

## Research Article

# Nonlinear Cosine Neighborhood Time Series-Based Deep Learning for the Prediction and Analysis of COVID-19 in India

Ashok Kumar Munnangi,<sup>1</sup> Ramesh Sekaran,<sup>1</sup> Sivaram Rajeyyagari ,<sup>2</sup>  
Manikandan Ramachandran ,<sup>3</sup> Suthendran Kannan,<sup>4</sup> and Subrato Bharati <sup>5</sup>

<sup>1</sup>Department of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

<sup>2</sup>Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra, Saudi Arabia

<sup>3</sup>School of Computing, SASTRA Deemed University, Tamil Nadu, India

<sup>4</sup>Department of Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India

<sup>5</sup>Institute of Information and Communication Technology (IICT), Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

Correspondence should be addressed to Subrato Bharati; [subratobharati1@gmail.com](mailto:subratobharati1@gmail.com)

Received 10 May 2022; Revised 3 June 2022; Accepted 10 June 2022; Published 29 June 2022

Academic Editor: Kalidoss Rajakani

Copyright © 2022 Ashok Kumar Munnangi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The widening of *coronavirus disease* (COVID-19) across the globe has put both the government and humanity at risk. The funds of part of the biggest recessions are stressed out due to the severe infectivity rates and highly communicable nature of this disease. Due to the expanding consequence of cases being registered and their successive significance on the civic body administration and health professionals, certain prediction methods are intended to be necessitated to predict the number of cases in the future. In this paper, nonlinear cosine-based time series learning (NCTL) is introduced for the prediction and analysis of COVID-19 in India. First, the *nonlinear least squares regressive feature selection* (NLS-RFS) model is used for choosing the relevant features by considering both the active cases with less prediction error. Next, the cosine-based neighborhood filter algorithm is applied to attain the optimum filtered features to select relevant features with minimum prediction time. Finally, cosine neighborhood-based LSTM is used for the prediction of the number of COVID-19 cases being registered in India to the fore and consequence of precautionary measures like social distancing, lockdown, and declaring containment zones on the outspread of COVID-19. The existing deep learning methods' prediction accuracy was not enhanced with lesser time. In order to overcome the issue, the nonlinear cosine-based time series learning (NCTL) method has been introduced. The aim of the proposed NCTL method is to predict the number of COVID-19 cases with less prediction time and prediction error. This helps to enhance the prediction accuracy for considering the time series with accurate prediction results. The experiment of the NCTL method is conducted using metrics such as accuracy, prediction error, prediction accuracy, and prediction time with respect to diverse samples. The simulation result illustrates that the NCTL method increases the prediction accuracy by 8%, reduces the prediction time by 18%, and minimizes the prediction error by 31% compared to state-of-the-art works in a computationally efficient manner.

## 1. Introduction

The pandemic of the coronavirus disease 2019 (COVID-19) rapidly escalated all over the globe in its initial stages. This was due to the maximal uncertainty of the elementary phase of the outbreak, the stage of the pandemic, the finite percep-

tion of the novel coronavirus by the medical association, and the absence of medical resources. Public understanding of spread avoidance has progressively escalated due to the significant measures taken for prevention and control.

First, as conventional outbreak models treat all individuals with coronavirus as acquiring a similar infection rate,

an improved susceptible infected (ISI) was proposed in [1] to evaluate the different types of rates of infection by considering the development tendency. Next, taking into account the aftermaths of prevention and control measures and the upsurge of the avoidance consciousness, the natural language processing (NLP) module and the long short-term memory (LSTM) network are ingrained into the ISI to construct the hybrid artificial intelligence (AI) model for successfully predicting COVID-19. With this, the prediction error results were minimized to a greater extent.

A linear regression was performed in [2] for predicting the spread of COVID-2019. Here, multilayer perceptron and vector autoregression were designed to forecast the ideological sample of the disorder and swiftness of COVID-2019 cases in India. With the data obtained using Kaggle, the possible samples of COVID-19 consequences in India were recorded. With this, the confirmed cases, death cases, and recovery cases covering India aided in forecasting and judging the future cases to a certain extent, therefore improving the prediction results. But, the deep learning methods were not used for forecasting time series data for getting better predictions.

*1.1. Problem Definition.* Most of the existing prediction methods have been designed for COVID-19. But, the accurate prediction was not enhanced, and time consumption was not reduced. In addition, the updated dataset was less focused. Then, the existing prediction method failed to recognize patients at different risks. But, the feature selection was not performed. To overcome the existing issue, the nonlinear cosine-based time series learning (NCTL) method is introduced to increase the accurate prediction with less time.

*1.2. Motivation and Our Contributions.* Coronavirus disease (COVID-19) is a risky disease caused by the SARS-CoV-2 virus. Several research works have been developed for COVID using AI models and machine learning models. However, it failed to use deep learning methods for forecasting time series data. But, the prediction time was not reduced. In conventional methods, the error rate was not minimized. In several existing methods, prediction accuracy was not improved. Motivated by this, the nonlinear cosine-based time series learning (NCTL) method is presented using enhanced deep learning to accurately predict the number of COVID-19 cases in a timely and accurate manner. NCTL method is designed with the novelty of NLS regressive feature selection, cosine-based neighborhood filter, and cosine time series-based LSTM. The contributions include the following:

- (i) To reduce the prediction time by utilizing a cosine-based neighborhood filter algorithm that outputs relevant features considering both the active cases
- (ii) To minimize the prediction error by proposing a nonlinear regressive feature selection that outputs more relevant features required for early prediction, to control the cases being registered and also the death rate to a certain extent

- (iii) To improve the prediction accuracy with the aid of cosine neighborhood-based LSTM that improves the accuracy rate of being predicted by considering the time series involved

*1.3. Article Structure.* The rest of the paper is organized as follows. Section 2 presents the related works. Section 3 proposes the method nonlinear cosine-based time series learning (NCTL) for the prediction and analysis of COVID-19. Section 4 provides discussions on the results and discussion with the aid of table and graph. Finally, Section 5 concludes the work.

