



Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model

R. Rathipriya¹ · Abdul Aziz Abdul Rahman² · S. Dhamodharavadhani¹ · Abdelrhman Meero² · G. Yoganandan³

Received: 5 January 2022 / Accepted: 22 September 2022

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Demand forecasting is a scientific and methodical assessment of future demand for a critical product. The effective Demand Forecast Model (DFM) enables pharmaceutical companies to be successful in the global market. The purpose of this research paper is to validate various shallow and deep neural network methods for demand forecasting, with the aim of recommending sales and marketing strategies based on the trend/seasonal effects of eight different groups of pharmaceutical products with different characteristics. The root mean squared error (RMSE) is used as the predictive accuracy of DFMs. This study also found that the mean RMSE value of the shallow neural network-based DFMs was 6.27 for all drug categories, which was lower than deep neural network models. According to the findings, DFMs based on shallow neural networks can effectively estimate future demand for pharmaceutical products.

Keywords Deep learning models, Demand forecasting · Pharmaceutical industry · Shallow neural network models

1 Introduction

The demand forecasting model is an essential machine learning tool used by the majority of international pharmaceutical corporations with a dual objective of matching the supply with any eventual rise or fall in the demand for the products and keeping the inventory as minimum as possible. If such a system is effectively deployed, it would

contribute to the efficient management of their supply chain [1]. This will lead to improvement in the overall profitability of the organization. It should be noted that the supply chain of pharmaceutical companies consists of a series of activities that are complex and sometimes intertwined. As such, these complexities add up to the numerous challenges that are faced by pharmaceutical companies around the globe. Efficient supply chain management is critical for any organization in the essential service sector, this holds very true in the case of the pharmaceutical industry too. The current covid-19 pandemic which was originally considered as an epidemic pertaining to China has become an unsolicited global pandemic. This pandemic is one of the unique situations for the last 100 years, has affected both the supply and demand dimensions of even the well-established global pharmaceutical industry. On the one hand, sudden outbreaks led to waves of rising demand, and on the other hand, lockdowns, discoveries of vaccines, pills, and medicines contributed to the unpredictability of supply for the pharmaceutical products. Demand Forecasting Models (DFMs) help the companies to stock or produce the required quantity of pharmaceutical products. Efficient DFMs greatly support pharmaceutical companies to accurately predict the demand and to make

✉ Abdul Aziz Abdul Rahman
a.abdulrahman@ku.edu.bh

R. Rathipriya
rathipriyar@gmail.com

S. Dhamodharavadhani
vadhanimca2011@gmail.com

Abdelrhman Meero
ar.meero@ku.edu.bh

G. Yoganandan
g_yoganandan@yahoo.com

¹ Department of Computer Science, Periyar University, Salem, India

² Kingdom University, Riffa, Bahrain

³ Department of Management Studies, Periyar University, Salem, India

appropriate decisions in their production and supply chain operations. In a nutshell, efficient DFMs not only enable to surmount the complexity in the supply chain but create a competitive advantage for global pharmaceutical companies [2].

Most pharmaceutical companies used machine learning-based forecasting techniques for identifying the future demand in advance [3]. The main advantage of using artificial neural networks or data mining techniques is its ability to give high-quality predictions while consuming lesser time. Practically it is impossible to obtain such higher accuracies using the existing standard statistical forecasting methods [4]. In [5], the authors mentioned that the companies were not able to deliver the right quantity of products on time to their suppliers. Foremost reason being, that these companies failed to identify the rapid changes in the quantity of demand and were unaware of the other competitor's demand in the supply chain management. Therefore, authors have suggested these organizations use a predictive tool for managing their demand forecasting and related functions management.

The very popular and successful linear forecasting time-series models are Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and Autoregressive Moving Average (ARMA) [6], but its prediction quality suffered from their assumption of direct actions. This leads to the rapid development of machine learning methods, such as data mining techniques and Artificial Neural Networks (ANN) for the time series data forecasting that provides better prediction accuracy [1]. Recently, the growth of the ANN-based time-series forecasting model was tremendous. In addition, it paves the way for making accurate predictions by handling nonlinear input and output data. Linear forecasting methods mentioned above are not suitable for modeling data containing non-linear behavior [7]. In recent times, ANN has been used widely in time-series forecasting for better prediction quality [8]. Therefore, numerous studies have proposed in the literature to study the performance of ANN-based time-series forecasting in comparison using conventional forecasting methods. On the other hand, these studies have shown that there was no benchmark neural network parameter setting for the time-series domain. Also, it is purely dependent on the problem domain [1].

Similarly, in this paper, shallow neural networks (Probabilistic Neural Network-P_NN, Generalized Regression Neural Network-GR_NN, and Radial Basis Function Neural Network-RBF_NN), and deep learning neural network (Long-Term Short Memory Neural Network) are used to develop ANN-based DFMs. Moreover, a comparative study will be illustrated to determine whether the shallow neural network-based demand forecasting

model performs better than the deep neural network or not for time series pharmaceutical data.

The main contribution of this work, in summary, is as follows:

- Development of a methodology to select the best demand forecasting models (DFMs) for pharmaceutical data using statistical and neural network models for a time-series sales database.
- Investigate whether shallow and deep neural network methods such as RBF_NN, P_NN, GR_NN, LSTM and Stacked LSTM lead to more accurate ATC thematic drug prediction than non-neural network models.
- Investigate the performance of ARIMA model, Shallow Neural Network and Deep Neural Network models in dealing with time series dataset.

The authors have appropriately segmented the contents of this research paper into 5 sections as follows. Section 1 was earmarked for providing information pertaining to the introduction of the study. Section 2 describes the review of the related literature and highlights the research gaps. Section 3 discusses a brief explanation of the time-series forecasting models and evaluative measures. Section 4 includes the proposed methodology and experimental analysis. At the last, Sect. 5 summarizes the present work with possible future work.

2 Review of literature

In general, demand forecasting is the biggest challenge for a big data application, which may use additional data information such as value and veracity for better prediction. This research mainly focused on demand forecasting for big data analytics [2]. It explains the incorporation of the various sources of data into a possible prediction process but needs a data analyst to accomplish the task, and adequate technical base, and funds in technology [9].

Past literature highlight the insufficiency of statistical methods for optimization of the pharmaceutical assembly line and a production facility. Keeping this constraint in mind, some researchers [10] have adopted an improvised approach that enabled combining simulation methods and analysis of data. It recommended the pharmaceutical unit should be optimized for performance optimization [11].

In another study [12], Convolutional Neural Network (CNN) method, NN-based Genetic Algorithm, and NN-based Particle Swarm Optimization (PSO) methods are executed to predict the demand for energy consumption. According to the simulation results, neural network-based energy forecasting techniques predicted the demand for energy consumption with greater accuracy than the CNN approach.

