Review Article

A Review on Deep Sequential Models for Forecasting Time Series Data

Dozdar Mahdi Ahmed,[1] **Masoud Muhammed Hassan** [iD],[1,2] **and Ramadhan J. Mstafa** [iD][1,3]

[1]*Department of Computer Science, University of Zakho, Duhok 42001, Kurdistan Region, Iraq*
[2]*College of Art and Science, American University of Kurdistan, Duhok, Kurdistan Region, Iraq*
[3]*Department of Computer Science, College of Science, Nawroz University, Duhok 42001, Kurdistan Region, Iraq*

Correspondence should be addressed to Masoud Muhammed Hassan; masoud.hassan@uoz.edu.krd

Deep sequential (DS) models are extensively employed for forecasting time series data since the dawn of the deep learning era, and they provide forecasts for the values required in subsequent time steps. DS models, unlike other traditional statistical models for forecasting time series data, can learn hidden patterns in temporal sequences and have the memorizing data from prior time points. Given the widespread usage of deep sequential models in several domains, a comprehensive study describing their applications is necessary. This work presents a comprehensive review of contemporary deep learning time series models, their performance in diverse domains, and an investigation of the models that were employed in various applications. Three deep sequential models, namely, artificial neural network (ANN), long short-term memory (LSTM), and temporal-conventional neural network (TCNN) along with their applications for forecasting time series data, are elaborated. We showed a comprehensive comparison between such models in terms of application fields, model structure and activation functions, optimizers, and implementation, with a goal of learning more about the optimal model used. Furthermore, the challenges and perspectives of future development of deep sequential models are presented and discussed. We conclude that the LSTM model is widely employed, particularly in the form of a hybrid model, in which the most accurate predictions are made when the shape of hybrids is used as the model.

1. Introduction

There are many different forms of data, and among them is time series data. This data type has the ability to predict future data at the same rate as the forecasting technique of analysis [1]. Time series frequently exhibit time dependencies, resulting in two comparable time points to be classified differently or to predict divergent behavior [2]. Time periods are represented by years, months, weeks, days, or hours [3]. Many real-world applications, such as biological sciences [4], healthcare [5], and financial and weather prediction [6], use time series data to capture data over time. Traditional approaches, such as exponential smoothing [7, 8], autoregressive (AR) [7, 9], or structural time series models [10], are focused on parametric models driven by

domain knowledge. Machine learning techniques have recently made it possible to learn chronological dynamics purely from data [11]. Machine learning has emerged as a vital component of the next generation of time series forecasting algorithms, as data availability and computer power have increased in recent years [12]. A variety of machine learning algorithms and models have been employed to forecast time series data. Some of them employ hybrid approaches for prediction, which mix multiple models or integrate any optimization process in the prediction process [3].

Deep learning (DL) models, a subset of machine learning that use deep neural networks, are employed in a range of research domains, including speech recognition [13], text mining, and image analysis [14]. DL models combine

numerous layers to represent the data abstraction to develop computational methods. Although the DL training model takes a long time to train due to the large number of parameters, it tests faster than other machine learning techniques [15]. This is because DL models can learn hidden patterns from data. Deep neural networks can learn complicated data representations [16] by including unique architectural assumptions [17], which take into account the small difference of the underlying datasets. This eliminates the requirement for model creations and human feature engineering. Additionally, the amount of available frameworks for back-propagation made the network training much easier by allowing for modification of network elements and loss functions [18, 19]. We emphasize diverse parts of sequential understanding and how they compare to one another in regard to deep learning. The goal of this paper is to give a comprehensive analysis of the three different deep sequential forecasting models that have been utilized to produce predictions. Our analysis may aid researchers in identifying the best model to use based on the application domain, as well as in the development of unique or hybrid algorithms.

With the widespread use of deep sequential models in several domains of science and application, one may be confused as to which model is appropriate for solving their problems. Our research is driven by a goal to provide a comprehensive review of current deep learning time series models. It is worth providing the user communities with the best model for their problem in terms of the model performance in various fields. This study intends to answer which models are typically used for which applications, which programs best implement the algorithms, how data splitting is utilized in the studies, and which optimization approaches are employed by the researchers. Our evaluation may aid researchers and practitioners in making the best model selection based on scope resemblance and system necessities, consequently assisting in the development of novel algorithms. The main contributions of this paper are summarized as follows:

(i) To the best of our knowledge, this is the first review work that covers the usage of deep sequential models for forecasting time series data to serve the research community.

(ii) We reviewed the three common deep sequential models.

(iii) A comprehensive review of deep sequential models in different applications, such as healthcare, finance and stocks predictions, weather forecasting, environment, and pollution, is provided.

(iv) Main challenges in deep sequential models are highlighted.

(v) Some research gaps and open issues in the field that need further investigation are discussed.

The reminder of the paper is organized as follows. Section 2 gives a quick overview of three popular models for interpreting deep sequential models with time series data. In Section 3, we reviewed selected articles in the last three years that used deep sequential models in various applications. Section 4 gives an in-depth discussion about the models, their implementation, activation functions used, and a comparison of applications based on the models used. Finally, in Section 5, we draw the matter to a close by making a last comment on the methodologies that have been examined.

2. Deep Sequential Model

Deep sequential models are deep learning techniques used when both the input and the output are sequence data [20]. Sequences are made up of data points that can be arranged so that observations at one point in the sequence provide meaningful information about observations at other places in the sequence. The sequence data require managing incalculable supervised learning tasks, and the sequence learning issue may occur when the input is a sequence and the output is a single data point, as in video activity detection, classification of sentiments, and stock price forecasting. On the other hand, the output might be a series of data points and the input is a single data point, as in picture captioning, music production, and speech synthesis. There are also the cases when both the output and the input are sequences, as in voice recognition, natural language understanding, and DNA sequence analysis; in this case, the lengths of the input and output sequences might be the same or different [21].

Deep learning neural networks have proven to be capable of simulating extremely complex input-output mappings, particularly in disciplines like computer vision, voice recognition, and natural language processing. This concept has led to the development of several deep learning techniques in the field of time series forecasting that outperform traditional approaches in terms of accuracy performance [22]. In this section, we will briefly describe the three most commonly used deep sequential models in the literature: ANN, LSTM, and TCNN.

2.1. Artificial Neural Network (ANN). ANN is a technique for processing data that mimics the human brain. A brain acquires knowledge from human experiment, and an ANN processes data similarly to the human brain. It is categorized as a supervised or unsupervised learning based on the amount of information about the values of the output variables [23]. The basis of an ANN is made up of nodes or neurons that are placed in parallel linked processing unit arrays. A multilayer perceptron (MLP) is an ANN model that feed-forwards datasets into a set of appropriate data outputs [24]. The ANNs are composed of three fundamental layers: input, hidden, and output. An input layer is the information source for networks. The hidden layer learns the nonlinear relationships amidst the input(s) and the output(s) by altering the weights. Layers are composed of varying numbers of neurons, which process input through activation functions. The output layer contains forecasted data [25], and the output of unit i can be given as follows:
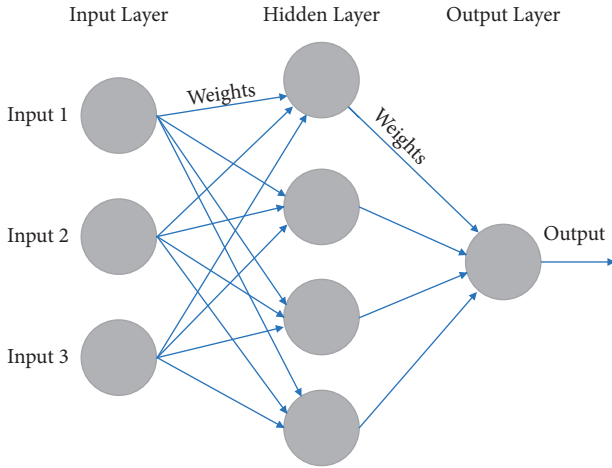
FIGURE 1: ANN architecture [27].

$$y_i = \sum_{i=1}^{p} w_i x_j + \theta_i, \qquad (1)$$

where $w_{ij} = (w_1, w_1, \ldots\ldots, w_p)$ is the parameter vector of unit i, p is the number of neurons in the unit i preceding layer, x_i is the input of unit i, and θ_i is the bias of unit i. This weighted sum y_i, referred to as the unit i incoming signal, is then transferred via a transfer function as shown in equation (1) [26]. The ANN's general architecture is described in Figure 1.

The output of a function is known as the activation function. Weights associated with the connections between two neurons are proportional to their strength. Typically, a neuron's computation is separated into two stages: aggregation and activation. Utilizing the aggregation function requires computing the total of all the inputs received by the neurons through all their incoming connections. The result values are sent to the activation function. There are several types of activation functions; the sigmoid and hyperbolic tangent are frequently used. The rectified linear unit is another function which has grown in popularity [2]. The mathematical formulas of these activation functions are as follows: equations (2)–(5).

(i) ReLU is the often used activation function, owing to its ease and effectiveness. A subset of neurons fires simultaneously, sparing the network and increasing efficiency.

$$R(x) = \text{Max}(0; x). \qquad (2)$$

(ii) Sigmoid is another fairly common method of activation, which constrains the output to a value between 0 and 1 [28],

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \qquad (3)$$

(iii) Tanh is the same as the sigmoid function, but tanh function allows for the output values within a specified range (−1, 1),

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \qquad (4)$$

(iv) Softmax: this function produces continuous values between (0, 1) and is often applied as a classifier at the output layer since it produces probabilities spread throughout the number of classes in multi-class predictions,

$$f(x) = \frac{e^x}{\sum_{i=1}^{n} e^x}. \qquad (5)$$

2.2. RNN. Deep learning algorithms are subsets of machine learning algorithms that aim to identify various representations for incoming data. One of the most widely used models for deep sequential learning is the recurrent neural network (RNN), which was developed in the 1980s [29]. RNN is used to handle sequence data [15]. RNNs contain memory cells, which enables them to remember data from the past, and that is important for forecasting future outcomes. They have been extensively utilized to solve supervised learning problems involving sequential data. Due to the sequential structure of time series, comparable designs have been utilized for predicting time series as well [22]. The back-propagation through time (BPTT) algorithm introduced by Werbos [30] is capable of training RNN. The drawback with RNN training is that the RNN design is backward dependent over time [31]. We will cover some RNN designs that have been proven to be effective at predicting time series data, such as LSTM model and gated recurrent unit (GRU) [32]. The GRU structure is similar to the LSTM structure; however, it is easier to compute and implement and has less parameters.