## 2. Related Works

An upsurge in COVID-19 detection is giving rise to an enigma for research persons, civic bodies, and healthcare workers. As of July 24, 2020, 15.7 million cases have been confirmed, over 638K is dead, and about 8.98 million are reported recovered. Not only economies are ramming but also the comprehensive toughness and morality of the deliberately affected nations are being conceded. To predict the disease accurately, considering the typical advancement of disease is very predominant.

A deep learning model was designed in [3] to enhance the reported cases' accuracy rate and predict the disease in a precise manner via convolutional neural networks (CNNs). With these, not only abnormalities in the structure were observed, but also the classification of disease was made in a significant manner forming keys to obtaining the hidden patterns. Besides, transfer learning was also applied for early detection. A study concentrating on enhanced mathematical modeling for early analysis and prediction of the epidemic was proposed in [4].

Also, a machine learning-based model was proposed to predict the potential threat of COVID-19. Finally, the efficient prediction was performed using the generalized inverse Weibull distribution. Technology developments have a swift consequence on all aspects of life, be it medical or any other real-time field. The application of artificial intelligence (AI) has exhibited optimistic results in the medical field using its decision-making by rigorous data analysis. COVID-19 has created a greater impact in more than 100 countries, with people around the globe being susceptible to its significance in the future.

Feature engineering was performed in [5] for better accuracy. A survey of the proposal for the people who are in the recent days struggling with the worldwide COVID-19 pandemic was presented in [6]. Practically, one and all across the world have been influenced by the COVID-19 widespread emerging from Wuhan China and require speedy solutions. AI in response to COVID-19 was proposed in [7].

In [8], a new susceptible-exposed-infected-recovered (SEIR) epidemic model was utilized to analyze the local and global stability by considering the influence of healthcare. A review of machine learning techniques was analyzed in [9] for COVID-19. To predict the death rate in India and to control the death rate, analysis of linear and

multiple regressions were made in [10] to enhance the predictive ability of regression in a significant manner.

Another method to avoid the spread of COVID-19 was proposed in [11] with the aid of artificial intelligence for tracking the death, case registered, and recovered. A broad-ranging assessment of coronavirus disease was investigated in [12]. In the previous few months, several research articles were printed in consideration of the strong and early COVID-19 detection with the aid of mathematical modeling and AI techniques. The objective of [13] remained in bestowing the research community with a complete outline of the models utilized in addressing early predictions of COVID-19.

It is now apparent that the world requires a swift and faster solution to accommodate and handle the supplementary spread of COVID-19. In [14], models about data mining were utilized for COVID-19 prediction patients in South Korea. Certain data mining models like decision trees, random forests, and  $K$ -nearest neighbor algorithms were related to the dataset. The model predicted the day required for recovery, high-risk age group patients, and so on in a precise manner. Some of the disadvantages and pitfalls involved in the application of AI for COVID-19 prediction were analyzed in [15].

AI-driven sensing prototypes to uproot surveillance in a real-time fashion acquired from social media users were proposed in [16]. An in-depth analysis using regression was conducted in [17] for February in China to control the spread. Based on the data collected via an official platform, the transmission process of the coronavirus disease 2019 (COVID-19) was analyzed in [18]. The error rate between the actual and measured was predicted by the epidemic situation. With this analysis results, efficient decision-making was said to be made. Supervised machine learning models were applied in [19, 20] for efficient forecasting of new cases, death rate, and recovery. The COVID-19 prediction model [21] is based on time-dependent SIRVD using deep learning. However, only using artificial intelligence methods for the prediction cannot capture the time change pattern of the transmission of infectious diseases. The hybridization of GCN and GRU models in the mRNA degradation fields [22] to predict mRNA sequences' stability/reactivity and degradation risk.

### 3. Methodology

With the increase in new cases in recent months in India, a prediction method to control the escalating rate of COVID-19 in India is the need of the hour. Though the fatality and recovery rates are higher, with the increase in new cases, fear of jobs being wiped out in every sector has created a panic among the public. Therefore, designing a prediction method would assist the health professionals and civic body administration in reducing the fear of the public. In this section, a nonlinear cosine-based time series learning (NCTL) method for the prediction and analysis of COVID-19 is presented. The proposed nonlinear cosine-based time series learning (NCTL) method comprises three models; there are (1) feature selection model, (2) filter module, and (3) local

trend prediction. The block diagram of the nonlinear cosine-based time series learning (NCTL) method is illustrated in Figure 1.

Figure 1 demonstrates the block diagram of the nonlinear cosine-based time series learning (NCTL) method. At first, the COVID-19 in India dataset is taken as input. Next, NLS regressive feature selection is applied to pick the important features, followed by, cosine-based neighborhood filter is to achieve an optimal filter by using significant features. Finally, cosine time series-based LSTM is utilized to enhance prediction accuracy for computationally efficient and accurate prediction.

*3.1. Nonlinear Least Squares (NLS) Regressive Feature Selection Model.* Based on the objective of the proposed NCTL method, experimental parameters such as prediction accuracy, prediction error, and prediction time are selected for experimental purposes. In our work, the nonlinear regressive feature selection algorithm is applied in the proposed NCTL method by choosing relevant features. Cosine-based neighborhood filter algorithm is used to select the relevant features to forecast the precise outcome. This process helps to improve prediction accuracy and reduce prediction time. We postulate that all the features are not associated with the prediction variable. As far as the COVID-19 dataset are concerned, three main interests have to be taken into account for the risk classification of a country. They are the number of cases ' $\alpha$ ,' the number of deaths ' $\beta$ ,' and the number of recovered ' $\gamma$ .' With these three interests concerned, in our work, relevant feature selection for risk classification is performed by using the nonlinear least-squares regressive feature selection (NLS-RFS) model. Figure 2 below shows the block diagram of the nonlinear least squares (NLS) regressive feature selection model.