Genetic Algorithms (GA), Bayesian structure, and Reinforcing Artificial Neural Networks (RANN) design have been developed by authors [13] to forecast the short-run demand for irrigation. The experimental study of these models highlighted the superiority of the ANN model as the results it produced showed a greater predictive quality when compared to the traditional algorithms.

In yet another research [14], Iran's annual energy usage was forecasted using three different combinations of the Autoregressive Integrated Moving Average Model (ARIMA) –Adaptive Neural Fuzzy Inference System model (ANFIS). The findings of these models indicated that hybrid models have higher accuracy than the ARIMA or ANFIS model individually for the forecasting of energy use.

Many researchers attempted to come out with near-perfect, zero-error forecasting models. Interestingly, in one of such studies, a neuro-fuzzy approach was employed [15] to forecast the future demand from past sales data. ANFIS was implemented along with ANN as a “neuro-fuzzy approach” despite the specification it shows the approach's effectiveness. There is a notable study [16] that addressed state-of-the-art approaches and significant forecasting issues concerning the demand for drugs of pharmaceutical companies. They have used three forecasting scenarios for demand. It concluded that the “symbolic regression-based forecasting model” had the better model fit and low error rate to demand observation data than the other models. Few authors attempted to forecast the revenue stream of pharmaceutical companies. One such study [17] aimed to introduce a new forecasting approach to the pharmaceutical companies' revenue estimates. This approach implemented a combination of NN-based prediction methods for the time series dataset.

Due to the COVID-19 pandemic, the breakdown of supply chain function has created a shortage of ambulatory medicines [18]. The pandemic managed to increase demand for pharmaceutical products and reported shortages during the early time of the outbreak [3]. The principal objective of this study was to find the most effective animal drugs based on drug distribution routes like those for humans. This generates demands for investment in the efficiency, health, and feasibility of drug delivery systems in testing [17–19]. The Long Short-Term Memory (LSTM) model was used to enhance the predictive quality of the time series dataset. The demand forecasting approach based on a multi-layer LSTM model was proposed in this paper [20] which has a robust capacity to predict extremely rising and falling demand data. The LSTM model is compared with the CNN model, which indicated that the LSTM model has a more robust prediction performance [21].

A theoretical Fuzzy Neural Network (FNN) for the development of forecasting techniques was employed by another study [22]. FNN has the capability to remove the insignificant weights to develop Fuzzy IF–THEN rules for further learning. The output of FNN is pooled with ANN results for the better prediction of time-series data. The statistical and computational intelligence approaches such as “exponential smoothing K-nearest Neighbors (KNN), Supporting Vector Machines (SVM), ARIMA, ANN, RNN, and LSTM are proposed in the literature as the forecasting techniques for various problem domains [23].

In the study [24], the authors introduced a forecasting approach to analyze the existing sales data and contrasted it with the various machine learning models to predict future demand. The demand forecasting method based on LSTM was developed to capture the nonlinearities present in the e-commerce business. The findings of this study showed that LSTM outperformed predictive univariate techniques. The study [25] proposed a hybrid approach that investigated the long time series of memory to integrate Autoregressive Fractionally Integrated Moving Average (ARFIMA) and feed-forward neural networks (FFNN). The authors said that the hybrid model provided better predictive accuracy than other standard approaches.

Table 1 lists some neural-network-based predictive models used for different demand forecasting operations. After a detailed investigation of different methods, strategies, techniques, and methodologies developed and implemented in previous studies, the ANN-based demand forecasting model is identified as the appropriate demand forecasting model for the time series dataset. Therefore, in this research work, ANN-based demand forecasting models will be developed for Anatomical Therapeutic Chemical (ATC) thematic drug sales dataset to predict the weekly demand.

3 Methods and materials

Autoregressive Integrated Moving Average Model (ARIMA) is a widely used forecasting model due to its simplicity and ability to generalize to non-stationary series. A detailed description of this model can be found in [14, 23, 35].

3.1 Shallow neural network model: SNN

The popular SNN models used for time-series data forecasting are “Radial Basis Function (RBF) Neural Network (RBF_NN) [40–45] Generalized Regression Neural Network (GR_NN) and Probabilistic Neural Network (P_NN)” [41, 46–51]. These models are used to build demand forecasting models for pharmaceutical products. It

Table 1 List of popular NN-based forecasting models

Model/Algorithm	Demand type and period	Factor	References
ANN model with an Adaptive backpropagation algorithm (ABPA)	Long-term forecasts and monthly	Electricity load demand	[26]
CNN- LSTM	Short-term forecasts and daily	Gold price	[27]
MLP and LSTM	Long term forecasts and weekly monthly and yearly	Vehicular traffic flow	[28]
ANN	Long-term forecasts and monthly	Monthly electricity demand	[29]
Deep learning neural networks (DLNN)	Long-term forecasts and monthly	Forecast aviation demand	[30]
Hybrid CNN with GRNN	NA	Predicting yarn strength index	[31]
Multiple linear regression (MLR); Symbolic regression (SR)	Weekly forecasting	Demand forecasting for the pharmaceutical industry	[32]
Machine learning tools	Yearly forecasting	Annual inflation rate forecasts	[33]
Operation research (OR) forecasting techniques	Short, Mid and long term horizon	Branded and generic drugs forecasting	[34]
Statistical methods such as ARIMA and holt' s-winter (HW) model	Short, Mid and Long term horizon forecasting	Demand forecasting for the pharmaceutical products	[35]
Feedforward backpropagation neural networks (FFBP_NN), GRNN and (RBF_NN)	Monthly	Accurate river flow forecasts	[36]
PSO-GRNN	Yearly	Railway freight volume prediction	[37]
MLP_NN, RBF_NN, GRNN	Daily	Daily temperature prediction	[38]
Machine learning-based approaches for forecasting	Short, Mid and Long term horizon	Air temperature forecasting	[39]
RBF_NN, condensed polynomial neural network (CPNN), and GA-based CPNN	Daily	Stock Index price forecasting	[40]

can be made up of an input layer(IL), 1 or 2 hidden layers (HL), and an output layer (OL). It performs regression where the target variable of the problem is continuous. The typical network configuration diagram for SNN is shown in Fig. 1. The IL has one neuron for each predictor variable (X_1, X_2, \dots, X_n). In HL, the predictor variables' value and its target value for each category in the dataset (H_1, H_2, \dots, H_n) were stored in each neuron. When the input data are given to the HL, it calculates the distance from the neuron's mean point for each test category using a specified