2.2.1. LSTM. LSTM algorithm is a type of RNN algorithms that was developed by Hochreiter and Schmidhuber [33] to model the structure of sequential data. The disadvantage of the normal RNNs is that their learning output reduces when data are moved away from the input. When the initial weight is no longer accessible, the vanishing gradient descent problem occurs. The LSTM is recommended for managing each memory cell for both state and output values throughout the learning process [31]. LSTM networks are specialized in learning and analyzing sequential data such as data classification [34], processing [35], and time series data forecasting with time differences of unknown sizes [36]. LSTM is a chain-like architecture suitable for memorizing the information and long term using four network layers as represented in Figure 2. The LSTM network is made up of memory blocks that look like cells. This cell state is crucial because it allows data to go ahead and stays unaltered. However, data can be added or wiped using sigmoid activation function gates. These gates consist of a sequence of matrix operations with dissimilar weights. With gates, it is possible to avoid long-term dependency challenges that exist with memorization of the LSTM. In the following paragraph, we will go through how to use LSTM [37].
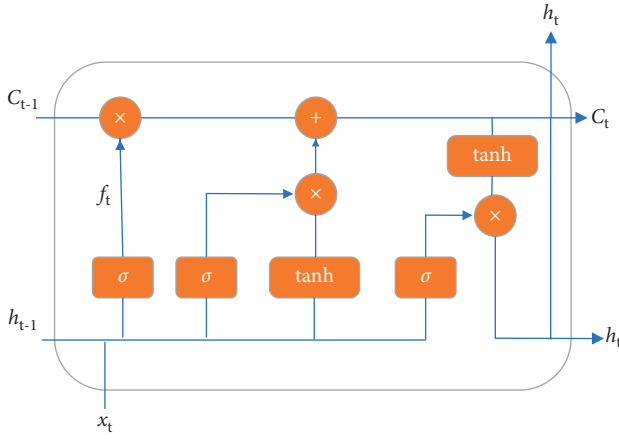
FIGURE 2: LSTM diagram [38].

In the LSTM procedure, the first step is to specify data that is unnecessary for the operation. This is accomplished by using the sigmoid function, which takes the output as (v_{t-1}) at time $(t-1)$ and the existing input as (x_t) at time (t). The sigmoid function assesses whether the portion of the old output should be deleted based on the old output. This process is referred to as the forget gate (f_t) as in equation (6). The vector values vary from 0 to 1, with one value for each digit in the cell (c_{t-1}).

$$f_t = \sigma\big(w_f[v_{t-1}, x_t] + b_f\big). \qquad (6)$$

The sigmoidal function is denoted by σ as in equation (3), the weights are denoted by w_f, and the bias of the forget gate is denoted by b_f.

The following equations include two cases: one that ignores, and one that stores the current state of input cell in x_t. Two layers comprise the process: a sigmoidal layer and a tanh layer. Using 0 or 1, the sigmoidal layer determines whether the latest information is updated or disregarded. The weights are updated and calculated via the tanh function and then transfer data between (1 to −1) in the second layer. Values are assigned weight based on their importance. As can be seen in (C_t), both values have been changed, and a new cell state is created.

$$
\begin{aligned}
m_t &= \sigma\big(w_f[v_{t-1}, x_t] + b_m\big), \\
N_t &= \tanh\big(w_f[v_{t-1}, x_t] + b_n\big), \\
C_t &= C_{i-1}f_t + N_t m_t.
\end{aligned}
\qquad (7)
$$

Finally, the output v_t is multiplied by the newly layer formed via $\tanh(C_t)$. It is based on the output of sigmoidal gates (Q_t).

$$
\begin{aligned}
Q_t &= \sigma\big(w_q[v_{t-1}, x_t] + b_q\big), \\
v_t &= Q * \tanh(C_n),
\end{aligned}
\qquad (8)
$$

where w_q and b_q denote the parameter within a neuron and biases of the output gates, respectively.

2.2.2. TCNN. TCNNs are generic convolutional networks proposed by [39]. The TCN network is generally utilized to perform the sequence modeling tasks with causal constraint. A sequence modeling network learns to predict the output sequence $(y_0, \ldots, y_t)$ given the input sequence $(x_0, \ldots, x_t)$, and the matching output sequence $(y_0, \ldots, y_t)$ by network training on some loss function between both the estimated and output sequences $(\widehat{y}_0, \ldots, \widehat{y}_t)$ may be obtained. The causal constraint imposed on the network means that the prediction $\widehat{y}_t$ is conditional on the inputs $(x_0, \ldots, x_t)$ but not on the future inputs $(x_{t+1}, \ldots, x_t)$. As can be seen in Figure 3, the TCNN is a hierarchical structure composed of many convolutional hidden layers of equal size to the input layer [39].

3. Application

To demonstrate the importance of the time series prediction problems, a state-of-the-art study is conducted by categorizing deep learning research works by different application domains, such as healthcare, finance, energy, traffic, and weather prediction, based on the most widely used network model designs (ANN, LSTM, TCNN). The following paragraphs provide a summary of the articles for each application domain, highlighting the aims achieved for each method and field.

3.1. Healthcare. Deep sequential models are frequently used in healthcare. Deep learning's predicting capabilities and automated feature detection make it a compelling tool for disease diagnosis. There are numerous studies in the field of healthcare applications.

Nikparvar et al. [40] suggested a multivariate and multi-time series long short-term memory (MTS-LSTM) network for forecasting the COVID-19 pandemic in terms of affirmed cases, mortality, and movement concurrently. The results indicated that including mobility as a variable and training the network with many samples improve predictive performance in terms of prediction bias and variance. Additionally, the study demonstrated that the projected outcomes are comparable in terms of accuracy and spatial patterns to those of a typical ensemble model employed as a benchmark [40].

TEG-net was presented by Hong et al. [41] as a revolutionary deep learning approach for physiologic diagnosis and explained time series with high precision. TEG-net constructs T-net (a multi-scale bidirectional TCNN) for directly modeling physiological time series, E-net (personalized linear model) for directly modeling expert features extracted from physiological time series, and G-net (gating neural network) for combining T-net and E-net for diagnosis. Through G-net, the combination of T-net and E-net enhances precision in diagnosing, and E-net may be used for the sake of clarification. TEG-net exceeds the second-best benchmark in terms of area under the receiver operating characteristic curve and area under the precision-recall curve by 13.68% and 11.49%, respectively [41].

In another work, the LSTM was used by Da Silva et al. [42] to anticipate the patient's vital signs and then to estimate the severity of the patient's health status using
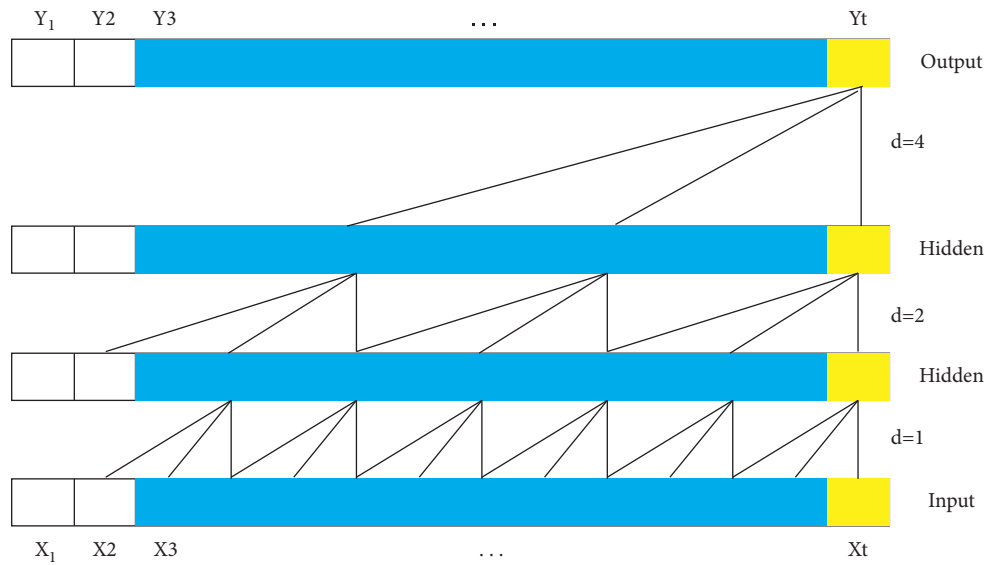
FIGURE 3: TCNN architecture [39].

prognostic indexes, which are extensively employed in medicine. According to experiments, it is possible to estimate vital signs with a high degree of accuracy (>80%) and so forecast prognostic indexes in advance, enabling patients to get treatment before they deteriorate [42].

To anticipate COVID-19 cases in India and Chennai, multivariate LSTM models were examined by Devaraj et al. [43]. Both short- and long-term infected cases were predicted, and the study showed that the stacked LSTM and LSTM models beat other models in terms of accuracy. When compared to other algorithms, the stacked LSTM models predicted more reliably and generated results with an error of less than 2% [43].

In the work of Kafieh et al. [44], different models include multi-layered perceptrons, random forests, and various versions of LSTM. Their proposed model was trained on three datasets, including COVID-19 cases, to predict the outbreak in nine different countries (Germany, Iran, Japan, Italy, Switzerland, Korea, China, Spain, and the USA), and their performances were evaluated using four metrics, including MAPE, RMSE, NRMSE, and R^2. According to the authors, promising results were discovered for predicting the pandemic's future trajectory by using a modified version of LSTM termed M-LSTM [44].

A unique deep learning architecture based on the LSTM was also presented by Balaji et al. [45] for the purpose of grading the severity of Parkinson's disease (PD) based on gait pattern. Three independent gait datasets were used to train the LSTM network. Each dataset included vertical ground reaction force (VGRF) records of distinct types of walking. The experimental findings demonstrated that Adam-optimized LSTM networks were capable of successfully learning gait kinematic data and provided a rate of accuracy for binary and multi-class classification, as well as an accuracy increase over similar approaches [45].

Shastri et al. [46] suggested three types of LSTM, including stacked LSTMs, bidirectional LSTMs, and convolutional LSTMs, for forecasting COVID-19 cases one month in advance in India and the USA. As a result, convolutional LSTM outperformed the other two models and forecasted the mortality rate of COVID-19 with high accuracy and very little error for all four datasets in both countries. Based on the mean absolute percentage error (MAPE), a convolutional LSTM model obtained an error rate of 2.0% to 3.3% [46].

Kara [47] presented a hybrid technique for predicting multi-step influenza outbreaks based on incorporated LSTM neural networks and genetic algorithms (GAs). They used weekly data on influenza-like illnesses (ILI), which obtained in the USA. The experimental findings indicated that the provided hybrid model outperforms other well-developed machine learning techniques, a statistical model, and a fully connected neural network when various performance measures were considered during peak times [47].

In the work of Shetty and Pai [48], they showed the capacity of the multilayer perceptron (MLP), an ANN model, to anticipate the number of infected cases in the Indian state of Karnataka. The forecasting model's parameters were chosen using the partial autocorrelation function, and their performance was compared to parameters chosen using the cuckoo search (CS) technique. The use of the CS algorithm resulted in improved prediction performance when MAPE was used. Additionally, to validate the model's effectiveness, data from Hungary's COVID-19 cases were employed, which produced a MAPE of 1.55%, confirming the stability of the developed ANN model for predicting COVID-19 instances in the Karnataka state [48].

Hawas [49] utilized the GRU version of the RNN structure to predict daily COVID-19 infections in Brazil using forecasting models trained on sparse raw data (30 time-steps and 40 time-steps alternatives). Since the GRU has a smaller number of parameters and has a simpler structure, it converged faster and provided better performance than LSTM, which has a lot more parameters and is

TABLE 1: A summary of published articles that used deep sequential models in healthcare predictions.