Figure 2 demonstrates the block diagram of the nonlinear least squares (NLS) regressive feature selection model. With the input COVID-19 in the India dataset provided as input, the first active cases are measured. Next, the residual function is computed. Followed by, the approximation is carried out for the selection of the relevant features. Let us consider the COVID-19 in the India dataset provided as input. Then, the active cases ' $AC$ ' are estimated as given below.

$$AC \longrightarrow \alpha - [\gamma + \beta]. \quad (1)$$

From the above equation (1), ' $AC$ ' indicates the active cases. The number of cases denotes ' $\alpha$ ,' the number of deaths is represented as ' $\beta$ .' Then, the essential features are selected using the reverse suppression model. We calculate ' $\delta$  value' of all features with ' $AC$ ' using nonlinear least squares (NLS) regression. Let us consider a set of data representations ' $(f_1, y_1), (f_2, y_2), \dots, (f_n, y_n)$ ' and a model function ' $y = \text{fun}(f, \delta)$ ' that in addition to the feature ' $f$ ' depends on ' $m$ ' factors ' $\delta = (\delta_1, \delta_2, \dots, \delta_m)$ ', where ' $n > m$ ,' it is required to identify the feature vector (relevant features) that the curve best fits the given data. This is mathematically formulated as given below.

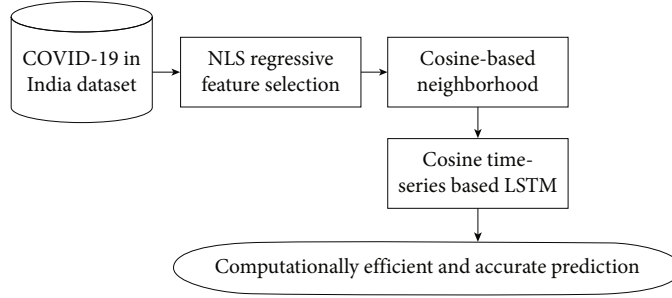


FIGURE 1: Block diagram of the nonlinear cosine-based time series learning (NCTL) method.

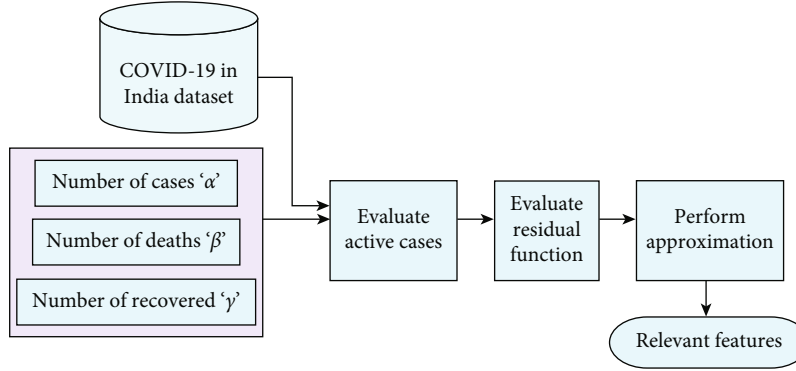


FIGURE 2: Block diagram of the nonlinear least squares (NLS) regressive feature selection model.

$$\text{SofS} = \sum_{i=1}^n r_i^2, \text{ where } i = 1, 2, \dots, n. \quad (2)$$

From the above equation (2), the sum of squares ‘SofS’, for each feature is arrived at based on the residual function ‘ $r_i$ .’ It is mathematically evaluated as given below.

$$r_i = y_i - \text{fun}(f_i, \delta). \quad (3)$$

With the assumption that the minimum value of ‘SofS’ occurs when the scalar differential function ‘fun’ of several features with respect to partial derivatives is ‘0,’ the gradient for ‘ $n$ ’ data representations is then formulated as given below.

$$\partial \text{SofS} = \partial \delta_j * \left[ 2 \sum r_i \left( \frac{\partial r_i}{\partial \delta_j} \right) \right], i = (1, 2, \dots, m). \quad (4)$$

From the above equation (4), as the model includes ‘ $m$ ’ data representations, in our case, ‘ $n = 7$ ’ (i.e., Age-GroupDetails, HospitalBedsIndia, ICMRTTestingLabs, IndividualDetails, StatewiseTestingDetails, Covid19India, and PopulationIndiaCensus), at each iteration approximation is performed to choose relevant features. This is mathematically formulated as given below.

$$\text{fun}(f_i, \delta) = \text{fun}(f_i, \delta^k) + \sum \frac{\partial \text{fun}(f_i, \delta^k)}{\partial \delta_j} \quad (5)$$

$$= \text{fun}(f_i, \delta^k) + \sum \text{DM}_{ij}. \quad (6)$$

From the above equation (5), the relevant features ‘ $\text{fun}(f_i, \delta)$ ’ are selected based on the input variable ‘ $\text{fun}(f_i, \delta^k)$ ’ and the data representation in the form of data matrix ‘ $\text{DM}_{ij}$ ,’ respectively. Here, the data matrix includes representations of seven files. The pseudocode representation of nonlinear regressive feature selection is given below.

As given in the above nonlinear regressive feature selection algorithm, with the number of cases, death, and recovered patients as input, the objective remains in obtaining remains relevant filters. The active cases are calculated for each case, death, and recovery, followed by the sum of the square is computed based on the residual function. The partial derivatives for ‘ $n$ ’ data representations are performed to select the relevant features. The key features, i.e., the number of new cases reported with respect to the previous day’s infections. This is obtained by calculating ‘ $\delta$  value’ which includes previous day infections (with respect to the number of cases, death, and recovered) for a specified region.

**3.2. The Cosine-Based Neighborhood Filter Function.** With the most relevant features obtained, the second step for early COVID-19 prediction in a specified area is the rate of spread or controlling the spread rate. In this work, a cosine-based neighborhood filter function is applied to the relevant feature set. The neighborhood-based filter evaluates the similarity between two cases (i.e., swab tests obtained from two

```

Input: number of cases ‘ $\alpha$ ,’ number of deaths ‘ $\beta$ ,’ and number of recovered ‘ $\gamma$ ’
Output: relevant features ‘ $RF = rf_1, rf_2, \dots, rf_n$ ’
1: Initialize data representations ‘ $(f_1, y_1), (f_2, y_2), \dots, (f_n, y_n)$ ,’ model function ‘ $y = \text{fun}(f, \delta)$ ’
2:   Begin
3:     For each case, ‘ $\alpha$ ,’ deaths ‘ $\beta$ ,’ and recovered ‘ $\gamma$ ’
4:       Evaluate active case ‘AC’ using equation (1)
5:       Evaluate the sum of squares ‘SofS’ using equation (2)
6:       Evaluate partial derivatives for ‘ $n$ ’ data representations using equation (4)
7:       Obtain relevant features using equation (6)
8:     Return relevant features (RF)
9:   End for
10:  End

```

ALGORITHM 1: Nonlinear regressive feature selection.