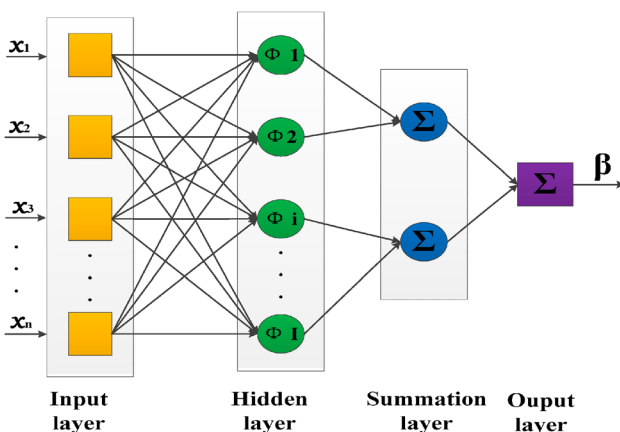


Fig. 1 Common architecture of SNN

distance measure. This distance value is given as the input to the RBF of HL. The output of the HL is given to the next layer (pattern/summation layer or output layer) for intermediate output. The decision layer of P_NN/ GR_NN or output layer of RBF_NN is the last layer which calculates the final output of the whole NN [The detailed descriptions of these neural networks are given below in Table 2.

3.2 The description of these shallow neural network models is given below:

3.2.1 RBF_NN

The input layer of the RBF_NN receives the input data and feeds it into the hidden layer of the RBF_NN. The calculation that takes place within the hidden layer uses the radial basis kernel function. The output layer performs the prediction task.

3.2.2 P_NN

The input layer of P_NN computes the euclidean distance from the input time-series data to the training input time-series data. It generates a vector whose components indicate how close the input is to the training input. The second layer computes the contribution for each class of inputs and

Table 2 Description of shallow neural network models

	RBF_NN	P_NN	GR_NN
Type	Feedforward neural network	Feedforward neural network, and a kind of RBF_NN	Variation to radial basis neural networks
Number of layers	3 layers namely “input layer, 1Hidden layer and output layer”	4 layers namely “input layer Pattern layer Summation layer Output layer”	4 layers namely “input layer/ Hidden layer, pattern layer/ Summation layer Output layer”
Transfer function /Kernel function in HL	RBF	RBF	RBF
Smoothing parameter (σ)	Yes	Yes	Yes
Application	Classification and forecasting	Pattern recognition, and forecasting	Regression, prediction, and classification

creates its net output as a vector of probabilities using the probability density function. Finally, an activation function on the second layer’s output selects the maximum probability neuron. It generates 1 for the target classes and 0 for non-target classes.

3.2.3 GR_NN

The input layer of the GR_NN receives the time series data and feeds it into the hidden layer of the GR_NN. The calculation that takes place within the hidden layer uses the radial basis kernel function.

In the summation/ pattern layer, the denominator summation unit adds up the weight values of the hidden neurons. The numerator summation unit adds up the weight values multiplied by the actual target value for each hidden neuron. The decision layer divides the value accumulated in the numerator summation unit by the value in the denominator summation unit and uses the result as the predicted target value. Finally, it performs the prediction task.

3.3 Deep Neural Network Model: LSTM

LSTM consists of an input layer, more hidden layers, and an output layer [52–54]. It is a Recurrent Neural Networks (RNN) with a deep learning architecture containing the LSTM units in hidden layers. The key benefit of RNNs is that they can be used for both classification and regression by using historical time points or data. This network has a single IL, more HLs, and a single OL. The neurons in the hidden layers are fully inner-connected which includes memory cells and related gate units. The general architecture diagram of LSTM is given in Fig. 2. An LSTM neural network consists of a cell, an input gate, an output gate, and a forgot gate. A cell remembers values at

arbitrary intervals. The input gate manages the flow of information into the cell. The output gate manages the flow of information out of the cell. Similarly, forget gates manage the necessary or unnecessary information. It is useful for predictions and classification problems.

3.4 Performance metrics

Root Mean Squared Error (RMSE) is defined as the square root of sum of the squared differences between predicted sales values and actual sales values. It is given in Eq. 5. The smaller RMSE value signifies the best model. Normalized RMSE is calculated as the RMSE of the dataset divided by the mean of the dataset. The percentage of error (PE) is defined as the sum of the differences between the forecasted sales values and the actual sales values, which is divided by the sum of the forecast sales values and multiplied by 100.

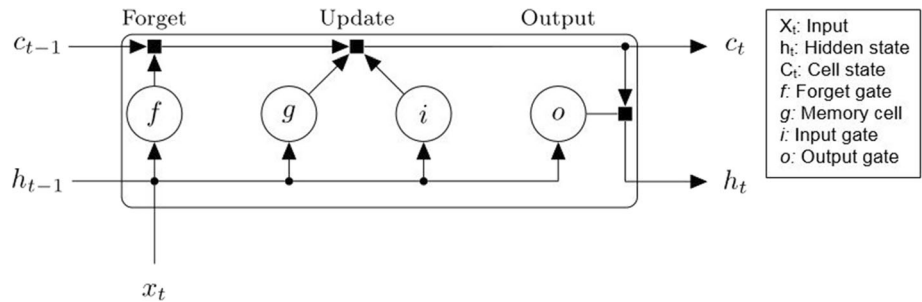
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Sales_P_i - Sales_A_i)^2} \tag{5}$$

$$Normalized\ RMSE = \frac{RMSE}{mean(Sales_A_i)} \tag{6}$$

$$PE = \left(\frac{\sum_{i=1}^n abs(Sales_P_i - Sales_A_i)}{\sum_{i=1}^n (Sales_P_i)} \right) * 100 \tag{7}$$

In Eqs. 5, 6 and 7, Sales_P denotes predicted sales values, n represents the number of instances in the dataset, Sales_A is the actual sales values.

Fig. 2 LSTM network architecture



4 Proposed methodology

In this work, ARIMA, neural network variants such as GR_NN, RBF_NN, P_NN, LSTM, and Stacked LSTM models are used to build the DFMs for pharmaceutical