Ref.	Method	Application	Results obtained	Metrics used	Data unit
[40]	MTS-LSTM	Healthcare COVID-19 prediction	Excellent spatial granularity superior predicted performance many weeks ahead.	RMSE	Daily
[41]	TEG-net	Healthcare (heart rate)	TEG-net exceeds the second-best baseline in terms of area under the receiver operating characteristic curve and area under the precision-recall curve.	ROC-AUC, PR-AUC, F1-score, precision, recall, accuracy	Every minute
[42]	LSTM	Healthcare	It is possible to estimate vital signs with a high degree of accuracy.	Accuracy MAE >80%	Hourly
[43]	Stacked LSTM, LSTM	Healthcare (forecasting pandemic)	The stacked LSTM models predicted more reliably and generated good results.	MAPE	Daily
[44]	M-LSTM	Healthcare (forecasting pandemic)	The greatest result was discovered for predicting the pandemic's future trajectory	MAPE, RMSE, NRMSE, and R^2 are 0.51, 458.12, 0.0016, and 0.9	Daily
[45]	LSTM	Healthcare (Parkinson's disease diagnosis)	Adam-optimized LSTM networks are capable of successfully learning gait kinematic data and provide an average accuracy	Accuracy improvement of 3.4%	Every 2 minutes
[46]	Stacked, bidirectional, and convolutional LSTM	Healthcare COVID-19 prediction	Convolutional LSTM outperformed the other two models	MAPE, accuracy, precision, recall, F-measure	Daily
[47]	LSTM-GA	Influenza outbreak forecasting	The experimental results show that the hybrid model presented here outperforms other well-developed machine learning approaches	RMSE	Weekly
[48]	ANN	Healthcare COVID-19 prediction	The developed ANN model for predicting COVID-19 cases was found to be stable	MAPE and RMSE	Daily
[49]	GRU	COVID-19 prediction	GRU can achieve better convergence and high performance with the simplest structure	R^2	Daily

more complicated to implement [49]. Table 1 summarizes the abovementioned studies that applied deep sequential models in healthcare predictions.

3.2. Energy.

Deep learning's recent success in a variety of domain applications has drawn researchers to the energy field, particularly in areas like demand forecasting, renewable energy, and building energy efficiency, and the breadth of methods presented, as well as the growing number of publications, demonstrates its enormous potential.

Two suggested methodologies for forecasting the power consumption of office buildings were validated by Chen et al. [50] using actual data from the University of Glasgow as a case study. To minimize the impacts of occupants' activities, the first suggested technique divided power consumption data into set time periods including working and nonworking hours. The second proposed approach combined ANNs and fuzzy logic approaches to match the building's base load, peak demand, and occupancy rate to many meteorological factors. The simulation findings indicated that, when compared to the conventional ANN technique, both suggested approaches had lower RMSE for forecasting power consumption [50].

In work of Lin et al. [51], for solar energy forecasting, the performance of TCNN was compared to multilayer feed-forward neural networks and recurrent networks, including the state-of-the-art LSTM and GRU recurrent networks. The

assessment was based on two Australian datasets that comprise historical heliacal and climate data, as well as future-day weather prediction data. The results indicated that TCNN beats other models in terms of precision and was capable of maintaining a lengthier history of effectiveness than recurrent networks [51].

Lin et al. [52] also proposed a new deep learning technique for probabilistic forecasting that combined attention processes with temporal convolutional attention network (TCAN) and shows its effectiveness in a case study of solar energy forecasting. TCAN extracts temporal dependence using the hierarchical convolutional structure of TCNN and then employs sparse attention to concentrate on critical time steps. Authors claimed that the TCAN beat various state-of-the-art deep learning prediction models, like TCNN, in terms of accuracy. They also conclude that the TCAN needs less convolutional layers for an enlarged receptive field, and it is quicker to train [52].

A method for estimating extremely short-term solar production relying on the LSTM with the temporal attention mechanism (TA-LSTM) was also suggested by Pan et al. [53]. The length of the time series was first determined using partial autocorrelation, which was then used as an input to the LSTM forecasting model. The trials were carried out to ensure that the proposed approach worked well. The predictive solar production of each forecasting technique from 1 May 2016 to 10 May 2016 showed that the suggested strategy was possible and successful [53].

TABLE 2: A summary of published articles that used deep sequential models in energy predictions.

Ref.	Method	Application	Results obtained	Metrics used	Data units
[50]	ANN + fuzzy logic	Energy (electricity demand)	The method was lowering the average.	RMSE = 42%	Hourly
[51]	TCNN, LSTM,	Energy (solar power)	In terms of accuracy and capability to maintain a longer effective history, TCNN beats other models.	MAE	Half-hourly
[52]	TCNN, TCAN	Energy (solar power)	In terms of accuracy, TCAN outperforms some state-of-the-art deep learning prediction models, such as TCNN.	MAE	Half-hourly
[53]	LSTM + TA	Solar generation (energy)	Employing partial autocorrelation to calculate the input lag, the TA method improves performance over regular LSTM.	RMSE = 0.25 MAE = 0.12	Daily
[54]	ETS + LSTM	Energy (electricity demand)	Superior performance and competitiveness with both classical models	MAPE 5%	Monthly
[55]	LSTM	Energy (electricity consumption)	Prediction results were obtained with the least degree of error.	RMSE	Every minute
[56]	LSTM	Energy	LSTM outperformed ARIMA and SARIMA	RMSLE, RMSE, MASE, and MAPE	Daily
[57]	NARX-ANN-PSO	Energy	The model is capable of determining the appropriate input and output lag terms.	MSE, RMSE, MAPE	Hourly
[58]	Combined ANN	Energy (electricity price)	The model's performance may be utilized as a baseline for making EV charging decision	MAPE	Hourly
[59]	LSTM + fuzzy	Energy (electricity demand and wind speed)	The nonstationary and irregular characteristics are effectively addressed by the CFML model. Under the MAPE metric, the model perfection wind speed and electrical power load by an average of 49% and 70%, respectively.	MAPE	-
[60]	ISO-TS-RBF-RFNN	Power load forecasting	The model performs the best in terms of long-term load forecasting accuracy.	MAPE	Monthly, daily, weakly, hourly

Dudek et al. of [54] developed an approach for deep learning that was hybrid and hierarchical for predicting midterm electricity request for 35 European countries. It was then combined with exponential time smoothing (ETS) and LSTM. Their experiments on monthly electricity demand time series demonstrated the proposed model's superior performance and competitiveness with both classical models such as ARIMA and ETS and the latest machine learning models [54].

On investigating the multi-step time series dataset, several models of LSTM were analyzed by Ghanbari and Borna [55]. The research examined if a model could be developed to forecast how much power a residence would use over the following seven days. The dataset they used contains four years' worth of data on residential power use. The intended prediction results were obtained with the minimum errors possible when contrasted with current state-of-the-art models [55].

Kumar Dubey et al. [56] used the LSTM model to contribute to the establishment of accurate predictions and forecasting for the energy unit supply. The dataset in their experiments was associated with energy consumption measurements from 5,567 London households that participated in the UK Power Network-led Low Carbon London initiative during November 2011 to February 2014. The findings showed that energy usage had a significant positive correlation with humidity and a strong negative correlation with temperature, and the LSTM outperformed ARIMA and

SARIMA models with an average mean absolute error (MAE) of 0.23.

Another work by Mustapa et al. [57] on energy consumption was estimated by using multiple linear regression (MLR) and nonlinear autoregressive with exogenous input (NARX-ANN). The NARX-ANN architecture was optimized using the particle swarm optimization (PSO) technique, and the anticipated values were compared to the actual values. It was shown that the NARX-ANN-PSO model had a smaller error [57].

Dang et al. [58] proposed a technique for forecasting next-day electricity prices to the 5-minute level using a model integrated with eight components of ANN. The combined ANN model was then used to forecast electricity or time-of-use (TOU) prices for the following day with a 5-minute accuracy. The authors concluded that the model's performance had a MAPE accuracy of roughly 13%, which might be used as a benchmark for EV charging decision-making [58]. Table 2 summarizes the abovementioned studies that applied deep sequential models in energy predictions.

Wei et al. [59] presented CFML (complementary ensemble empirical mode decomposition (CEEMD)-fuzzy time series (FTS)-multi-objective grey wolf optimizer (MOGWO)-long short-term memory (LSTM)) as a novel hybrid prediction system. Four wind speed datasets and two electrical power load datasets were used for energy forecasting; the forecasting models CEEMD, FTS, and MOGWO

demonstrated the ability to carry the strength of each component while electively improving the prediction performance of the CFML predicting model on the basis of stabilization and precision [59].

Another work by Wen et al. [60] proposed a novel approach for predicting power load. In this approach, stages of weather station were selected and predicted using the Takagi–Sugeno (TS) fuzzy model, which is an upgraded self-organizing radial basis function recurrent neural network using the TS fuzzy model (ISO-TS-RBF-RFNN). The suggested ISOTS-RBF-RFNN model developed a technique for determining the current firing strength of fuzzy modes by using a new type of activation mechanism and robust-type fuzzy rules to enhance the prediction model's resilience. In comparison with the other five models (MLR, SVR, FIR, GA-LSTM, and classic SO-TS-RBF-RFNN), the suggested model outperforms in comparison with the other five models when confronted with a variety of uncertainty about feature components [60].

3.3. Environment and Pollution.

Global climate change and its associated consequences have created a great deal of issues for farmers and land management. Not only does climate change have a noteworthy influence on the resources, environment, and socioeconomic situations of all areas, but it also influences water resources, air quality, agriculture, rural development, and health. Therefore, its mitigation requires precise forecasting.

By combining the unique cooperation search algorithm (CSA) into the ANN learning process, Feng and Niu [61] presented a hybrid ANN model for river flow forecasting. The suggested model was evaluated using data of two real-world hydrological sites in China. During both the training and testing stages, the hybrid technique based on ANN and CSA consistently exceeded the effectiveness of control systems and improved predicting outcomes [61].

To forecast hourly $PM_{2.5}$ (particles that have diameter less than 2.5 micrometers) concentrations, a unique hybrid forecasting model was constructed on the basis of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and deep temporal convolutional neural network (deep-TCNN) was developed by Jiang et al. [62]. Using the $PM_{2.5}$ concentrations in Beijing as a sample, experimental results revealed that the proposed CEEMDAN-deep-TCNN model had the highest forecasting accuracy when compared to classical time series models, ANN, and the prevalent DL model [62].

With the purpose of improving forecasts for smart agriculture production, the authors of [63] employed LSTM model and compared it to back-propagation method. Experimental results based on the smart agriculture dataset, which was sophisticated by the Electronic Systems Lab of Gadjah Mada University, showed that the LSTM outperformed back-propagation in terms of prediction accuracy [63].

In another work by Belavadi et al. [64], the LSTM and RNN models were presented as an approach that is scalable for monitoring and collecting real-time data on air pollution concentrations from diverse locations and for forecasting future air pollutant concentrations. The data were obtained through a wireless sensor network that collects and transmits pollutant concentrations to a server, with sensor nodes distributed around Bengaluru, South India, and from the Government of India's real-time air quality data as part of its Open Data project. They found that there is no such thing as a "one-size-fits-all" strategy that will work in every environment or circumstance [64].

Hamami and Dahlan [65] developed a construct LSTM model for forecasting air pollution levels. The model was able to forecast five measures of air pollution, including PM10, SO_2, CO, O_3, and NO_2. The findings indicated that the LSTM model was efficient with a root mean square error that equals 5.58 [65].