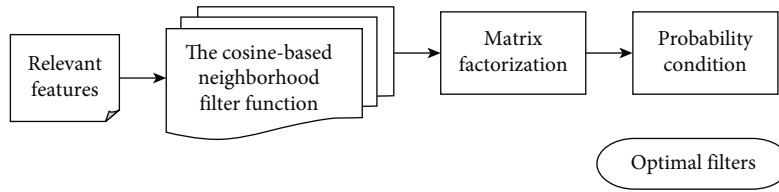


FIGURE 3: Block diagram of cosine-based neighborhood filter function.

different cases residing in a similar region) so that the global outbreak can be tracked in a significant manner.

With this filter, by considering the weighted average of all the total samples, the similarity between two neighboring users is obtained. With this, as by controlling the neighborhood cases, the global outbreak can be brought to control. Hence, a cosine-based neighborhood filter function is applied to the acquired relevant features so that optimal filters are obtained. Figure 3 shows the block diagram of the cosine-based neighborhood filter.

Figure 3 explains the block diagram of the cosine-based neighborhood filter function. With relevant features provided as input. First, the similarity between two swab tests is determined by using cosine-based neighborhood filter function. Next, the matrix factorization is used to find the matrix representation of cases and individual details. Finally, the conditional probability functions are employed to consider the next outcome and the previous outcome results to obtain the optimum filters.

Let us assume the relevant features as input. First, the similarity between two swab tests ‘ $S$ ’ obtained from ‘ $P$ ’ and ‘ $Q$ ’ is mathematically evaluated [23] as given below.

$$\begin{aligned}
 \text{SIM}(P, Q) &= \text{COS}(P', Q') = \frac{P' * Q'}{|P'| * |Q'|} \\
 &= \frac{\sum_{i \in \text{RF}_{P,Q}} S_{P,i} S_{Q,i}}{\sqrt{\sum_{i \in \text{RF}_P} S_{P,i}^2} \sqrt{\sum_{i \in \text{RF}_Q} S_{Q,i}^2}}.
 \end{aligned} \quad (7)$$

From the above equation (7), the similarity between two swab tests ‘ $S$ ’ obtained from ‘ $P$ ’ and ‘ $Q$ ’ is acquired based on the set of relevant features ‘ $\text{RF}_{P,Q}$ ’ sampled from two cases. The above function evaluates the similarity between two

swab tests and generates a prediction for the samples by considering the weighted average of all the related features. We follow the neighborhood filter-based proposition; the matrix representation of cases and individual details are broken down by matrix factorization (MF). By representing a set of values corresponding to a set of cases with index ‘ $i = \{1, 2, 3, \dots, N\}$ ’ to a set of individual details with index ‘ $j = \{1, 2, 3, \dots, M\}$ ’, each entry of the value ‘ $A_{ij}$ ’ is then indicated as the inner product of a cases matrix and individual details matrix as given below.

$$A_{ij} = (P_i, Q_j) = \text{SIM}(P, Q) \left[ \sum P_{ik} Q_{kj} \right]. \quad (8)$$

From the above equation (8), ‘ $P_i, Q_j$ ’ represent the ‘ $k$ ’ case-specific and individual detail of each case-specific vectors of swab test of the case ‘ $i$ ’ and individual details ‘ $j$ ’, respectively. Here ‘ $P_{ik}$ ’ represents the 9 columns (from covid\_19\_India) and ‘ $Q_{kj}$ ’ represents the 12 columns (from Individual Details), respectively. This work reduces the time and overhead involved by using conditional probability functions, which include the case and individual detail matrix as given below.

$$\text{Prob}A_{ij} = \text{Prob} [A_{ij}(\text{next}) | A_{ij}(\text{prev})]. \quad (9)$$

From the above equation (9), ‘ $\text{Prob}A_{ij}$ ’ denotes the probability function. By utilizing the above function, the parameter sets are predicted parameters utilizing probability function involving both next outcome ‘ $A_{ij}(\text{next})$ ’ and the previous outcome ‘ $A_{ij}(\text{prev})$ ’, respectively, therefore forming optimum filters. The pseudocode representation of a cosine-

```

Input: relevant features 'RF = rf1, rf2, ..., rfn'
Output: optimal filtered features 'FF = ff1, ff2, ..., ffn'
1: Initialize swab 'S, where P, Q ∈ S'
2:   Begin.
3:     For each relevant feature 'RF'
4:       Evaluate cosine-based neighborhood filter function using equation (7)
5:       Evaluate probability conditional functions using equation (9)
6:     Return (Optimal filtered features)
7:   End for
8:   End

```

ALGORITHM 2: Cosine-based neighborhood filter.

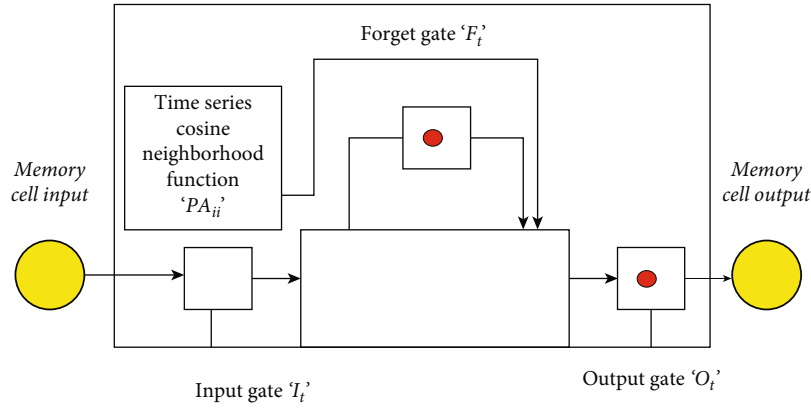


FIGURE 4: Structure of time series cosine LSTM.

based neighborhood filter for acquiring optimal filters is given below.