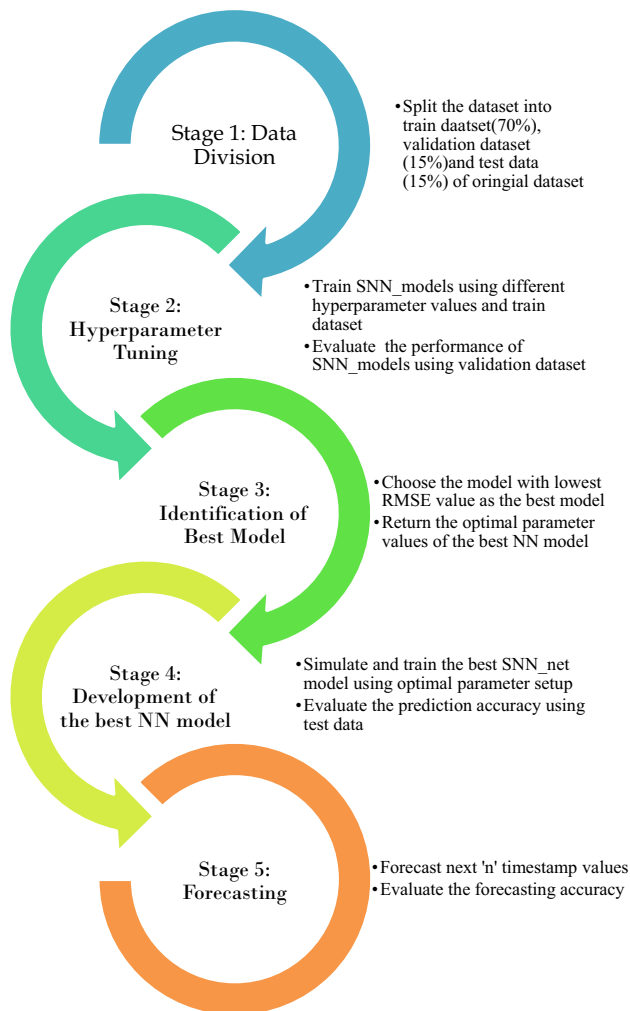


Fig. 3 Workflow of proposed methodology

time-series sales data. The graphical abstract of the proposed methodology is given in Fig. 3. Their performance was compared based on key performance indicators such as RMSE and PE.

Algorithm 1 lists the sequences of the step for the Neural Network-based Prediction Model for Demand Forecasting. The first step is the initialization of the model parameters and hyperparameters of RBF_NN, GR_NN, P_NN, and LSTM models. For the initial run of these models, the parameters of the neural network model such as weights, biases, and other hyper-parameters are initialized randomly. In this work, dataset split ratio is 70%:15%:15% with 10 K-fold cross-validation. Epoch represents the maximum number of iterations. The objective of the run is to minimize the RMSE value of these models. Secondly, fix the value for threshold (θ) which denotes the RMSE value of the model. It is purely dependent on the dataset and is user-defined. Thirdly, simulate the NN_net model with initial values and training dataset. After that, evaluate its performance. If desired performance is reached then stop the process the training process. Otherwise, repeat the training process with the different values of hyperparameters as specified in Table 3 until the termination criterion is reached. Fourth, build the NN_net model using optimal parameter values. Fifth, test the NN_net model using testing data. Sixth, estimate the testing performance of the NN_net model. In the next step, identify the best model from the list of NN_models using a benchmark performance indicator (called RMSE). And then simulate the best model using optimal neural network configuration parameter values for the epidemiological data prediction. Finally, forecast the values for the next 'n' time stamp.

Table 3 Neural network model parameters setup

Model parameters	Shallow NN model	Deep NN model
Models	RBF_NN, P_NN and GR_NN	LSTM and stacked LSTM
Hidden layer (HL)	1	3–5
Number of neurons in HL	10 – 20	15–20
Learning algorithm	Bayesian regularization	ADAM
Kernel functions	Radial basis function	Gaussian process with radial basis function
Activation/Transfer function	Hyperbolic tangent activation function	State activation function: hyperbolic tangent activation function Gate activation function: sigmoid
Performance metric	RMSE	RMSE

Algorithm 1: Neural Network based Prediction Model for Demand Forecasting

Input: Dataset Y

Output: Predicted Values, Forecast Values, Performance Metrics

Step 1: Input the historical data for forecasting

Step 2: Set number of nodes in the input layer= number of time points in the dataset

Step 3: Set number of hidden layer=1; Initialize number of nodes in the hidden layer

Step 4: Set number of nodes in the output layer= 1

Step 5: Set Threshold= θ and Epoch= max_iteration

Step 6: NN_models={ 'P_NN', 'RBF_NN', 'GR_NN', 'LSTM', 'Stacked LSTM' }

Step 7: For each NN_model in NN_model

Initialize weight and bias of NN models

Initialize hyperparameters of NN models

Perform validation of the initial NN models

While(performance(NN_net) > threshold)

Select number of nodes in hidden layer(NH) [Range is 5:25]

Select feedback delay(FD) [Range is 1:5]

Select learning algorithm (LA) ['trainbr', 'trainlm', 'trainscg']

Select activation function for hidden layer

(AFH) ['tansig', 'logsig', 'elliotsig', 'purelin']

Select activation function for output layer

(AFO) ['tansig', 'logsig', 'elliotsig', 'purelin']

Fix the default values for other parameters

Train NN_net(NH,FD,LA,AFH,AFO)

Perform validation of NN_net model

end

Identify the optimal values of model parameters and hyperparameters

Build NN_net model using optimal values of model parameters and hyperparameters

Test NN_net model using test dataset

Evaluate the performance of NN_net model

RMSE_NN{NN_model} = RMSE(NN_net)

end

Step 8: Choose NN_model with the lowest RMSE value as the best NN_net model

Step 12: Simulate the best NN_net model with optimal model parameters and hyperparameter values for prediction

Step 13: Evaluate the prediction

Step 16: Forecast the values for new data

4.1 Data description

The weekly sales data for pharmaceutical products is taken from the Kaggle website [55]. This database contains 600,000 sales instances of 57 drugs during the period 2014–2019 and includes attributes such as date and time of sale, brand name, and amount of these drugs. In this work, the demand for specific types of drugs is predicted, rather than specific products. Therefore, the selected group of 57 drugs is classified into eight sections of anatomical treatment chemical (ATC) thematic analysis and it is classified as shown in Fig. 4. The sales trends of all categories of drug datasets are shown graphically in Fig. 5. It provides a clear picture of the nature of the dataset taken for the study.

4.2 Result and discussions

Table 3 showed model parameter and hyperparameter setup for implementing shallow and deep neural network models for demand forecasting. The details of the hyperparameters of demand forecasting models such as ARIMA, RBF_NN, GR_NN, P_NN, LSTM and stacked LSTM models are shown in Table 4.

Tables 5 and 6 showed the RMSE value of the ARIMA model, Shallow neural network-based demand forecasting model, and deep neural network-based demand forecasting model respectively.

From Table 5, it is observed that all three shallow neural network models have relatively similar values for all eight drug categories. The values highlighted in the red shade indicate the lowest RMSE value. From the results, it can be seen that GR_NN-based DFM works better for most drug types and this reduces the RMSE from 5–23%.