In order to improve the accuracy of water quality forecasting, an LSTM-BP combination model technique was presented by Jia and Zhou [66]. This algorithm used both LSTM and BP neural networks. A framework for time series prediction was created using the water temperature data from the No.6 large-scale integrated observation signal in the Yangtze estuary as an example. The proposed approach was compared to the LSTM and the BP models, and it was determined that the time series predicted by LSTM-BP were more accurate [66].

Four ANN models were calibrated by Sharma et al. [67] to evaluate the effect of different variables on river flow prediction. When model 3 (without river flows) was compared to model 4 (with river flows), the results were greatly improved (with including lag 1 river flow as one of the inputs). In terms of the performance metrics, it was concluded that the developed ANN model 4 was more accurate at modeling river flow, except for extremely high peaks, resulting in solving highly complex nonlinear river flow prediction problems [67].

The hybrid models provided by Zhao et al. [68] to forecast monthly streamflow data first improved complete ensemble empirical mode decomposition with adaptive noise (ICEEWT) and then used the GRU to learn the relationship between historical and future streamflow series. The model was then used in conjunction with the improved grey wolf optimizer (IGWO) to identify the approximate parameter combination while balancing underfitting and overfitting. For two stations, the proposed ICEEWT-IGWO-GRU model outperformed the single GRU model average, decreasing MAE and RMSE by 50% and 52%, respectively, confirming its efficiency with a more reliable prediction [68]. Table 3 summarizes the abovementioned studies that applied deep sequential models in environment and pollution predictions.

3.4. Finance and Stocks.

Stock market time series forecasting is a concept used in quantitative finance. Finance prediction is a traditional but difficult topic that has attracted the interest of both economists and computer scientists. Researchers are using machine deep sequential models for predicting stock market. For example, three machine learning methods were used by Zekić-Sušac et al. [69] to

TABLE 3: A summary of published articles that used deep sequential models in environment and pollution predictions.

Ref.	Method	Application	Results obtained	Metrics used	Data unit
[61]	Hybrid ANN + CSA	Environment pollution	The model was consistently outperforming control models.	Error = 24.1%	Monthly
[62]	Hybrid CEEMDAN-deep-TCNN	Environment pollution	The model shown to perform the best compared to time series models, ANN, and popular deep learning models.	MAPE, RMSE, MAE	Hourly
[63]	LSTM	Environment pollution	LSTM outperformed back-propagation in terms of prediction accuracy.	RMSE = 0.08	Air pressure parameters
[64]	LSTM	Environment pollution	There is no "one-size-fits-all" approach that is effective in any city under any circumstance.	RMS = 30–40 ppm	Hourly
[65]	LSTM	Environment pollution	The LSTM model's root mean square error is 5.58	Average error was 5.58 by RMSE	Monthly
[66]	LSTM-BP	Environment pollution	The time series predicted by LSTM-BP was found to be more accurate	RMSE = 1.03. MAE = 0.68 MAPE = 3.37	Daily
[67]	ANN	River flow	The model significantly increases when the lag 1 river flow is included as an input.	R^2, MIA, RMSE, MNSE	Daily and monthly
[68]	Hybrid ICEEWT-IGWO-GRU	Streamflow	The proposed model outperformed the single GRU	MAE, RMSE	Monthly

forecast the energy costs of public buildings, and an actual dataset of general buildings of the Croatian public sector's localized database was developed. The prediction models were developed in two stages: the first stage involved a comparison of the use of ANNs and two types of regression trees, CART and random forest (RF); the second stage involved the combination of RF-Boruta feature and machine learning approach for forecasting. The ANN and regression tree topology were optimized using cross-validation approach. The findings indicated that when Boruta and machine learning are combined, the highest precise forecasting in terms of symmetric mean absolute percentage error (sMAPE) is attained [69].

Ahmed et al. [70] suggested a strategy which incorporates the forex loss function (FLF) into an LSTM model named FLFLSTM, which reduces the discrepancy among the actual and forecast averages of forex candles. Data were used from 10,078 four-hour candles for the EURUSD pair, and they discovered that when compared to the standard LSTM model, the suggested FLF-LSTM system showed a 10.96% percent reduction in total mean absolute percentage error (MAPE). Additionally, they stated that the predicting error for high and low prices was lowered. The research demonstrated that by introducing domain expertise into the methods of training machine learning models, substantial improvements in forex price prediction were obtained [70].

In another work conducted by Dave et al. [71], Indonesia's future export forecasts with LSTM were applied to the nonlinear component of the data. Autoregressive integrated moving average (ARIMA) model was also applied to the linear part, and the two most widely used time series prediction algorithms (ARIMA and LSTM) were combined to generate the best model. In comparison with the separate models, the hybrid (LSTM-ARIMA) model produced the lowest mean absolute percentage error (MAPE) and root mean square error (RMSE) values [71].

In another work conducted by Hryhorkiv et al. [72] to forecast time series of stock indices, they suggested a

sophisticated hybrid forecasting model based on a combination of ARIMA and ANN, which combines the benefits of both approaches. Mean square error (MSE) calculations demonstrated that the suggested algorithm produced more accurate projections of the 500 stock indexes [72].

Livieris et al. [73] suggested a CNN-LSTM model for accurate gold price and movement prediction. The model made use of convolutional layers' ability to extract relevant information and learns the internal representation of time series data, as well as the LSTM layers' efficacy in recognizing short- and long-term relationships. The initial experimental study demonstrated that combining the LSTM with extra convolutional layers might significantly improve predicting ability [73].

Farahani and Hajiagha [74] proposed a framework for forecasting stock price indices using an artificial neural network (ANN) and training it using various novel metaheuristic algorithms including social spider optimization (SSO) and the bat algorithm (BA). They selected features using genetic algorithm (GA) like a heuristic technique. The standard artificial neural network was compared to a hybrid metaheuristic-based ANN. In comparison with earlier approaches, SSO and BA had least errors, indicating that they could better forecast stock prices. However, when error is used as a metric of superior, the social spider algorithm outperformed the others. The ANNs outperformed ARIMA-based time series models in predicting stock prices. Their experimental results demonstrated that the hybrid models achieved a high degree of accurately explaining the model [74].

A new intuitionistic fuzzy time series (IFTS) forecasting method based on LSTM is proposed by Kocak et al. [75]. The model named as DIFTS-LSTM is applied to the Giresun temperature data and the Nikkei 225 stock exchange time series forecasting. Experimental results of the proposed model demonstrated better forecasting performance than some alternative methods, namely, fundamental fuzzy time series models and the classical models of time series [75].

TABLE 4: A summary of published articles that used deep sequential models in finance and stock predictions.

Ref.	Method	Application	Results obtained	Metrics used	Data units
[69]	ANN	Finance (energy cost building)	Used models showed an increase in accuracy.	SMAPE = 22.35%	Monthly
[70]	FLFLSTM, LSTM	Finance (forex price)	Considerable improvements in forex price forecasting are obtained.		Four hours
[71]	Hybrid (ARIMA + LSTM)	Finance	The model has the lowest error metric compared to all models examined.	MAPE = 7.38% RMSE = 1.66×1013	Monthly
[72]	ARIMA_ANN	Finance	MSE calculations on the hybrid model demonstrated high accuracy.	MSE = 1.58	Yearly
[73]	CNN-LSTM	Gold price	Combining LSTM with extra convolutional layers has significantly improved predicting accuracy	MAE, RMSE, ACC, AUC, SEN, SPE	Daily
[74]	ANN-metaheuristic	Stock price	Experiments reveal that hybrid models correctly describe the model to a high degree.	MSE	Close-open stock price
[75]	DIFTS-LSTM	Stock exchange	The model has better forecasting performance. IFTS forecasting models consider both membership and nonmembership values in solving a time series	RMSE and MAE	Daily

Table 4 summarizes the abovementioned studies that applied deep sequential models in finance and stock predictions.

3.5. Weather Forecasting. Meteorological forecasting is the process of predicting the condition of the atmosphere at a particular area using a variety of weather characteristics. Weather predictions are created by collecting data on the status of the atmosphere at any given time.

Farmers often rely on further precise short- to medium-range predicted from regional forecasting models with greater accuracy. For this purpose, Hewage et al. [76] investigated current state-of-the-art models (TCN and LSTM), with the goal of designing and testing a lightweight and innovative weather forecasting systems. According to the findings, the suggested model that uses TCN delivered superior prediction than the LSTM and other conventional machine learning techniques [76].

In another work, an ANN model was created by Yadav and Malik [77] to forecast wind speed at 10, 20, 30 minutes, and 1 hour ahead of time, in the mountainous region of India. They proposed a graph to depict wind speed data that have been averaged over a 10-minute period. The study concluded that their experiment can be used to track wind power over the Internet [77].

The application of ANN, a computational intelligence approach, was proposed by Rahul et al. [78] as a significant step forward in the creation of an intuitive framework capable of comprehending and predicting nonlinear weather phenomena. The suggested research focused on creating a user-friendly framework that can accurately forecast the weather with the least amount of mistake and a more appropriate design [78].

In another study by Bou-Rabee et al. [79], a hybrid ANN technique was suggested for wind power capacity and electricity generation at coastal sites, based on wind-related dependent variable's properties. Data from three Kuwaiti coastal sites were used to confirm the suggested approach. For the computation of wind speed, the hybrid model, whose integrated ANN and particle swarm optimization (PSO), forecasted wind direction one month ahead of time. Before finding the best ANN architecture, the NN started by assessing its output using a variable neuron in the hidden layer. With RMSE and MSE, the hybrid ANN-PSO framework outperformed marginal systems based on ANN [79].

In another study on forecasting Indian summer monsoon rainfall (ISMR), Johny et al. [80] examined an adaptive hybrid modeling framework called the adaptive ensemble empirical mode decomposition-artificial neural network (AEEMD-ANN) for forecasting ISMR. The AEEMD-ANN method performed reasonably well in capturing hydrologic extremes when compared to the EEMD-ANN forecasting model, with a high degree of precision when trying to capture dry seasons [80].

Unnikrishnan and Jothiprakash [81] developed an integrated SSA-ARIMA-ANN model for daily rainfall forecasting. Singular spectrum analysis (SSA) was used to preprocess the data before combining it with ARIMA and ANN models. The statistical performance of the proposed model revealed that the hybrid SSA-ARIMA-ANN model was capable of forecasting daily rainfall in the catchment with a high degree of confidence [81].

Niu et al. [82] introduced a new sequence-to-sequence model based on the attention-based gated recurrent unit (AGRU) that significantly enhances the quality of wind power forecasting (WPF) systems. AGRU model was assessed using National Renewable Energy Laboratory (NREL) data. The model successfully handles the error accumulation issue associated with a recursive approach. The authors concluded that the GRU model can be used instead of the LSTM in multi-step-ahead WPF problems because the GRU can predict the same results as the LSTM while taking less time to compute [82]. Table 5 summarizes the abovementioned studies that applied deep sequential models in weather predictions.

3.6. Traffic Flow. Traffic flow forecasting is a critical activity that predicts the amount of time required to flow the traffic in advance. Numerous intriguing studies using deep learning have been conducted in this field. For traffic flow

TABLE 5: A summary of published articles that used deep sequential models in weather predictions.