As given in the above cosine-based neighborhood filter, for each relevant feature given as input, the objective remains to obtain optimum filters. In other words, only with the aid of relevant features prediction of COVID-19 is made for different cases all over the country and hence across the globe. In addition to relevant features, two important factors have to be considered to minimize the complexity involved in time and overhead. They are similarities between neighboring users, and the next-previous outcomes should also be considered. This is done in our work by introducing a cosine-based neighborhood filter function and then performing probability conditions. With these two factors, optimum filters are obtained in a computationally efficient manner.

**3.3. Cosine Neighborhood-Based LSTM for COVID-19 Prediction.** Time series analysis for COVID-19 case prediction assists in forecasting the possible future cases that could be registered based on the last diagnosis data. In this work, a model called cosine neighborhood-based LSTM for COVID-19 prediction is proposed. The model uses the time series cosine neighborhood function to predict future health risks. For predicting any disease at an early stage, LSTMs are regarded to be among the most practicable solutions. They predict the upcoming forecasts depending on different underscored features present in the dataset. With the assistance of conventional LSTMs [24], the data (i.e., optimal fil-

tered features) advances via components referred to as cell states.

In our work, COVID-19 in India dataset is considered an input as the epidemic outbreak is increasing day by day and also with the varying time frames. In this work, the time series cosine neighborhood function is applied to cell state in conventional LSTM, forming time series cosine LSTM. In this sort of design, the model passes the previous state to the next stage of the arrangement, following temporal patterns. The structure of time series cosine LSTM consists of four gates. They are input gate ' $I_t$ ', forget gate ' $F_t$ ', control gate ' $C_t$ ', and output gate ' $O_t$ ' which is shown in Figure 4.

As shown in the figure with three gates forming the time series cosine LSTM, the input gate is mathematically expressed as given below.

$$I_t = \text{Sigmoid}(W_I[h_{t-1}, FF_t]) + B_I. \quad (10)$$

From the above equation (10), ' $I_t$ ' forms the input gate function and ' $W_I$ ' and ' $B_I$ ' represents the weight and bias of neuron at ' $I$ ' with results from the previous step obtained via ' $h_{t-1}$ ' and the input filtered features represented by ' $FF_t$ ', respectively, at the time ' $t$ .' Next, form the forget gate that resolves which filtered feature to be moved to the cell. The information or the filtered feature from the input of preceding memory which is unused by the forget gate is mathematically formulated as given below.

$$F_t = \text{Sigmoid}(W_F[h_{t-1}, FF_t]) + B_F. \quad (11)$$

```

Input: optimal filtered features 'FF = ff1, ff2, ..., ffn'
Output: accurate prediction results
1: Initialize weight 'WI, WF, WO', bias 'BI, BF, BO', 'α', 'β', 'γ', 'A', 'TotPop'.
2:   Begin
3:     For each optimal filtered feature 'FF'
4:       Form input gate function using equation (10)
5:       Form forget gate function using equation (11)
6:       Form control gate function using equation (12)
7:       Form output gate function using equation (13)
8:       Evaluate death rate using equation (15)
9:       Evaluate the case registered using equation (16)
10:      Evaluate recovery rate using equation (17)
11:     Return (prediction results)
12:   End for
13: End

```

ALGORITHM 3: Cosine neighborhood-based LSTM.

From the above equation (11), ' $F_t$ ' forms the forget gate function and ' $W_F$ ' and ' $B_F$ ' represent the weight and bias of neuron at ' $F$ '. Also, the cell update is under the control of the control gate. It is mathematically formulated as given below.

$$C_t = F_t * C_{t-1} + PA_{ij}. \quad (12)$$

From the above equation (12), the cell update is performed by the control gate ' $C_t$ ' using the forget gate results ' $F_t$ ' and the time series cosine neighborhood function ' $PA_{ij}$ ', respectively. Finally, the hidden layer ' $h_{t-1}$ ' is updated by the output layer which is also accountable for upgrading the output and is mathematically formulated as given below.

$$O_t = \text{Sigmoid}(W_O[h_{t-1}, FF_t]) + B_O, \quad (13)$$

$$h_t = O_t * \text{Tanh}(C_t). \quad (14)$$

As given in the above equations (13) and (14), the output gate ' $O_t$ ' integrates the cell state information output and the forget gate information output at the time ' $t$ '. Finally, reasoning-based risk categorization is modeled as given below based on death rate 'DR,' case registered 'CR,' and recovery rate 'RR,' respectively. It is mathematically formulated as given below.

$$\text{DR} = \frac{\alpha}{\beta} \quad (15)$$

$$\text{CR} = \frac{\text{TotPop}}{\text{AC}}, \quad (16)$$

$$\text{RR} = \frac{\text{AC}}{\gamma}. \quad (17)$$

From the above equations (15)–(17), with the aid of a number of cases ' $\alpha$ ,' number of deaths ' $\beta$ ,' number of recovered ' $\gamma$ ,' active cases 'AC,' and total population 'TotPop', the risk is analyzed. The pseudocode representation of cosine neighborhood-based LSTM is given below.

TABLE 1: Performance measurements for prediction error for different prediction methods.

Samples obtained	Prediction error (%)		
	NCTL	Hybrid AI model	Linear regression forecasting model
250	8	12	16
500	9.25	14.35	19.35
750	9.45	15.25	20.15
1000	9.85	18.15	23.25
1250	11.35	20.35	24.55
1500	12.45	22.45	25.35
1750	14.35	25.45	30.15
2000	14.85	26.15	33.45
2250	16.25	30.35	35.15
2500	19.15	33.45	40.23

As given above, the cosine neighborhood-based LSTM algorithm improves the accuracy of COVID-19 prediction. The outbreak in any state or district in a country takes typically place at distinct extents of immensity with respect to time, time series is first considered. Next, with the optimal filtered features provided as input for varied time factors, an LSTM is designed, and according to three different factors, i.e., death rate, case registered, and recovery rate, prediction analysis is made. This helps to improve accurate prediction results in a significant manner.

#### 4. Experiments and Results

The proposed nonlinear cosine-based time series learning (NCTL) method is implemented using the JAVA SDK toolkit with customized security packages. To evaluate the appropriateness of the NCTL method, the method was applied to a benchmark dataset COVID-19 in the India dataset obtained from <https://www.kaggle.com/sudalairajkumar/covid19-in-india>.