M01AB	Anti-inflammatory and antirheumatic products, non-steroids, Acetic acid derivatives and related substances
M01AE	Anti-inflammatory and antirheumatic products, non-steroids, Propionic acid derivatives
N02BA	Analgesics and antipyretics, Salicylic acid and derivatives
N02BE/B	Analgesics and antipyretics, Pyrazolones and Anilides
N05B	Psycholeptics drugs, Anxiolytic drugs
N05C	Psycholeptics drugs, Hypnotics and sedatives drugs
R03	Drugs for obstructive airway diseases
R06	Antihistamines for systemic use

Fig. 4 Components of drug dataset

As for ARIMA based DFM concerned, average RMSE for the ARIMA model for all drug categories is 15.02. The overall percentage of prediction accuracy for all drug categories using the ARIMA model is 83.43%. If the prediction accuracy of the model is close to 100%, it is considered to be the best model. However, for real-time datasets, achieving 90% forecast accuracy is always desirable.

It is observed from Table 6 that there was a little difference in RMSE of LSTM and stacked LSTM for all drug types. But there was a slight difference in the remaining RMSE value. As said above, the values highlighted in the red shade indicate the lowest RMSE value for the Deep NN-based DFM. For M01AE and R06 drugs, Deep_NN-based DFMs worked better than Shallow_NN-based DFMs. For the rest of the drugs, Shallow_NN-based DFMs showed the best performance. These best models have a predictable accuracy of over 90% for all drugs. This is also one of the interesting findings of this research work.

LSTM and stacked LSTM have shown almost equal efficacy for all drug types except M01AE. Another interesting fact is that shallow NN-based DFM outperformed deep NN-based DFM. This is because shallow NN has a strong ability to tolerate noise in datasets and learn patterns from small databases and is finally easy to create a model with minimal parameters.

Table 7 showed the PE values of the shallow neural network-based demand forecast model and the deep neural network-based demand forecast model. The PE of the best model ranges from 0.2% to 13.57%. Therefore, these DFMs have good predictive accuracy for the (ATC) thematic drug dataset.

The performance of DFMs for eight drug types is compared using a benchmark performance indicator called the RMSE value. This is illustrated graphically in Fig. 6. Figure 7 illustrates the comparison of the minimum RMSE values of all eight drug types obtained using shallow NN and deep NN-based demand forecasting models.

Table 8 displayed the mean RMSE value of all demand forecasting models for eight drug types. It is found that shallow NN models (such as GR_NN, RBF_NN, and P_NN) have almost identical mean RMSE values for all eight categories of the drug. Therefore, it is concluded that the shallow NN-based demand forecasting model worked best in predicting demand for pharmaceutical products.

5 Summary and scope

The main scope of demand forecasting models is stated as follows:

- Achieving the planned production target
- Stabilize production against demand

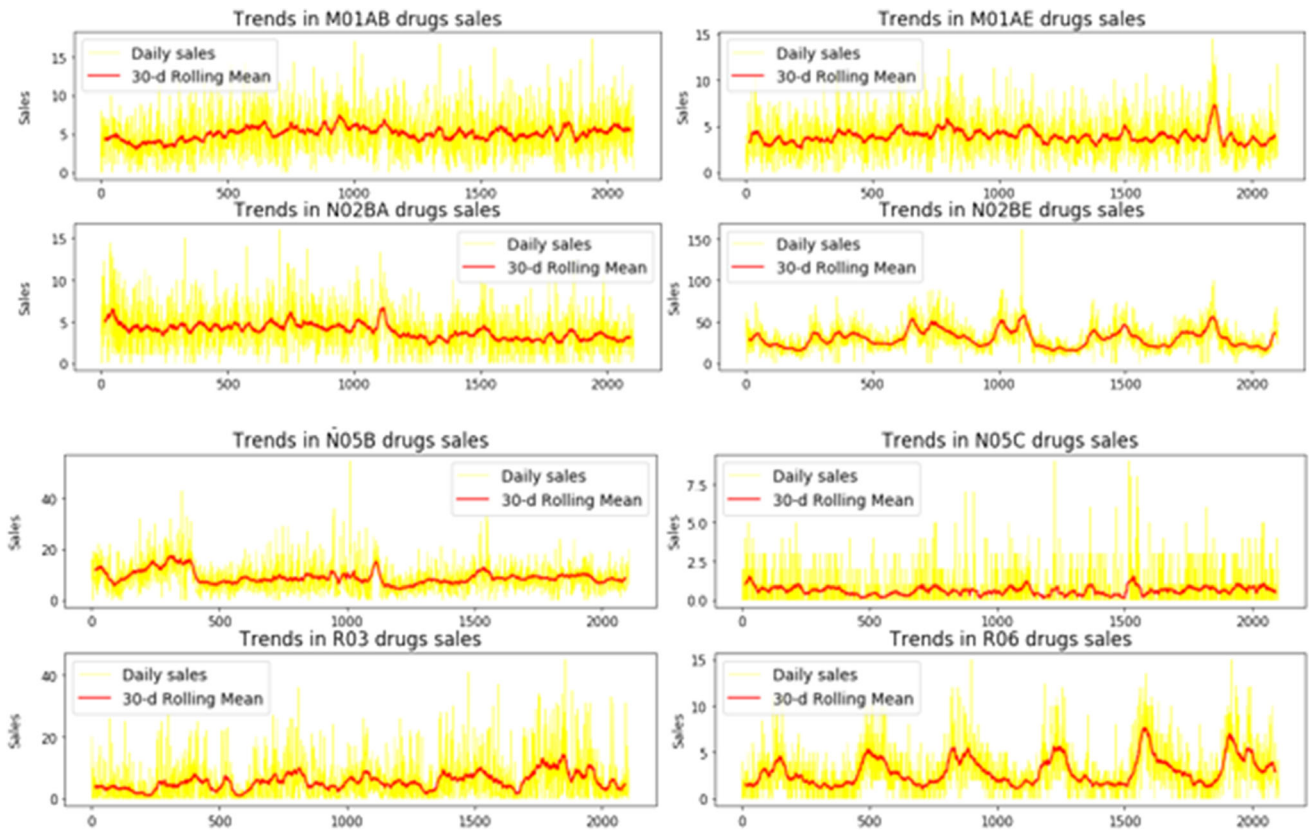


Fig. 5 Sales trends of the eight categories of the ATC drugs

Table 4 List of hyperparameters for neural network model

DFM	Hyperparameters
ARIMA	α , Alpha; β , Beta; γ , Gamma
GR_NN	Smoothing factor(σ), Count of hidden neurons, feedback delays
RBF_NN	Smoothing factor(σ), Count of hidden neurons, feedback delays
P_NN	Smoothing factor(σ), Count of hidden neurons, feedback delays
LSTM	Count of hidden neurons in HL, Count of hidden layers, and dropout layer
Stacked LSTM	Count of hidden neurons in HL, Count of hidden layers, and dropout layer