Ref.	Method	Application	Results obtained	Metrics used	Data unit
[76]	MINO-TCNN, MISO-TCNN	Weather forecasting	TCNN produced better forecasting. It can be applied as an effective method localized climate forecasting tool. It is executed on a stand-alone personal computer.	MSE	Every 15 minutes
[77]	ANN	Weather	The error in guessing 10 minutes in advance is the smallest statistically.	MSE, MAE, RMSE, and ME	10, 20, 30 minutes and 1 hour
[78]	ANN	Weather	System is capable of forecasting weather with a low error rate and a more acceptable structure	MSE ranges from 0.9325 to 3.5321	Hourly and daily
[79]	Hybrid ANN + PSO	Weather (wind speed)	The model was quite accurate.	MAPE (3–6%)	Daily
[80]	AEEMD-ANN	Rain fall	The model was shown to be effective in capturing very low SWM rainfall.	R, MAE, NRMSE, IA	Monthly
[81]	SSA-ARIMA-ANN	Rain fall	The hybrid model was capable of forecasting the catchment with a high degree of confidence.	R^2, RMSE, MAE, MPE, MNE	Daily
[82]	AGRU	Wind power forecasting (WPF)	The AGRU model proposed provides competitive capabilities in power system WPF. The attention method increases computing time every learning epoch, and hyperparameter tuning takes time	NRMSE and MAPE	Every 5 minutes

forecasting at isolated sites, Lu et al. [83] developed a novel LSTM network supplemented with temporal-aware convolutional context (TCC) blocks and a novel loss-switch mechanism (LSM). Extensive trials on two widely used benchmark traffic flow datasets from California's metropolitan regions demonstrated that the suggested approaches are effective and reliable when compared to state-of-the-art traffic flow forecasting methods [83].

To improve the accuracy of short-term traffic flow forecasting, a model based on traffic flow time series analysis was presented by Ma et al. [84], with an upgraded LSTM. To begin, time series analysis was done on traffic flow data to produce a reliable time series. Then, an enhanced LSTM model was created using LSTM and bidirectional LSTM networks. The bidirectional long-term memory network (BILSTM) was incorporated into the prediction model by combining the benefits of sequential data and the long-term reliance of forwarding and reversing LSTM. Finally, the suggested method's performance was assessed by comparing anticipated results to real traffic data. With respect to precision and stability, the suggested strategy outperformed both examined methods [84].

Wang et al. [85] analyzed the road section's traffic flow data and weather circumstances in details, and produced a model for short-term traffic flow prediction based on the attention mechanism and the 1DCNN-LSTM network. The proposed model combined the temporal expansion capabilities of the CNN with the long-term memory benefits of the LSTM. The experimental findings indicated that the 1DCNN-LSTM-attention model's prediction effect was superior to that obtained without incorporating the weather element. The suggested model's prediction effect indicated a quicker convergence rate and greater forecast accuracy [85].

Bohan and Yun [86] proposed a model to predict traffic flow data. A database was collected using GPS data from K5, 32, and 73 buses received from the Hohhot Bus Corporation. Their data were collected from 6 : 00 to 20 : 00 on weekdays between May and October 2017. They chose a bidirectional

recurrent neural network (BRNN) model inside a recurrent neural network (RNN) and compared it to the LSTM model and a gated recurrent unit (GRU) model. Their results showed that BRNN outperforms the other two models in terms of prediction accuracy and model performance [86].

George and Santra [87] proposed a new hybrid model based on agglomerated hierarchical K-means (AHK) clustering and fuzzy optimum long short-term memory (FOLSTM) called AHK-FOLSTM. In the LSTM model, the whale optimization algorithm (WOA) was used to better manage the fuzzy rule parameters and calculated the best weight value. Experimental results demonstrated that their proposed method outperforms other state-of-the-art approaches on different evaluation metrics used [87]. Table 6 summarizes the abovementioned studies that applied deep sequential models in traffic flow predictions.

3.7. Other Fields. Deep sequential models are applied in different other fields for forecasting time series data. For instance, a new hybrid method based on biased random key genetic algorithms was suggested by Cicek and Ozturk [88] to identify the optimal network design and parameter combinations. The BRKGA-NN method estimated the number of hidden neurons, their bias values, and the weights of ties among nodes. The proposed BRKGA-NN model was assessed on some of the most well-known time series datasets. The performance of the BRKGA-NN was compared against genetic algorithm-based ANNs, ANNs with back-propagation, support vector regression, and autoregressive integrated moving average. Forecasting findings indicated that the BRKGA-NN algorithm delivered more accurate predictions than the other approaches [88].

A deep model structure was created by Hong et al. [31] to improve the prediction accuracy of the turbofan engine's remaining useful life (RUL). The proposed model improved performance by sequentially stacking one-dimensional convolutional neural network (1D-CNN), LSTM, and

TABLE 6: A summary of published articles that used deep sequential models in traffic flow predictions.

Ref.	Method	Application	Results obtained	Metrics used	Data unit
[83]	LSTM (TCC + LSM)	Traffic flow	Showed the efficacy and robustness of the proposed methods	MSE	Every 5 and 10 minutes
[84]	Bidirectional LSTM, LSTM, and LSTM_BILSTM hybrid	Traffic (traffic flow prediction)	The suggested strategy outperformed both examined methods for accuracy and stability	RMSE = 16.7, MSE = 12.6, R^2 = 0.86, MAPE = 6.3	Every 5 minutes
[85]	1DCNN-LSTM	Traffic flow	The prediction impact of the proposed model demonstrated a faster convergence rate and increased forecast accuracy.	MAE, MSE, RMSE	Daily
[86]	BRNN, LSTM, GRU	Traffic flow	The BRNN model forecasted velocity at three, six, and twelve time steps, BRNN the most advanced option, and provides with the lowest MAE and RMSE	MAE, RMSE	Every 20 minutes
[87]	AHK-FOLSTM	Traffic speed prediction	FOLSTM determines the exact traffic speed from anomalous traffic data in order to overcome nonlinear features.	MSE, MAPE, MAE, and RMSE	5 minutes

bidirectional LSTM algorithms. Additionally, it integrated a residual network and a dropout strategy to enhance the proposed model's learning ability and minimizes model complexity via correlation analysis. When compared to former studies, the outcomes indicated that the model performed better at predicting RUL [31]. In another study by Fu et al. [89], to predict RUL, they created a deep residual LSTM featuring domain-invariance (DIDR-LSTM) to optimize RUL estimation's prognostic efficacy among domains. Experimental results showed that the DIDR-LSTM achieved a superior cross-domain RUL prediction accuracy [89].

Chen et al. [90] provided a probabilistic forecasting framework for multiple linked time series forecasting using TCNN model. The proposed approach was applied to estimate both parametric and nonparametric estimations of probability density. In estimating both point and probabilistic forecasting, the framework outperformed existing state-of-the-art approaches [90].

In another study by Pandey and Wang [91], a CNN model with an encoder-decoder architecture was suggested with an extra temporal convolutional module (TCM) placed among the encoder and decoder for real-time domain voice enhancement, termed TCNN. The suggested model was trained independently of speaker and noise. The suggested model consistently outperformed the state-of-the-art real-time convolutional recurrent model in terms of enhanced outcomes [91].

Another technique was developed by Golshani and Ashtiani [92] to improve forecasting and decision-making accuracy in comparison with the existing literature using TCNN and RNN for cloud service workload prediction and usage prediction. The suggested approach's decisions set up a degree of compromise between the choice criteria. The order of preference by similarity to ideal solution technique was used, which is one of the most well-known multi-criteria decision-making procedures. It was offered as a technique for deciding on scalability [92].

In the work of Smyl [93], the M4 forecasting competition's winning entry model was proposed. The proposed model employed a dynamic computational graph neural network architecture that provides a hybrid and hierarchical forecasting technique by combining traditional exponential smoothing with an LSTM model [93].

In the work of Li et al. [94], a unique multi-objective speech enhancement algorithm called a stacked and temporal convolutional neural network (STCNN) was developed. The STCNN architecture outperformed other neural network models in terms of feature extraction and sequential modeling. The proposed STCNN also outperformed LSTM, TCNN, and CRNN in all noise types [94].

Fan et al. [95] devised another novel hybrid model for efficient and precise well production forecast, which is critical for prolonging the life of a well and optimizing reservoir recovery. This approach combined ARIMA and LSTM models. Three well real production time series were evaluated to compare the efficiency of the suggested models ARIMA-LSTM-DP (daily production time series) and ARIMA-LSTM with the LSTM, ARIMA, and LSTM-DP models. In comparison with other models, the hybrid ARIMA-LSTM-DEEP model outperformed and its results were more reliable.

He et al. [96] used a SARIMA-CNN-LSTM model to predict daily visitor demand, which is critical for the tourism industry's functioning in 6 countries, including China, Hong Kong, Singapore, Korea, Philippine, and Taiwan. The SARIMA-CNN-LSTM model beats other models in terms of prediction accuracy and was capable of extracting more information from high-frequency data throughout the forecasting process.

A novel fuzzy time series approach based on an ARMA-type recurrent pi-sigma artificial neural network was presented by Kocak et al. [97], and particle swarm optimization (PSO) was used to execute the optimization. The experimental results of using proposed model (FTS-ARMATPS-ANN) with a recurrent structure based on both ARMA-type and PSO algorithms revealed improvement in predicting performance for numerous real-life time series [97].

To develop a novel forecasting model, Egrioglu et al. [98] developed new fuzzy time series and intuitionistic fuzzy time series definitions. The data were fuzzified with an

TABLE 7: A summary of published articles that used deep sequential models in different applications.

Ref.	Method	Application	Results obtained	Metrics used	Data unit
[88]	BRKGA-NN	Computer science	Generated more accurate predictions. A lot of parameters. Bad running times.	MAE = 6.47 RMSE = 6.51 MAPE = 0.01	Various public datasets
[31]	Bi-LSTM, LSTM	Forecasting the turbofan engine's remaining service life	The model performs better at predicting RUL	RMSE, MSE, MAE, MAPE	Sensor data
[89]	DIDR-LSTM	RUL prediction	DIDR-LSTM allows more precise cross-domain RUL prediction	RMSE	Sensor data
[90]	Deep-TCNN	Electricity, traffic, and spare parts of cars demand	The framework outperformed existing state-of-the-art approaches.	NRMSE, SMAPE, MASE	Hourly and monthly
[91]	TCNN	Speech enhancement	Significantly outperforms existing real-time systems in the frequency domain.	STOI and PESQ scores	Hourly
[92]	TCNN	Computer science	It was suggested that a technique for deciding on scalability be developed.	MSE	-
[93]	Hybrid ES-LSTM	M4 competition	The ES retains key elements like seasonality. The prediction of forex price LSTM networks allows for nonlinear trends and cross-learning from various related series.	MAPE = 0.45–0.49	Monthly and hourly
[94]	STCNN	Speech enhancement	STCNN architecture outperformed comparable neural network models	-	Hourly
[95]	ARIMA-LSTM, ARIMA-LSTM-DP	Well production	Hybrid ARIMA-LSTM-DEEP model is more reliable.	RMSE, MAE, MAPE, Sim	Daily
[96]	SARIMA-CNN_LSTM	Tourism demand	SARIMA-CNN-LSTM model beats other individual models.	RMSE, MAPE	Daily
[97]	Pi-sigma ANN	Computer science	The FTS-ARMA TPS-ANN, which is based on both ARMA and PSO algorithms, considerably improves prediction performance.	MAPE, RMSE	Various time series data
[98]	Pi-sigma ANNs + fuzzy *c*-means	Temperature and stock exchange forecasting	The best forecasts were generated by our suggested technique for all test data of each dataset	RMSE, MAE	Daily

intuitionistic fuzzy *c*-means approach, and the fuzzy relations were created with pi-sigma ANNs, which were trained with an artificial bee colony approach. The temperature dataset and stock exchange were used to predict future data in the suggested technique. Among all the competitor methods, the proposed method produced the best forecasts for all test sets of investigated data [98]. Table 7 summarizes the abovementioned studies that applied deep sequential models in different applications.