This dataset possesses information about the population with respect to age group, details regarding hospital beds made ready for the usage of COVID-19 patients,

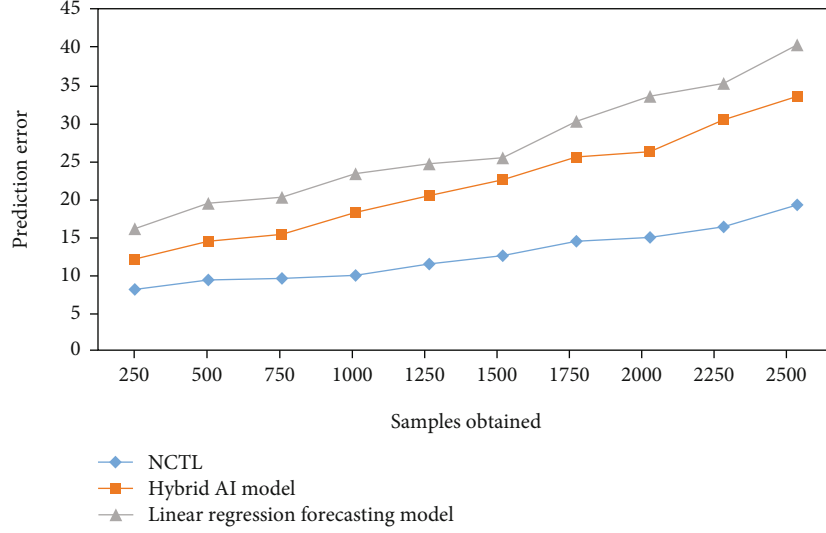


FIGURE 5: Graphical representation of prediction error.

information regarding Indian Council of Medical Research Lab, and individual details of the COVID-19 patients, including information like diagnosis date, age, gender, detected city, detected district, detected state and so on, population details, and testing details made in state wise manner obtained from ministry of health and family welfare. The number of new cases is enhancing day by day around the world. This dataset has information from the states and union territories of India at the daily level. The dataset includes 8067 samples. The dataset comprises 5 attributes, including date, state, total samples, negative, and positive. Experimental evaluation was carried out on certain factors such as the COVID-19 prediction accuracy and COVID-19 prediction time for the different number of samples obtained at varied time periods. Based on the objective of the proposed method (i.e., focused on accurate COVID-19 prediction with minimum prediction time) the existing methods, such as the AI model [1] and linear regression forecasting model [2] are taken as base paper. These two base papers are explained to understand the proposed method. The proposed method concept is derived by considering the problems of these base papers. The drawbacks of these methods are effectively convinced by implementing the proposed method.

*4.1. Case 1: Prediction Error.* Prediction errors can be assessed in different types, depending on the application used. For COVID-19 prediction and diagnosis, prediction error plays a major role because by reducing the prediction error, the greater number of fatalities and recovery cases may also be minimized. In other words, it is an estimate of how well samples obtained are predicted to the correct category. This is mathematically expressed as given below.

$$PE = \sum_{i=1}^n \frac{[MV - PV]}{S_i} * 100. \quad (18)$$

From the above equation (18), prediction error ‘PE’ is measured based on the measured value ‘MV,’ predicted

TABLE 2: Performance measurements for prediction accuracy for different prediction methods.

Samples obtained	Prediction accuracy (%)		
	NCTL	Hybrid AI model	Linear regression forecasting model
250	98	96.8	96
500	97.85	93.82	90.15
750	97.55	93.45	89.35
1000	97.25	93	88.15
1250	97	92.85	88
1500	96.85	92.65	86.25
1750	97	93.35	86
2000	97	93.55	85.35
2250	97.35	94	85
2500	97.25	94.25	84.15

value ‘PV’ with respect to the samples ‘ $S_i$ ’ considered. It is measured in terms of percentage (%) and summaries of the performance measurement for prediction error for the three methods as mentioned above, NCTL, hybrid AI model [1], and linear regression forecasting model [2], are found in Table 1.

The findings in the above table indicate that the proposed NCTL method outperforms the hybrid AI model [1] and linear regression forecasting model [2] in terms of prediction error.

Figure 5 above depicts the prediction error with respect to 2500 swab samples collected from the state of Tamil Nadu within two months for different age groups between 20 and 40 years. These samples were also obtained from different periods. From the figure with the increase in the swab samples collected, the prediction error is also found to be increasing linear using all three methods. This is because the reason that the swab samples were not obtained in a single day but collected at different periods and also due to the varying climatic conditions in different districts in a



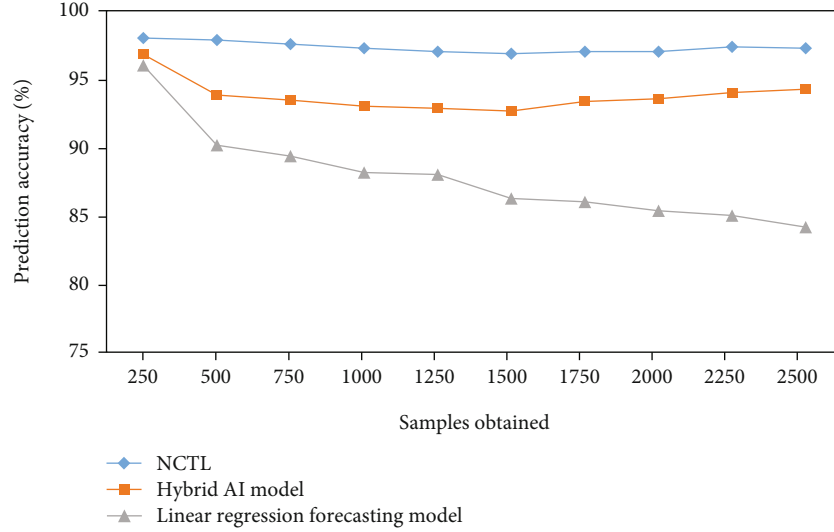


FIGURE 6: Graphical representation of prediction accuracy.

specified state itself (i.e., Tamil Nadu), though relevant and optimal filtered features are utilized for prediction, the error is said to be present. However, with simulations conducted for 250 different samples collected from both government laboratories and private laboratories, out of 30 predicted values, the measured value was observed to be 32, 33, and 34 using the NCTL method, [1, 2], respectively. From these predicted values, it is inferred that the error is significantly lower using the NCTL method when compared to [1, 2]. This is because the application of the nonlinear regressive feature selection algorithm uses the proposed NCTL method. By applying this algorithm, key features are obtained for prediction. This algorithm considers both the new cases reported and also the previous day's infection reports, followed by active cases are computed. A reverse suppression model is used to choose the important features. The sum of squares is measured for every feature depending on the residual function. Finally, every iteration approximation is carried out to select relevant features. With this, relevant features required for prediction were evolved. This, in turn, minimized the prediction error using the NCTL method by 42% compared to [1] and 19% compared to [2].