Table 5 RMSE and normalized RMSE for ARIMA and shallow NN- DFM

Drug category	RMSE				NormalizedRMSE			
	ARIMA	P_NN	GR_NN	RBF_NN	ARIMA	P_NN	GR_NN	RBF_NN
M01AB	8.65	7.24	5.8	7.55	0.246	0.206	0.165	0.215
M01AE	8.75	16.11	16.32	16.99	0.322	0.593	0.601	0.625
N02BA	5.65	5.65	5.6	5.93	0.209	0.209	0.207	0.219
N02BE	51.13	0.16	0.16	0.17	0.245	0.001	0.001	0.001
N05B	12.3	0.5	0.66	0.15	0.199	0.008	0.011	0.002
N05C	2.82	2.43	2.04	3.24	0.681	0.587	0.492	0.782
R03	26.04	1.52	1.44	1.77	0.677	0.04	0.037	0.046
R06	8.66	16.03	16.04	16.92	0.428	0.793	0.793	0.837

- Future growth of sales
- Longstanding investment planning
- Budget preparation and Sales budget
- Control of raw materials

Table 6 RMSE and normalized RMSE for deep NN- DFM

Drug category	RMSE		Normalized RMSE	
	LSTM	Stacked LSTM	LSTM	Stacked LSTM
M01AB	9.24	9.2	0.214	0.213
M01AE	9.56	11.5	0.499	6.085
N02BA	6.25	5.95	0.179	0.170
N02BE	50.85	51.8	0.329	0.335
N05B	13.22	13.59	0.323	0.332
N05C	3.04	3.03	0.103	0.103
R03	26.22	26.49	0.239	0.241
R06	8.47	8.18	0.566	0.546

It is evident that in most cases, shallow neural network models showed better performance in terms of predictive accuracy. The most important fact identified from in this study is that a deep neural network has not shown the best performance against this drug sales dataset because the number of samples used for building the demand forecasting model was less. In short, shallow neural network-based time-series demand forecasting models have offered potentially useful recommendations to the pharmaceutical

companies. The weekly sale data analysis of ATC drug products has proven to be a useful to identify the days for implementation of special marketing campaigns to promote their sales.

6 Conclusion and future work

The main objective of this work was to provide a methodology for demand forecasting to help pharmaceutical companies to manufacture or stock the right quantity of pharm products. Shallow Neural networks (GRNN, RBF_NN, and P_NN) and Deep Neural Network (LSTM and Stacked LSTM) models for eight categories of the ATC thematic drug. In this work, shallow neural network models based DFM have performed well for five drug categories out of eight, while ARIMA model has worked best for remaining three drug categories. From the experimental study, it has been known that the single demand forecasting model is not optimal for all drug categories. For ATC thematic drug time-series dataset, shallow NN-based DFM performed better with a mean RMSE value of 6.27 for all eight drug types. Future research work to enhance the predictive accuracy of the time series forecasting model

Table 7 Percentage of error for NN- DFMs

Drug category	Statistical DFM	Shallow NN- DFM			Deep NN- DFM	
	ARIMA	P_NN	GR_NN	RBF_NN	LSTM	Stacked LSTM
M01AB	7.02	5.87	5.8	6.02	7.5	7.46
M01AE	12.42	22.86	22.88	25.04	13.57	165.31
N02BA	4.38	4.38	4.34	4.55	4.85	4.61
N02BE	69.03	0.22	0.21	0.23	68.65	69.93
N05B	18.54	0.75	0.75	2.28	19.92	20.48
N05C	0.39	0.34	0.33	0.36	0.43	0.42
R03	9.11	0.53	0.53	0.56	9.18	9.27
R06	11.70	21.65	21.41	22.55	11.44	11.05

Fig. 6 RMSE value for each category

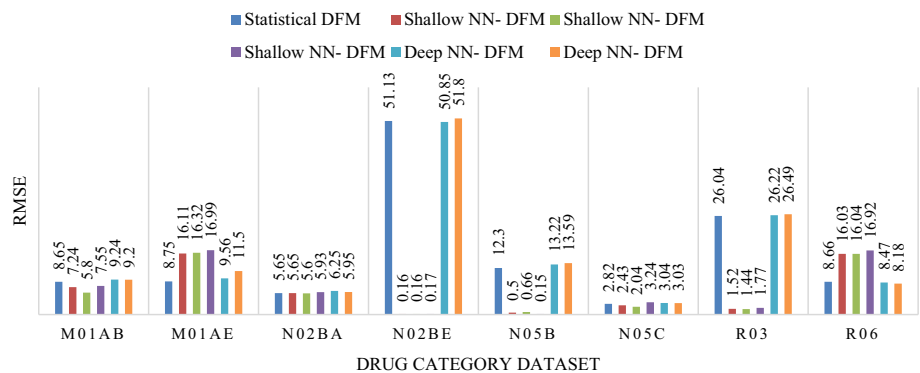


Fig. 7 Minimal RMSE value for eight categories

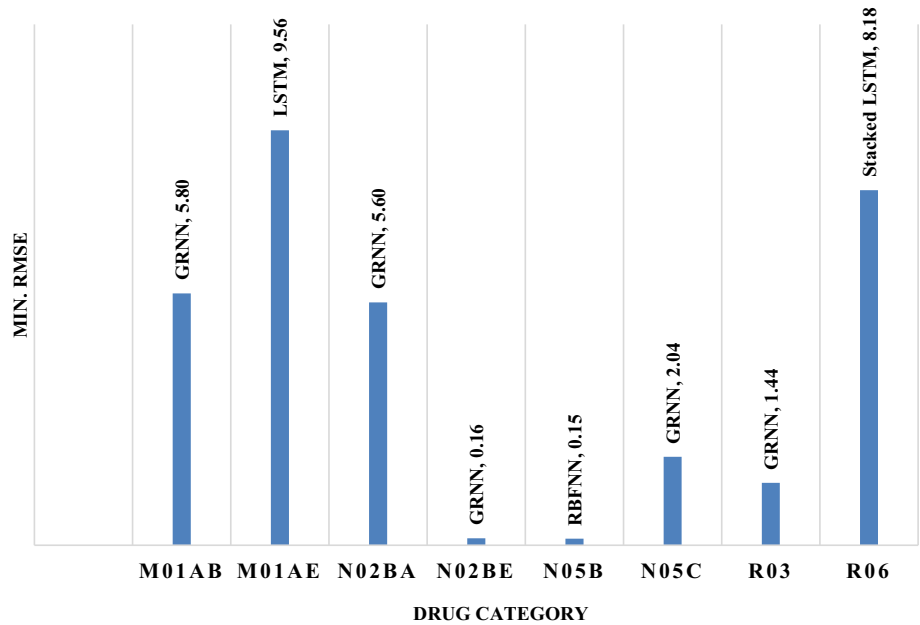


Table 8 Overall performance of all DFM

DFM	Mean RMSE
ARIMA	15.50
P_NN	6.21
GR_NN	6.01
RBF_NN	6.59
LSTM	15.86
Stacked LSTM	16.22

will include more cases, explore different accuracy measurements, and develop shallow and deep neural network models with optimal hyperparameters.