4. Discussion

In this section, we will go over the reviewed articles, discuss paper's implementation settings, compare each model's application, and focus on the significant differences.

4.1. Models. We looked at 60 papers from the last three years, in which most of them are from 2021, as shown in Figure 4. The models we reviewed in this study were LSTM, ANN, and TCNN, with LSTM (26 articles) being the most popular, followed by ANNs (15 articles) and TCNN (9 articles), as illustrated in Figure 5. The LSTM network offers considerable benefits for processing long-term sequences, which is why it is used in the majority of the examined publications. This is owing to its ability to independently
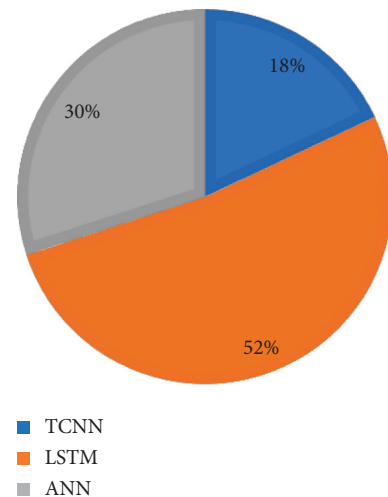

FIGURE 4: Published papers by year.

correlate information on the time axis, as well as its high memory learning capacity [99].

Using the TCNN model, on the other hand, is uncommon because data storage during the assessment of the TCNNs models must store the raw sequence up to the duration of the effective history, which may necessitate
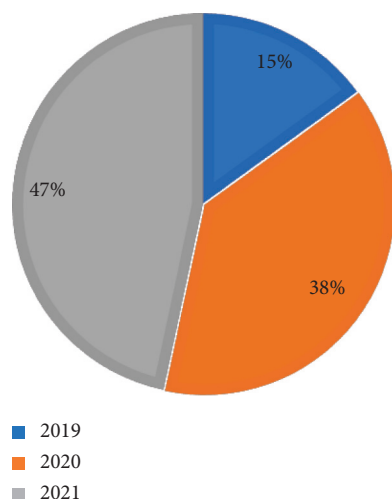
Figure 5: Number of articles based on model.

additional memory during the assessment. This includes the possibility of parameter modification in the event of a domain transfer, and different domains may have different requirements for the amount of history required for prediction. Thus, TCNN may underperform when transferring a model from a domain where just a limited amount of memory is required (i.e., small filter sizes (k) and dilation factors (d)) to a domain where considerably more memory is required (i.e., much bigger k and d) [39].

It is also worth noticing that ANN models are not as often used as LSTM models for sequential data prediction. This is due to the ANNs disadvantages, which depends on how weight values are initialized, local minima, and overfitting problems, as well as the difficulty of generalizability [100].

4.2. Implementation.

Deep sequential (DS) models are difficult to implement and require a high level of technical understanding as well as a significant amount of time. Several firms have focused their efforts on developing frameworks for implementing, training, and using DS models in order to make them more accessible. The basic purpose of DS frameworks is to provide an interface that allows models to be implemented regardless of their mathematical complexity [101]. The most often utilized frameworks in the literature are listed in Table 8. It refers to ANN, TCNN, and LSTM models, demonstrating that Python and the Keras package are the most popular programming languages for building deep learning models. ANN was primarily implemented in MATLAB, whereas TCNN was implemented in PyTorch. Only one paper used the R programming language.

4.3. Application Comparison.

An ultramodern study was undertaken to demonstrate the problem's importance, categorizing deep learning research works by application domain (such as finance and stocks, energy, healthcare, weather, environment, traffic flow, and others), and the most often utilized network architectures are LSTM, TCNN, and ANN. The LSTM was frequently used for sequential data in most fields, as shown in Figure 6. For weather prediction, the ANN model was preferred compared to other models, and the same result was obtained by another survey prepared by Jaseena and Kovoor [102]. According to the publications we evaluated, the ANN models were not used to predict traffic or healthcare, and the TCNN models were not used for weather forecasting. We could not come up with any points that proved the differences in other fields.

4.4. Optimization Methods.

Deep learning models require optimization method as a fundamental component. We believe that optimization for neural networks is an important topic for theoretical study because, despite its nonconvexity, it is tractable and can considerably improve our understanding of tractable problems. Furthermore, traditional optimization theory is insufficient to account for a wide range of events [103]. Table 9 shows several optimization approaches based on deep sequential models that have been examined, with Adam and stochastic gradient descent being the most commonly utilized optimizers (SGD). We may conclude from this review that a variety of optimization algorithms are used, including Adam, SGD, global optimizer, cooperation search algorithm (CSA), PSO, cross-entropy, BRKGA, and NADAM. The Adam optimizer was employed in practically all deep sequential models, most notably in LSTM, then TCNN, and only infrequently in ANN. In LSTM model optimization, the SGD optimizer was utilized, and an ANN was optimized using several types of optimizers.

The Adam optimizer has a number of advantages over other optimizers that make it more useful. The magnitudes of parameter updates are invariant to gradient rescaling, its step sizes are roughly bound by the step size hyperparameter, it does not require a stationary objective, it works with limited gradients, and it generally conducts a type of step size annealing. After that, SGD established itself as an effective and dependable optimization strategy that was crucial in the success of various machine learning initiatives [104]. We noticed that Adam proved to be dependable and well suited to a wide range of nonconvex optimization problems in deep learning models.

4.5. Activation Function.

The activation functions are the fundamental decision-making units of any neural network. They also assess the output of the network's neural nodes, making them critical for the network's overall performance. As a result, while computing neural networks, it is critical to select the most appropriate activation function [105].

The frequently utilized activation functions in the reviewed studies are ReLU, sigmoid, Tanh, scaled exponential linear unit (SeLU), and Softmax. ReLU is the most commonly used function, as shown in Table 10, especially in LSTM and TCNN models. This could be due to the fact that it is the most widely used function for hidden layers. Lee [106] found that ReLU6 (a subtype of ReLU) outperforms other activation functions [106]. Additionally, the sigmoid

TABLE 8: Implementation framework for the reviewed papers.

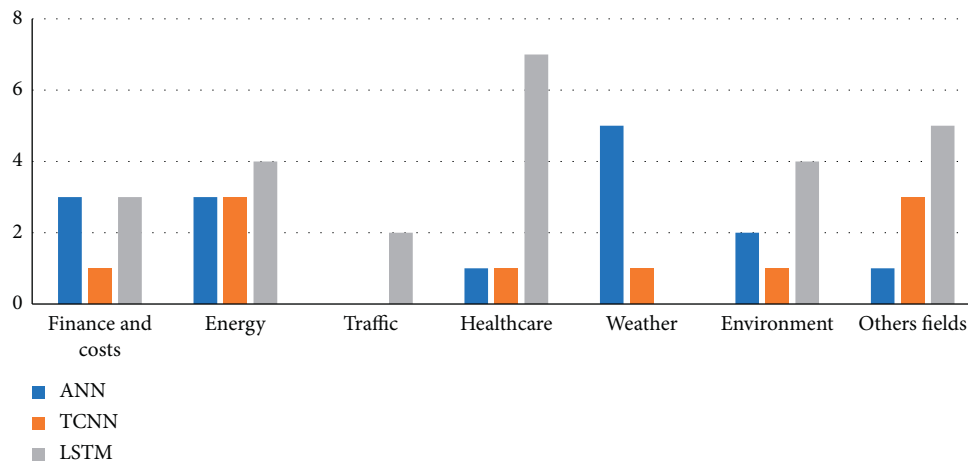| Models	Language	Framework	References
ANN	Python	—	[88]
	MATLAB	—	[48, 57, 58, 67, 74, 78–80, 97, 98]
TCNN	Python	PyTorch	[41, 51, 52, 90, 92]
	Python	Keras	[76]
LSTM	Python	Keras	[40, 42–47, 56, 70, 73, 84, 85]
	C++, DyNet	—	[93]
	MATLAB	—	[54, 75]
	Python	—	[53, 55, 64, 66, 71]
	Python	PyTorch	[89]
	R	—	[54]

FIGURE 6: Application according to models used.

TABLE 9: Optimization method used in each model in the reviewed articles.

Optimization	Number of articles that used model-based optimization		
	ANN	LSTM	TCNN
Adam	1	14	5
SGD	1	6	1
CSA	1	0	0
Global optimizer	1	0	0
BRKGA	1	0	0
PSO	3	0	0
NADAM	0	1	0
Cross-entropy loss function	0	0	1

TABLE 10: Activation function used in each model in the reviewed articles.

Activation Function	The number of articles that used model based on activation function		
	ANN	LSTM	TCNN
ReLU	1	9	6
Sigmoid	6	1	1
Tanh	0	8	1
Softmax	0	2	0
SeLU	0	1	0

was commonly employed in ANN models. This is due to the fact that it is not zero-centered, and it is known to be prone to the vanishing gradient function and zigzagging during training [105]. Because each activation function is unique, we cannot readily state which one is the best; the only option we can do is pick an activation function that is appropriate for our purposed network structure [107].

4.6. Model Evaluation. In this work, three prominent deep sequential models (ANN, LSTM, and TCNN) were investigated along with other hybrid models that outperformed standard models in forecasting accuracy. At least two deep learning algorithms are combined in hybrid models. These models are more resilient because they often increase forecast accuracy by enhancing the benefits of the particular methods applied [100]. The widely used architectures are the hybrid of LSTM models in energy applications, followed by the LSTM model, as found in the results of Alkhayat and Mehmood [108]. When compared to the LSTM model, the TCNN model produced highly accurate predictions when used to forecast solar power [51]. One of the initial preprocessing steps necessary in the context of deep learning is splitting the dataset into training and testing data. The evaluation of model performance is made possible by the creation of different training and testing samples. Instead of

segmenting data into training-testing set, best practices advocated segmenting data into training, validation, and testing sets. By splitting dataset, we can assess the model's generalization performance while also determining the model's hyperparameters. The validation set is used to test the model, whereas the training set is used to fit it [109]. Almost all of the evaluated articles in this review study have used a significant portion of the data for training. About ten studies, for example, used more than 80% of their data for training and the rest for testing. Validation data, on the other hand, are only employed in a few of the studies reviewed.

Evaluation metrics are used to assess and compare performance models [110, 111]. There are four sorts of evaluation measures for sequential models in general: classification metrics, regression metrics, profit analysis, and significance analysis [112]. Only one publication utilized a classification measure, while the others used regression measures, suggesting that these metrics are more efficient than others.