**4.2. Case 2: Prediction Accuracy.** An apparent manner to evaluate the COVID-19 prediction quality of the learned model is to see how long-term the predictions given by the method are accurate. In our work, the COVID-19 prediction accuracy refers to the patients with both symptoms and asymptomatic (i.e., infected with corona) are being detected as the same. This is mathematically expressed as given below.

$$\text{PAcc} = \sum_{i=1}^n \frac{\text{SCP}}{S_i} * 100. \quad (19)$$

From the above equation (19), the prediction accuracy 'PAcc' is evaluated based on the samples involved for testing 'S<sub>i</sub>' and the swab samples being correctly predicted 'SCP.' It is measured in terms of percentage (%). Summaries of the

TABLE 3: Performance measurements for prediction time.

Samples obtained	Prediction time (ms)		
	NCTL	Hybrid AI model	Linear regression forecasting model
250	36.26	41.25	45
500	41.35	54.15	73.25
750	55.25	70.25	90.55
1000	73.55	85.35	113.25
1250	82.55	100.45	125.55
1500	102.35	115.35	140.35
1750	115.25	125.35	175.35
2000	130.35	145.55	200.25
2250	145.55	160.13	215.35
2500	160	180.35	225.35

performance measurement for prediction accuracy for the abovementioned three methods, NCTL, hybrid AI model [1], and linear regression forecasting model [2], are found in Table 2.

The inferences from the above table precisely reveal that the proposed NCTL method comparatively performs better than the hybrid AI model [1] and linear regression forecasting model [2] in terms of prediction accuracy. Figure 6 depicts a graphical representation of prediction accuracy.

In Figure 6, we talk about the prediction accuracy of the proposed method NCTL and compare it with two other methods [1, 2]. The final cosine neighborhood-based LSTM-based classification depends on the death rate, case registered, and recovery rate. First, the NCTL method selected for COVID-19 in the India dataset is trained to predict these three values. We have calculated the time series cosine function on the validation data to evaluate the methods. We have compared using baseline methods such as [1, 2]. From comparison analysis, it is inferred that the prediction accuracy of COVID-19 is comparatively better

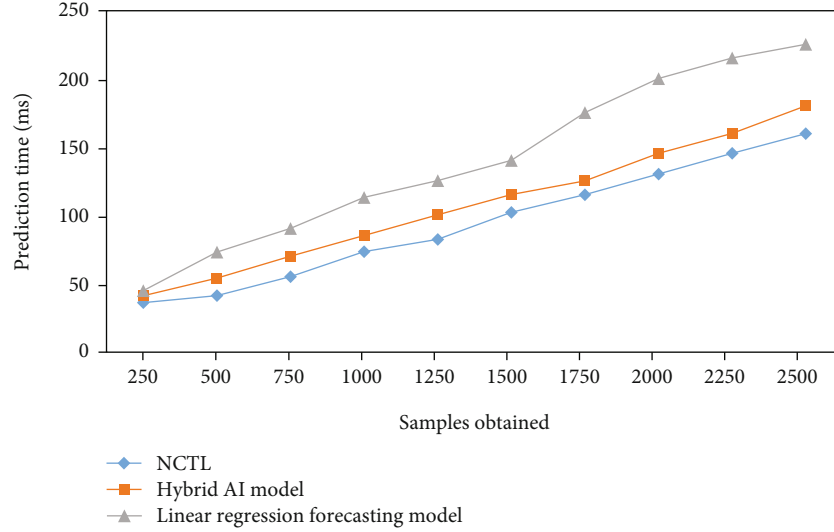


FIGURE 7: Graphical representation of prediction time.

when applied using the NCTL method, while simulations are conducted with 250 samples. It was found that the swab samples being correctly predicted using the NCTL method was found to be 245, whereas in the case of [1, 2] it was found to be 242 and 240. From these results, it is inferred that the prediction accuracy using the NCTL method is found to be better than [1, 2]. This was due to the application of the cosine neighborhood-based LSTM algorithm in the proposed NCTL method. First, optimal filtered features were given as input in the LSTM based on three different factors, namely, death rate, case registered, and recovery rate, and prediction analysis. Second, time series analysis was made considering conditional probability functions, in addition to filtered features also considering preceding and succeeding outcomes. With this, the prediction accuracy using the NCTL method was found to be better by 4% compared to [1] and 11% compared to [2], respectively.

**4.3. Case 3: Prediction Time.** Finally, prediction time refers to the time consumed in predicting any disease. In our work, the time taken for predicting the coronavirus pandemic is mathematically expressed as given below.

$$PT = \sum_{i=1}^n S_i * \text{Time [Prediction]}. \quad (20)$$

From the above equation (20), the prediction time ‘PT’ is measured based on the samples obtained ‘ $S_i$ ’ and the time consumed for predicting COVID-19 ‘Time [Prediction]’ in India. It is measured in terms of milliseconds (ms). Summaries of the performance measurement for prediction time for three methods, NCTL, hybrid AI model [1], and linear regression forecasting model [2], are listed in Table 3.