Funding statement The authors received no specific funding for this study.

Data availability The datasets analysed during the current study are available on the kaggle website. These datasets were derived from the following public domain resources: <https://www.kaggle.com/milanzdravkovic/pharma-sales-data-analysis-and-forecasting/data>

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Zadeh NK, Sepehri MM, Farvaresh H (2014) Intelligent sales prediction for pharmaceutical distribution companies: a data mining based approach. *Math Probl Eng* 2014:1–15. <https://doi.org/10.1155/2014/420310>

2. Hofmann E, Rutschmann E (2018) Big data analytics and demand forecasting in supply chains: a conceptual analysis. *Int J Logist Manag* 29:739–766. <https://doi.org/10.1108/ijlm-04-2017-0088>

3. Anirudh A (2020) Mathematical modeling and the transmission dynamics in predicting the Covid-19 - what next in combating the pandemic. *Infect Dis Model* 5:366–374. <https://doi.org/10.1016/j.idm.2020.06.002>

4. Alguliyev RM, Aliguliyev RM, Sukhostat LV (2020) Efficient algorithm for big data clustering on single machine. *CAAI Trans Intell Technol* 5(1):9–14. <https://doi.org/10.1049/trit.2019.0048>

5. Chen Z, Zhao B, Wang Y et al (2020) Multitask learning and GCN-based taxi demand prediction for a traffic road network. *Sensors* 20:3776. <https://doi.org/10.3390/s20133776>

6. Wang L, Zou H, Su J et al (2013) An ARIMA-ANN hybrid model for time series forecasting. *Syst Res Behav Sci* 30:244–259. <https://doi.org/10.1002/sres.2179>

7. Alpaslan F, Eğrioğlu E, AladağÇH TE (2012) An statistical research on feed forward neural networks for forecasting time series. *Am J Intell Syst* 2:21–25. <https://doi.org/10.5923/j.ajis.20120203.02>

8. Yu L, Wang S, Lai K (2005) A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Comput Oper Res* 32:2523–2541. <https://doi.org/10.1016/j.cor.2004.06.024>

9. Chen Y, Shen L, Li R et al (2020) Quantification of interfacial energies associated with membrane fouling in a membrane bioreactor by using BP and GRNN artificial neural networks. *J Colloid Interface Sci* 565:1–10. <https://doi.org/10.1016/j.jcis.2020.01.003>

10. Azar AT, Bastan M, Habibifar N, Hamid M (2019) Performance optimisation of a pharmaceutical production line by integrated simulation and data envelopment analysis. *Int J Simul Process Model* 14:360. <https://doi.org/10.1504/ijspm.2019.10025218>

11. Carrasco R, Vargas M, Soto I, Fuertes G, Alfaro M (2015) Copper metal price using chaotic time series forecasting. *IEEE Lat Am Trans* 13(6):1961–1965. <https://doi.org/10.1109/tla.2015.7164223>