5. Conclusion

Deep sequential models are frequently employed in a variety of real-world applications for forecasting time series data. In this review paper, three deep sequential models for time series data forecasting were extensively investigated in 60 research projects over the last three years. This study provided an in-depth examination of the most effective and widely utilized deep sequential models for forecasting. We offered a comprehensive overview and comparison of current state-of-the-art deep sequential models. The LSTM model was the most widely utilized in a number of industries, including healthcare, financial, weather, energy, environment, and other disciplines. Hybrid models, on the other hand, were able to forecast extremely well. Furthermore, the use of time series forecasting methods like TCNN and LSTM in conjunction with optimization techniques like Adam worked quite well. Python and the Keras package were the most commonly used programming languages for the implementation of deep sequential models for real-world problems. The ReLU activation function was employed in the majority of the publications we looked at, but we could not say which one was the best. In addition, we explored a variety of topics, including common deep model applications, in-depth feature discussions, and substantial research gaps, before offering several possible future research pathways in this subject. We believe that this review article will benefit researchers from a variety of fields and will serve as a guideline for using time series data in deep learning-based applications.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] L. Qadrini, A. Asrirawan, N. Mahmudah, M. Fahmuddin, and I. F. Amri, "Forecasting bank Indonesia currency inflow and outflow using ARIMA, time series regression (TSR), ARIMAX, and NN approaches in lampung," *Journal Matematika, Statistika dan Komputasi*, vol. 17, no. 2, pp. 166–177, 2021.

[2] J. C. B. Gamboa, "Deep learning for time-series analysis," 2017, https://arxiv.org/abs/1701.01887.

[3] G. Mahalakshmi, S. Sridevi, and S. Rajaram, "A survey on forecasting of time series data," in *Proceedings of the 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)*, Kovilpatti, India, January 2016.

[4] D. S. Stoffer and H. Ombao, "Editorial: special issue on time series analysis in the biological sciences," *Journal of Time Series Analysis*, vol. 33, no. 5, pp. 701–703, 2012.

[5] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.

[6] M. Mudelsee, "Trend analysis of climate time series: a review of methods," *Earth-Science Reviews*, vol. 190, pp. 310–322, 2019.

[7] E. S. Gardner Jr., "Exponential smoothing: the state of the art," *Journal of Forecasting*, vol. 4, no. 1, pp. 1–28, 1985.

[8] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management Science*, vol. 6, no. 3, pp. 324–342, 1960.

[9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, Wiley, Hoboken, NJ, USA, 2015.

[10] R. Fildes, A. C. Harvey, M. West, and J. Harrison, "Forecasting, structural time series models and the kalman filter," *Journal of the Operational Research Society*, vol. 42, no. 11, p. 1031, 1991.

[11] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny, "An empirical comparison of machine learning models for time series forecasting," *Econometric Reviews*, vol. 29, pp. 594–621, 2010.

[12] B. Lim and S. Zohren, "Time-series forecasting with deep learning: a survey," *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 379, no. 2194, Article ID 20200209, 2021.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, MIN, USA, June 2019.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012, https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

[15] Y. Xin, L. Kong, Z. Liu et al., "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.

[16] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[17] J. Baxter, "A model of inductive bias learning," *Journal of Artificial Intelligence Research*, vol. 12, pp. 149–198, 2000.

[18] M. Abadi, A. Agarwal, P. Barham et al., "Tensorflow: large-scale machine learning on heterogeneous distributed systems," 2016, https://arxiv.org/abs/1603.04467.

[19] A. Paszke, S. Gross, F. Massa et al., "Pytorch: an imperative style, high-performance deep learning library," 2019, https://arxiv.org/abs/1912.01703.

[20] B. Shickel and P. Rashidi, "Sequential interpretability: methods, applications, and future direction for understanding deep learning models in the context of sequential data," 2020, https://arxiv.org/abs/2004.12524.

[21] S. Zargar, "Introduction to sequence learning models: RNN, LSTM, GRU," 2021, https://www.researchgate.net/publication/350950396_Introduction_to_Sequence_Learning_Models_RNN_LSTM_GRU.

[22] K. Agarwal, L. Dheekollu, G. Dhama, A. Arora, S. Asthana, and T. Bhowmik, "Deep learning based time series forecasting," in *Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA, December 2020.

[23] G. Toğa, B. Atalay, and M. D. Toksari, "COVID-19 prevalence forecasting using autoregressive integrated moving average (ARIMA) and artificial neural networks (ANN): case of Turkey," *Journal of Infection and Public Health*, vol. 14, no. 7, pp. 811–816, 2021.

[24] M. Hajihassani, R. Kalatehjari, A. Marto, H. Mohamad, and M. Khosrotash, "3D prediction of tunneling-induced ground movements based on a hybrid ANN and empirical methods," *Engineering with Computers*, vol. 36, no. 1, pp. 251–269, 2020.

[25] A. Mahmoud and A. Mohammed, "A survey on deep learning for time-series forecasting," in *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*, A. E. Hassanien and A. Darwish, Eds., vol. 77, pp. 365–392, Springer International Publishing, Cham, Switzerland, 2021.

[26] T. Rajaee and H. Jafari, "Two decades on the artificial intelligence models advancement for modeling river sediment concentration: state-of-the-art," *Journal of Hydrology*, vol. 588, Article ID 125011, 2020.

[27] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of long short-term memory (LSTM) neural network for flood forecasting," *Water*, vol. 11, no. 7, p. 1387, 2019.

[28] D. Oliva, E. H. Houssein, and S. Hinojosa, *Metaheuristics in Machine Learning: Theory and Applications*, Springer International Publishing, Cham, Switzerland, 2021.

[29] J. Schmidhuber, "Deep learning: our miraculous year 1990-1991," 2020, https://arxiv.org/abs/2005.05744.

[30] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[31] C. W. Hong, K. Lee, M.-S. Ko, J.-K. Kim, K. Oh, and K. Hur, "Multivariate time series forecasting for remaining useful life of turbofan engine using deep-stacked neural network and correlation analysis," in *Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Busan, Republic of Korea, February 2020.

[32] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: encoder-decoder approaches," 2014, https://arxiv.org/abs/1409.1259.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] A. A. Abdullah, M. M. Hassan, and Y. T. Mustafa, "A review on bayesian deep learning in healthcare: applications and challenges," *IEEE Access*, vol. 10, pp. 36538–36562, 2022.

[35] I. H. Sarker, "Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective," *SN Computer Science*, vol. 2, no. 3, p. 154, 2021.

[36] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, "A sequential deep learning application for recognising human activities in smart homes," *Neurocomputing*, vol. 396, pp. 501–513, 2020.

[37] W. Yang, R. Wang, and B. Wang, "Detection of anomaly stock price based on time series deep learning models," in *Proceedings of the 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID)*, Guangzhou, China, December 2020.

[38] T. Santosh, D. Ramesh, and D. Reddy, "LSTM based prediction of malaria abundances using big data," *Computers in Biology and Medicine*, vol. 124, Article ID 202059, 2020.

[39] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, https://arxiv.org/abs/1803.01271.

[40] B. Nikparvar, M. M. Rahman, F. Hatami, and J.-C. Thill, "Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network," *Scientific Reports*, vol. 11, no. 1, Article ID 21715, 2021.

[41] S. Hong, C. Wang, and Z. Fu, "Gated temporal convolutional neural network and expert features for diagnosing and explaining physiological time series: a case study on heart rates," *Computer Methods and Programs in Biomedicine*, vol. 200, Article ID 105847, 2021.

[42] D. B. da Silva, D. Schmidt, C. A. da Costa, R. da Rosa Righi, and B. Eskofier, "Deepsigns: a predictive model based on deep Learning for the early detection of patient health deterioration," *Expert Systems with Applications*, vol. 165, Article ID 113905, 2021.

[43] J. Devaraj, R. Madurai Elavarasan, R. Pugazhendhi et al., "Forecasting of COVID-19 cases using deep learning models: is it reliable and practically significant?" *Results in Physics*, vol. 21, Article ID 103817, 2021.

[44] R. Kafieh, R. Arian, N. Saeedizadeh et al., "COVID-19 in Iran: forecasting pandemic using deep learning," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6927985, 16 pages, 2021.

[45] E. Balaji, D. Brindha, V. K. Elumalai, and R. Vikrama, "Automatic and non-invasive parkinson's disease diagnosis and severity rating using LSTM network," *Applied Soft Computing*, vol. 108, Article ID 107463, 2021.

[46] 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, Article ID 110227, 2020.

[47] A. Kara, "Multi-step influenza outbreak forecasting using deep LSTM network and genetic algorithm," *Expert Systems with Applications*, vol. 180, Article ID 115153, 2021.

[48] R. P. Shetty and P. S. Pai, "Forecasting of COVID 19 cases in Karnataka state using artificial neural network (ANN)," *Journal of the Institution of Engineers (India): Serie Bibliographique*, vol. 102, no. 6, pp. 1201–1211, 2021.

[49] M. Hawas, "Generated time-series prediction data of COVID-19's daily infections in Brazil by using recurrent neural networks," *Data in Brief*, vol. 32, Article ID 106175, 2020.

[50] S. Chen, Y. Ren, D. Friedrich, Z. Yu, and J. Yu, "Prediction of office building electricity demand using artificial neural network by splitting the time horizon for different occupancy rates," *Energy and AI*, vol. 5, Article ID 100093, 2021.

[51] Y. Lin, I. Koprinska, and M. Rana, "Temporal convolutional neural networks for solar power forecasting," in *Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Glasgow, UK, July 2020.

[52] Y. Lin, I. Koprinska, and M. Rana, "Temporal convolutional attention neural networks for time series forecasting," in *Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Shenzhen, China, July 2021.

[53] C. Pan, J. Tan, D. Feng, and Y. Li, "Very short-term solar generation forecasting based on LSTM with temporal attention mechanism," in *Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, pp. 267–271, Chengdu, China, December 2019.

[54] G. Dudek, P. Pelka, and S. Smyl, "A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2021.

[55] R. Ghanbari and K. Borna, "Multivariate time-series prediction using LSTM neural networks," in *Proceedings of the 2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, pp. 1–5, Tehran, Iran, March 2021.

[56] A. Kumar Dubey, A. Kumar, V. García-Díaz, A. Kumar Sharma, and K. Kanhaiya, "Study and analysis of SARIMA and LSTM in forecasting time series data," *Sustainable Energy Technologies and Assessments*, vol. 47, Article ID 101474, 2021.

[57] R. F. Mustapa, N. Y. Dahlan, I. M. Yassin, A. Hamizah Mohd Nordin, M. E. Mahadan, and S. Mohamad, "Particle swarm optimization for NARX-ANN baseline energy modelling in educational buildings," in *Proceedings of the 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pp. 19–24, Selangor, Malaysia, June 2019.

[58] Q. Dang, D. Wu, and B. Boulet, "EV charging management with ANN-based electricity price forecasting," in *Proceedings of the 2020 IEEE Transportation Electrification Conference & Expo (ITEC)*, pp. 626–630, Chicago, IL, USA, June 2020.

[59] D. Wei, J. Wang, K. Ni, and G. Tang, "Research and application of a novel hybrid model based on a deep neural network combined with fuzzy time series for energy forecasting," *Energies*, vol. 12, no. 18, p. 3588, 2019.