Finally, Figure 7 above shows the time involved in predicting COVID-19 in India based on the previous cases registered. With 2500 samples considered for samples from previous cases registered, the figure shows the time incurred in predicting COVID-19. With minimum time perceived in

predicting, higher amounts of cases being registered in several districts in Tamil Nadu or the death rate can be reduced to a greater extent with maximum recovery rate also. From the figure, it is inferred that the prediction time using the NCTL method is comparatively lesser than comparing [1, 2], though time is found to be in the increasing trend with the increasing swab test samples considered. However, the NCTL method was found to be comparatively better than [1, 2]. This is because of the application of the cosine-based neighborhood filter algorithm in the proposed NCTL method that assists in obtaining the optimal filtered features via neighborhood-based filter function, matrix factorization, and conditional probability functions. First, the neighborhood-based filter function is applied to compute the similarity between two neighboring users. With the aid of matrix factorization, matrix representation of cases and individual details are obtained extensively. Next, conditional probability functions are used for considering both the next outcome and the previous outcome results. The drudgery involved in acquiring parameters for filtering is said to be reduced. Also, this method achieves the optimum filter. With this, the prediction time involved in COVID-19 using the NCTL method is reduced by 14% when compared to [1] and 22% compared to [2], respectively.

## 5. Conclusion

In this paper, we have proposed a deep learning method for predicting the number of COVID-19 in Indian states. We have combined the number of cases and individual details to predict different parameters for a risk classification task. Based on the number of cases registered with a recovery rate and death rate, they are categorized into high risk, medium risk, and low risk so that lockdown measures can be made according to the results and to have the minimum impact on economic issues. First, nonlinear regressive feature selection is used as a relevant feature selection model. Next, based on the neighborhood function, the previous day’s outcomes were used as a measure for predicting the succeeding day’s

outcomes. Finally, time series were also included with cosine neighborhood-based LSTM for timely and accurate prediction in India. The proposed NCTL method is designed for accurate COVID-19 prediction with minimal prediction time. In future work, the proposed method will be further implemented for accurate COVID-19 prediction with higher prediction accuracy and minimum time consumption by using various feature selection models. The results of these predictions will assist civic bodies and healthcare workers in managing services and organizing medical groundwork appropriately. In the future, the effects of temperature, humidity, and the environment on the COVID-19 spread in cities and countries were considered.

### Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare no conflicts of interest.

### References

- [1] N. Zheng, S. Du, J. Wang et al., "Predicting COVID-19 in China using hybrid AI model," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 2891–2904, 2020.
- [2] R. Sujath, J. M. Chatterjee, and A. E. Hassanien, "A machine learning forecasting model for COVID-19 pandemic in India," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 7, pp. 959–972, 2020.
- [3] S. Vaid, R. Kalantar, and M. Bhandari, "Deep learning COVID-19 detection bias: accuracy through artificial intelligence," *International Orthopaedics*, vol. 44, no. 8, pp. 1539–1542, 2020.
- [4] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," *Internet of Things*, vol. 11, p. 100222, 2020.
- [5] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *International Journal of Information Technology*, vol. 12, no. 3, pp. 731–739, 2020.
- [6] G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, and A. E. Hassanien, "Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art," *SN Computer Science*, vol. 1, no. 4, pp. 1–15, 2020.
- [7] J. C. Sipior, "Considerations for development and use of AI in response to COVID-19," *International Journal of Information Management*, vol. 55, p. 102170, 2020.
- [8] S. Çakan, "Dynamic analysis of a mathematical model with health care capacity for COVID-19 pandemic," *Chaos, Solitons & Fractals*, vol. 139, p. 110033, 2020.
- [9] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review," *Chaos, Solitons & Fractals*, vol. 139, p. 110059, 2020.
- [10] S. Ghosal, S. Sengupta, M. Majumder, and B. Sinha, "Linear regression analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases - March 14th 2020)," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 4, pp. 311–315, 2020.
- [11] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial intelligence (AI) applications for COVID-19 pandemic," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 4, pp. 337–339, 2020.
- [12] M. R. H. Mondal, S. Bharati, and P. Podder, "Diagnosis of COVID-19 using machine learning and deep learning: a review," *Current Medical Imaging*, vol. 17, no. 12, pp. 1403–1418, 2021.
- [13] Y. Mohamadou, A. Halidou, and P. T. Kapen, "A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19," *Applied Intelligence*, vol. 50, no. 11, pp. 3913–3925, 2020.
- [14] L. J. Muhammad, M. Islam, S. S. Usman, and S. I. Ayon, "Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery," *SN Computer Science*, vol. 1, no. 4, pp. 1–7, 2020.
- [15] W. Naudé, "Artificial intelligence vs COVID-19: limitations, constraints and pitfalls," *AI & Society*, vol. 35, no. 3, pp. 761–765, 2020.
- [16] M. T. Rashid and D. Wang, "CovidSens: a vision on reliable social sensing for COVID-19," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 1–25, 2021.
- [17] K. Roosa, Y. Lee, R. Luo et al., "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020," *Infectious Disease Modelling*, vol. 5, pp. 256–263, 2020.
- [18] L. Li, Z. Yang, Z. Dang et al., "Propagation analysis and prediction of the COVID-19," *Infectious Disease Modelling*, vol. 5, pp. 282–292, 2020.
- [19] F. Rustam, A. A. Reshi, A. Mehmood et al., "COVID-19 future forecasting using supervised machine learning models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020.
- [20] D. P. Misra, V. Agarwal, A. Y. Gasparian, and O. Zimba, "Rheumatologists' perspective on coronavirus disease 19 (COVID-19) and potential therapeutic targets," *Clinical Rheumatology*, vol. 39, no. 7, pp. 2055–2062, 2020.
- [21] Z. Liaoa, P. Lana, X. Fanb, B. Kellyc, A. Innesc, and Z. Liao, "SIRVD-DL: a COVID-19 deep learning prediction model based on time-dependent SIRVD," *Computers in Biology and Medicine*, vol. 138, p. 104868, 2021.
- [22] A. Muneer, S. M. Fati, N. Arifin Akbar et al., "iVaccine-Deep: prediction of COVID-19 mRNA vaccine degradation using deep learning," *Journal of King Saud University-Computer and Information Science*, 2021.
- [23] H. Kalhori, M. M. Alamdari, and L. Ye, "Automated algorithm for impact force identification using cosine similarity searching," *Measurement*, vol. 122, pp. 648–657, 2018.
- [24] S. Shastri, K. Singh, S. Kumar, P. Kour, and V. Mansotra, "Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study," *Chaos, Solitons & Fractals*, vol. 140, p. 110227, 2020.