passengers data on evacuation flights, mdpicom, 9(2), 419.

O. O.-M. S. and U. 2020, Coronavirus disease 2019 (COVID-19): Diagnosis and management, gogalecom, 288(2), 192–206.

Ratnawati, D. E., Marjono, M., Widodo, N., & Anam, S. (2019). Features selection for classification of SMILES codes based on their function. *International Seminar on Research of Information Technology and Intelligent Systems* (pp. 103–108). IEEE. https://doi.org/10.1109/ISRITI48646.2019.9034619

Rockløv, J., Sjödin, H., & A. W.-S.-J. of travel Medicine, and U. (2020). COVID-19 outbreak on the diamond princess cruise ship: Estimating the epidemic potential and effectiveness of public health countermeasures. *Journal of travel medicine*, 27(3). taaa030. https://doi.org/10.1093/jtm/taaa030

Roosa, K., Lee, Y., Luo, R., & A. K.-J. of clinical, and U. 2020. Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. *Journal of clinical medicine*, 9(2), 596.

Rumelhart, D., Hinton, G., & R. W.- Nature, and U. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

Wang, S.-H. and Y.-D. of Cancer, and U. 2020. Deep rank-based average pooling network for Covid-19 recognition. *Computers, Materials and Continua*, 70(2), 2797–2813. https://doi.org/10.32604/CMC.2022.020140

Wang, S.-H., Zhang, X., & Zhang, Y.-D. (2021). DSSAE: Deep stacked sparse autoencoder analytical model for COVID-19 diagnosis by fractional Fourier entropy. *ACM Transactions on Management Information Systems*, 13(2), 1–20. https://doi.org/10.1145/3451357

WHO. 2020, Coronavirus (COVID-19) Dashboard/WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. *https://covid19.who.int/* (accessed May 06, 2022).

**AUTHOR BIOGRAPHIES**

Balakrishnama Manohar is a full time PhD, Research Scholar (Institute Fellowship Category) in the Department of Mathematics at Vellore Institute of Technology Vellore, Tamilnadu–632014, India. He has completed his Masters in Mathematics from Sree Vidyanikethan PG College, Tirupati, A.P. (Affiliated to S. V. University), India. He is M.Sc. Gold Medalist. His research interest includes Machine learning and Deep Learning.
Raja Das is working as an Assistant Professor (Department of Mathematics) at Vellore Institute of Technology Vellore, Tamilnadu –632014, India. He had published more than 68 papers in various International Journal and Conferences. His research interest includes Medical Imaging, Electrical Discharge Machining, Artificial Intelligence, Machine Learning and Deep Learning.

How to cite this article: Manohar, B., & Das, R. (2023). Artificial neural networks for prediction of COVID-19 in India by using backpropagation. Expert Systems, 40(5), e13105. https://doi.org/10.1111/exsy.13105