[60] Z. Wen, L. Xie, Q. Fan, and H. Feng, "Long term electric load forecasting based on TS-type recurrent fuzzy neural network model," *Electric Power Systems Research*, vol. 179, Article ID 106106, 2020.

[61] Z.-K. Feng and W.-J. Niu, "Hybrid artificial neural network and cooperation search algorithm for nonlinear river flow time series forecasting in humid and semi-humid regions," *Knowledge-Based Systems*, vol. 211, Article ID 106580, 2021.

[62] F. Jiang, C. Zhang, S. Sun, and J. Sun, "A novel hybrid framework for hourly $PM_{2.5}$ concentration forecasting using CEEMDAN and deep temporal convolutional neural network," 2020, https://arxiv.org/abs/2012.03781.

[63] P. S. Budi Cahyo Suryo, I. Wayan Mustika, O. Wahyunggoro, and H. S. Wasisto, "Improved time series prediction using LSTM neural network for smart agriculture application," in *Proceedings of the 2019 5th International Conference on Science and Technology (ICST)*, pp. 1–4, Yogyakarta, Indonesia, July 2019.

[64] S. V. Belavadi, S. Rajagopal, R. Ranjani, and R. Mohan, "Air quality forecasting using LSTM RNN and wireless sensor networks," *Procedia Computer Science*, vol. 170, pp. 241–248, 2020.

[65] F. Hamami and I. A. Dahlan, "Univariate time series data forecasting of air pollution using LSTM neural network," in *Proceedings of the 2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, pp. 1–5, Lombok, Indonesia, October 2020.

[66] H. Jia and X. Zhou, "Water quality prediction method based on LSTM-BP," in *Proceedings of the 2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pp. 27–30, Hangzhou, China, 2020.

[67] P. Sharma, S. Singh, and S. D. Sharma, "Artificial neural network approach for hydrologic river flow time series forecasting," *Agricultural Research*, 2021.

[68] X. Zhao, H. Lv, S. Lv, Y. Sang, Y. Wei, and X. Zhu, "Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer," *Journal of Hydrology*, vol. 601, Article ID 126607, 2021.

[69] M. Zekić-Sušac, A. Has, and M. Knežević, "Predicting energy cost of public buildings by artificial neural networks, CART, and random forest," *Neurocomputing*, vol. 439, pp. 223–233, 2021.

[70] S. Ahmed, S.-U. Hassan, N. R. Aljohani, and R. Nawaz, "FLF-LSTM: a novel prediction system using forex loss function," *Applied Soft Computing*, vol. 97, Article ID 106780, 2020.

[71] E. Dave, A. Leonardo, M. Jeanice, and N. Hanafiah, "Forecasting Indonesia exports using a hybrid model ARIMA-LSTM," *Procedia Computer Science*, vol. 179, pp. 480–487, 2021.

[72] V. Hryhorkiv, L. Buiak, A. Verstiak, M. Hryhorkiv, O. Verstiak, and K. Tokarieva, "Forecasting financial time sesries using combined ARIMA-ANN algorithm," in *Proceedings of the 2020 10th International Conference on Advanced Computer Information Technologies (ACIT)*, pp. 455–458, Deggendorf, Germany, September 2020.

[73] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Computing & Applications*, vol. 32, no. 23, pp. 17351–17360, 2020.

[74] M. S. Farahani and S. H. R. Hajiagha, "Forecasting stock price using integrated artificial neural network and meta-heuristic algorithms compared to time series models," *Soft Computing*, vol. 25, no. 13, pp. 8483–8513, 2021.

[75] C. Kocak, E. Egrioglu, and E. Bas, "A new deep intuitionistic fuzzy time series forecasting method based on long short-term memory," *Journal of Supercomputing*, vol. 77, no. 6, pp. 6178–6196, 2021.

[76] P. Hewage, A. Behera, M. Trovati et al., "Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station," *Soft Computing*, vol. 24, no. 21, pp. 16453–16482, 2020.

[77] A. K. Yadav and H. Malik, "Short-term wind speed forecasting for power generation in hamirpur, Himachal Pradesh, India, using artificial neural networks," in *Applications of Artificial Intelligence Techniques in Engineering*, pp. 263–271, Springer, Singapore, 2019.

[78] G. K. Rahul, S. Singh, and S. Dubey, "Weather forecasting using artificial neural networks," in *Proceedings of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 21–26, Coimbatore, India, June 2020.

[79] M. Bou-Rabee, K. A. Lodi, M. Ali, M. Faizan Ansari, M. Tariq, and S. Anwar Sulaiman, "One-month-ahead wind speed forecasting using hybrid AI model for coastal locations," *IEEE Access*, vol. 8, pp. 198482–198493, 2020.

[80] K. Johny, M. L. Pai, and S. Adarsh, "Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall

forecasting," *Theoretical and Applied Climatology*, vol. 141, pp. 1–17, 2020.

[81] P. Unnikrishnan and V. Jothiprakash, "Hybrid SSA-ARIMA-ANN model for forecasting daily rainfall," *Water Resources Management*, vol. 34, no. 11, pp. 3609–3623, 2020.

[82] Z. Niu, Z. Yu, W. Tang, Q. Wu, and M. Reformat, "Wind power forecasting using attention-based gated recurrent unit network," *Energy*, vol. 196, Article ID 117081, 2020.

[83] H. Lu, Z. Ge, Y. Song, D. Jiang, T. Zhou, and J. Qin, "A temporal-aware LSTM enhanced by loss-switch mechanism for traffic flow forecasting," *Neurocomputing*, vol. 427, pp. 169–178, 2021.

[84] C. Ma, G. Dai, and J. Zhou, "Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2021.

[85] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, and S. Dong, "A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction," *Physica A: Statistical Mechanics and Its Applications*, vol. 583, Article ID 126293, 2021.

[86] H. Bohan and B. Yun, "Traffic flow prediction based on BRNN," in *Proceedings of the 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 320–323, Beijing, China, July 2019.

[87] S. George and A. K. Santra, "An improved long short-term memory networks with takagi-sugeno fuzzy for traffic speed prediction considering abnormal traffic situation," *Computational Intelligence*, vol. 36, no. 3, pp. 964–993, 2020.

[88] Z. I. E. Cicek and Z. K. Ozturk, "Optimizing the artificial neural network parameters using a biased random key genetic algorithm for time series forecasting," *Applied Soft Computing*, vol. 102, Article ID 107091, 2021.

[89] S. Fu, Y. Zhang, L. Lin, M. Zhao, and S.-s. Zhong, "Deep residual LSTM with domain-invariance for remaining useful life prediction across domains," *Reliability Engineering & System Safety*, vol. 216, Article ID 108012, 2021.

[90] Y. Chen, Y. Kang, Y. Chen, and Z. Wang, "Probabilistic forecasting with temporal convolutional neural network," *Neurocomputing*, vol. 399, pp. 491–501, 2020.

[91] A. Pandey and D. Wang, "TCNN: temporal convolutional neural network for real-time speech enhancement in the time domain," in *Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6875–6879, Brighton, UK, May 2019.

[92] E. Golshani and M. Ashtiani, "Proactive auto-scaling for cloud environments using temporal convolutional neural networks," *Journal of Parallel and Distributed Computing*, vol. 154, pp. 119–141, 2021.

[93] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," *International Journal of Forecasting*, vol. 36, no. 1, pp. 75–85, 2020.

[94] R. Li, X. Sun, T. Li, and F. Zhao, "A multi-objective learning speech enhancement algorithm based on IRM post-processing with joint estimation of SCNN and TCNN," *Digital Signal Processing*, vol. 101, Article ID 102731, 2020.

[95] D. Fan, H. Sun, J. Yao, K. Zhang, X. Yan, and Z. Sun, "Well production forecasting based on ARIMA-LSTM model considering manual operations," *Energy*, vol. 220, 2021.

[96] K. He, L. Ji, C. W. D. Wu, and K. F. G. Tso, "Using SARIMA-CNN-LSTM approach to forecast daily tourism demand," *Journal of Hospitality and Tourism Management*, vol. 49, pp. 25–33, 2021.

[97] C. Kocak, A. Z. Dalar, O. Cagcag Yolcu, E. Bas, and E. Egrioglu, "A new fuzzy time series method based on an ARMA-type recurrent pi-sigma artificial neural network," *Soft Computing*, vol. 24, no. 11, pp. 8243–8252, 2020.

[98] E. Egrioglu, U. Yolcu, and E. Bas, "Intuitionistic high-order fuzzy time series forecasting method based on pi-sigma artificial neural networks trained by artificial bee colony," *Granular Computing*, vol. 4, no. 4, pp. 639–654, 2019.

[99] X. Fu, B. Yang, and S. Wang, "Modeling and control of plasma horizontal displacement for HL-2A tokamak based on LSTM," *Fusion Engineering and Design*, vol. 162, Article ID 112107, 2021.

[100] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902–924, 2017.

[101] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep learning for time series forecasting: a survey," *Big Data*, vol. 9, no. 1, pp. 3–21, 2021.

[102] K. U. Jaseena and B. C. Kovoor, "Deterministic weather forecasting models based on intelligent predictors: a survey," *Journal of King Saud University—Computer and Information Sciences*, 2020.

[103] R.-Y. Sun, "Optimization for deep learning: an overview," *Journal of the Operations Research Society of China*, vol. 8, no. 2, pp. 249–294, 2020.

[104] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2017, https://arxiv.org/abs/1412.6980.

[105] T. Szandała, "Review and comparison of commonly used activation functions for deep neural networks," in *Bio-Inspired Neurocomputing*, A. K. Bhoi, P. K. Mallick, C.-M. Liu, and V. E. Balas, Eds., vol. 903, pp. 203–224, Springer, Singapore, 2021.

[106] D. Lee, "Comparison of reinforcement learning activation functions to improve the performance of the racing game learning agent," *Journal of Information Processing Systems*, vol. 16, no. 5, pp. 1074–1082, 2020.

[107] J. Feng and S. Lu, "Performance analysis of various activation functions in artificial neural networks," *Journal of Physics: Conference Series*, vol. 1237, no. 2, Article ID 22030, 2019.

[108] G. Alkhayat and R. Mehmood, "A review and taxonomy of wind and solar energy forecasting methods based on deep learning," *Energy and AI*, vol. 4, Article ID 100060, 2021.

[109] D. E. Birba, *A Comparative Study of Data Splitting Algorithms for Machine Learning Model Selection*, KTH Royal Institute of Technology, Stockholm, Sweden, 2020.

[110] H. Engbers, A. A. Alla, M. Kreutz, and M. Freitag, "Applicability of algorithm evaluation metrics for predictive maintenance in production systems," in *Proceedings of the 2020 6th IEEE Congress on Information Science and Technology (CiSt)*, pp. 413–418, Agadir-Essaouira, Morocco, June 2020.

[111] T. Salman, A. Ghubaish, D. Unal, and R. Jain, "Safety score as an evaluation metric for machine learning models of security applications," *IEEE Networking Letters*, vol. 2, no. 4, pp. 207–211, 2020.

[112] W. Jiang, "Applications of deep learning in stock market prediction: recent progress," *Expert Systems with Applications*, vol. 184, Article ID 115537, 2021.