12. Muralitharan K, Sakthivel R, Vishnuvarthan R (2018) Neural network based optimization approach for energy demand

- prediction in smart grid. *Neurocomputing* 273:199–208. <https://doi.org/10.1016/j.neucom.2017.08.017>
13. PereaRG PEC, Montesinos P, Díaz JAR (2019) Optimisation of water demand forecasting by artificial intelligence with short data sets. *Biosys Eng* 177:59–66. <https://doi.org/10.1016/j.biosyseng.2018.03.011>
 14. Barak S, Sadegh SS (2016) Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm. *Int J Electr Power Energy Syst* 82:92–104. <https://doi.org/10.1016/j.ijepes.2016.03.012>
 15. Candan G, Taskin M, Yazgan HR (2014) Demand forecasting in pharmaceutical industry using neuro-fuzzy approach. *J Mil Inf Sci* 2:41. <https://doi.org/10.17858/jmisci.06816>
 16. Bandara K, Shi P, Bergmeir C et al (2019) Sales demand forecast in e-commerce using a long short-term memory neural network methodology. *Neural Inf Process Lect Note Comput Sci*. https://doi.org/10.1007/978-3-030-36718-3_39
 17. Silva CFD, Almeida T, Barbosa RDM et al (2020) New trends in drug delivery systems for veterinary applications. *Pharm Nanotechnol*. <https://doi.org/10.2174/2211738508666200613214548>
 18. Tian C, Ma J, Zhang C, Zhan P (2018) A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies* 11:3493. <https://doi.org/10.3390/en11123493>
 19. Rahman M, Islam D, Mukti RJ, Saha I (2020) A deep learning approach based on convolutional LSTM for detecting diabetes. *Comput Biol Chem* 88:107329. <https://doi.org/10.1016/j.compbiolchem.2020.107329>
 20. Abbasimehr H, Shabani M, Yousefi M (2020) An optimized model using LSTM network for demand forecasting. *Comput Ind Eng* 143:106435. <https://doi.org/10.1016/j.cie.2020.106435>
 21. Lakshmanan B, Raja PSNV, Kalathiappan V (2020) Sales demand forecasting using LSTM network. *Adv Intell Syst Comput Artif Intell Evolut Comput Eng Syst*. https://doi.org/10.1007/978-981-15-0199-9_11
 22. Kuo R, Wu P, Wang C (2002) An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination. *Neural Netw* 15:909–925. [https://doi.org/10.1016/s0893-6080\(02\)00064-3](https://doi.org/10.1016/s0893-6080(02)00064-3)
 23. Aladag CH, Egrioglu E, Kadilar C (2012) Improvement in forecasting accuracy using the hybrid model of ARFIMA and feed forward neural network. *Am J Int Syst* 2:12–17. <https://doi.org/10.5923/j.ajis.20120202.02>
 24. Kumar M, Thenmozhi M (2014) Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *Int J Bank Acc Financ* 5:284. <https://doi.org/10.1504/ijbaaf.2014.064307>
 25. Takahashi Y, Aida H, Saito T (2000) ARIMA model's superiority over f-ARIMA model. In: WCC 2000 - ICCT 2000 international conference on communication technology proceedings (Cat No00EX420). <https://doi.org/10.1109/icct.2000.889171>
 26. Mohammed NA, Al-Bazi A (2021) An adaptive backpropagation algorithm for long-term electricity load forecasting. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-021-06384-x>
 27. Livieris IE, Pintelas E, Pintelas P (2020) A CNN–LSTM model for gold price time-series forecasting. *Neural Comput Appl* 32(23):17351–17360. <https://doi.org/10.1007/s00521-020-04867-x>
 28. Oliveira DD, Rampinelli M, Tozatto GZ, Andreão RV, Müller SM (2021) Forecasting vehicular traffic flow using MLP and LSTM. *Neural Comput Appl* 33(24):17245–17256. <https://doi.org/10.1007/s00521-021-06315-w>
 29. Hamzaçebi C, Es HA, Çakmak R (2017) Forecasting of Turkey's monthly electricity demand by seasonal artificial neural network. *Neural Comput Appl* 31(7):2217–2231. <https://doi.org/10.1007/s00521-017-3183-5>
 30. Kanavos A, Kounelis F, Iliadis L, Makris C (2021) Deep learning models for forecasting aviation demand time series. *Neural Comput Appl* 33(23):16329–16343. <https://doi.org/10.1007/s00521-021-06232-y>
 31. Hu Z, Zhao Q, Wang J (2018) The prediction model of worsted yarn quality based on CNN–GRNN neural network. *Neural Comput Appl* 31(9):4551–4562. <https://doi.org/10.1007/s00521-018-3723-7>
 32. Merkurjeva G, Valberga A, Smirnov A (2019) Demand forecasting in pharmaceutical supply chains: a case study. *Procedia Comput Sci* 149:3–10. <https://doi.org/10.1016/j.procs.2019.01.100>
 33. Thomson ME, Pollock AC, Önkal D, Gönül MS (2019) Combining forecasts: performance and coherence. *Int J Forecast* 35(2):474–484. <https://doi.org/10.1016/j.ijforecast.2018.10.006>
 34. Nikolopoulos K, Buxton S, Khammash M, Stern P (2016) Forecasting branded and generic pharmaceuticals. *Int J Forecast* 32(2):344–357. <https://doi.org/10.1016/j.ijforecast.2015.08.001>
 35. Siddiqui R, Azmat M, Ahmed S, Kummer S (2021) A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry. *Supply Chain Forum Int J*. <https://doi.org/10.1080/16258312.2021.1967081>
 36. Tayyab M, Zhou J, Zeng X, Adnan R (2016) Discharge forecasting by applying artificial neural networks at the Jinsha river basin, China. *Eur Sci J ESJ* 12(9):108. <https://doi.org/10.19044/esj.2016.v12n9p108>
 37. Sun Y, Lang M, Wang D, Liu L (2014) A PSO-GRNN model for railway freight volume prediction: empirical study from China. *J Ind Eng Manag*. <https://doi.org/10.3926/jiem.1007>
 38. Ustaoglu B, Cigizoglu HK, Karaca M (2008) Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteorol Appl* 15(4):431–445. <https://doi.org/10.1002/met.83>
 39. Cifuentes J, Marulanda G, Bello A, Reneses J (2020) Air temperature forecasting using machine learning techniques: a review. *Energies* 13(16):4215. <https://doi.org/10.3390/en13164215>
 40. Nayak SC, Misra BB (2018) Estimating stock closing indices using a GA-weighted condensed polynomial neural network. *Financ Innov*. <https://doi.org/10.1186/s40854-018-0104-2>
 41. Li L, Xu Y, Yan L et al (2020) A regional NWP tropospheric delay inversion method based on a general regression neural network model. *Sensors* 20:3167. <https://doi.org/10.3390/s20113167>
 42. Ai S, Chakravorty A, Rong C (2019) Household power demand prediction using evolutionary ensemble neural network pool with multiple network structures. *Sensors* 19:721. <https://doi.org/10.3390/s19030721>
 43. Ng WW, Xu S, Wang T et al (2020) Radial basis function neural network with localized stochastic-sensitive autoencoder for home-based activity recognition. *Sensors* 20:1479. <https://doi.org/10.3390/s20051479>
 44. Chao Z, Kim H-J (2019) Removal of computed tomography ring artifacts via radial basis function artificial neural networks. *Phys Med Biol* 64:235015. <https://doi.org/10.1088/1361-6560/ab5035>
 45. Yu Q, Hou Z, Bu X, Yu Q (2020) RBFNN-based data-driven predictive iterative learning control for nonaffine nonlinear systems. *IEEE Trans Neural Netw Learn Syst* 31:1170–1182. <https://doi.org/10.1109/tnnls.2019.2919441>
 46. Qiu R, Wang Y, Wang D et al (2020) Water temperature forecasting based on modified artificial neural network methods: two cases of the Yangtze River. *Sci Total Environ* 737:139729. <https://doi.org/10.1016/j.scitotenv.2020.139729>
 47. Foroughi M, Azghandi MHA, Kakhki S (2020) Bio-inspired, high, and fast adsorption of tetracycline from aqueous media using Fe₃O₄-g-CN@PEI-β-CD nanocomposite: modeling by response surface methodology (RSM), boosted regression tree

- (BRT), and general regression neural network (GRNN). *J Hazard Mater* 388:121769. <https://doi.org/10.1016/j.jhazmat.2019.121769>
48. Tkachenko R, Izonin I, Kryvinska N et al (2020) An approach towards increasing prediction accuracy for the recovery of missing IoT data based on the GRNN-SGTM ensemble. *Sensors* 20:2625. <https://doi.org/10.3390/s20092625>
49. Chu W-L, Lin C-J, Kao K-C (2019) Fault diagnosis of a rotor and ball-bearing system using DWT integrated with SVM, GRNN, and visual dot patterns. *Sensors* 19:4806. <https://doi.org/10.3390/s19214806>
50. Xie H, Li G, Zhao X, Li F (2020) Prediction of limb joint angles based on multi-source signals by GS-GRNN for exoskeleton wearer. *Sensors* 20:1104. <https://doi.org/10.3390/s20041104>
51. Baliarsingh SK, Vipsita S, Gandomi AH et al (2020) Analysis of high-dimensional genomic data using map reduce based probabilistic neural network. *Comput Method Program Biomed* 195:105625. <https://doi.org/10.1016/j.cmpb.2020.105625>
52. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
53. Li D, Lin C, Gao W et al (2020) Short-term rental forecast of urban public bicycle based on the HOSVD-LSTM model in smart city. *Sensors* 20:3072. <https://doi.org/10.3390/s20113072>
54. Khan ZA, Hussain T, Ullah A et al (2020) Towards efficient electricity forecasting in residential and commercial buildings: a novel hybrid CNN with a LSTM-AE based framework. *Sensors* 20:1399. <https://doi.org/10.3390/s20051399>
55. MilanzdravKovic (2020) Pharma sales data analysis and forecasting. Available: <https://www.kaggle.com/milanzdravkovic/pharma-sales-data-analysis-and-forecasting/data>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.