Predicting temperature of erbil city applying deep learning and neural network

Sardar M. R. K Al-Jumur[1], Shahab Wahhab Kareem[2], Raghad z.yousif[3]
[1,3]College of Science Department of Physic, Salahaddin University, Kurdistan Iraq
[2]Erbil Technical Engineering College, Erbil Polytechnic University, Kurdistan Iraq
[1]paitaxt technical institute Erbil, Kurdistan Iraq
[3]Department of IT, Catholic University in Erbil, Kurdistan Iraq
[2]College of Engineering and Computer Science, Lebanese French University, Kurdistan Iraq

Article Info

Article history:

Received Sep 20, 2020
Revised Mar 21, 2021
Accepted Mar 31, 2021

Keywords:

Artificial neural network
Deep learning
Prediction models
Weather

ABSTRACT

One of the most significant and daunting activities in today's world is temperature prediction. The meteorologists traditionally predict temperature via some statistical models aimed to forecast the fluctuations that might have happened to atmospheric parameters such as temperature and humidity. The main objective of this paper is to build an intelligent temperature prediction model of Erbil city in KRG/Iraq based on a historical dataset from 1992 to 2016 in each year there are twelve months' average temperature readings from (January to December). Hence to resolve this prediction problem an up-to-date deep learning neural network has been used, the network model is based on long short-term memory (LSTM) as an artificial recurrent neural network (RNN) architecture which employed to estimate the future average temperature. The implementing model uses the dataset from real-time 30 weather stations deployed in the area of the city. The prediction performance of the proposed recurrent neural network model has been compared with some state of art algorithms like Adeline neural network, Autoregressive neural network (NAR), and generalized regression neural network (GRNN). The results show that the proposed model based on deep learning gives minimum prediction error.

Corresponding Author:

Shahab Wahhab Kareem
Department of Information System Engineering
Erbil Polytechnic University
Email: shahab.kareem@epu.edu.iq

1. INTRODUCTION

The prediction of future behavior is important issues in the sciences and engineering, to need it in the areas of all of life, such as the prediction price, weather, and temperatures, most countries rely on its plans and development programs on the foundations and advanced scientific methods to design more effective plan [1]-[2]. Through the statistical analysis know the past and disadvantages and predict the future and needs according to the possibilities available and where it cannot reach an accurate prediction for the future without knowing the cons and past shortcomings. So States registering a series of economic and social changes [3]. Forecasting the weather has received the attention of many researchers from different investigation areas due to its influence on public social life [4]-[5]. The popular availability is widespread Weather monitoring data including information development computer technology in the past decade has caused several searches to investigate the pattern hidden in the huge data set of prediction weather [6]. Weather forecasting is one of the catalytic research problems with most potential applications varying from aeronautics to agriculture and tourism. Learning to predict the weather using a broad dataset of weather data and designing a robust weather forecast form that uses fundamental designs covered in the large weather data collection are among the challenges of weather forecasting [7].

Many useful applications of statistical modeling techniques, especially machine learning, for forecasting weather requirements have been published over the last decade [5], [6], [8]. Based on deep belief networks [4] (DBN), proceeding deep networks [5], energy-based forms should display developing support of deep learning as structural designs organized in-depth within the prior decade [9]. The word "deep" in deep learning refers to the fact that this neural network (NN) has more layers than those used in machine learning models. Multilayer NN has drawn a lot of research interest after a successful implementation of an unsupervised pre-training tool which is used to solve simple training problems. When it comes to simple forms, "deep" structure is very important, and NN has a deep structure that can produce a more effective learning experience.

The goal of the paper is to examine some feasibility of deep learning methods for weather prediction applying strong hierarchical weather descriptions learned from the weather data set. Models have tested using the amount of weather data provided by Erbil weather station (ministry of transport and communications) was compiled from about 30 weather stations from 1992 to 2016. In the aforementioned research, The long short-term memory (LSTM), adaline neural network, autoregressive neural network (NAR), and generalized regression neural network (GRNN) are some of the algorithms for climate modeling.

2. RELATED WORK

There is much research to handle the problem of predicting the weather in this section, some of them will be focused on. In [10] temperature forecasting system for past weather parameters has been proposed which is based on fuzzy time series. In [7] a conclusion has been made, which states that using a combination of ANN leads to enhance a learning process in weather predications models for different weather parameters. An ANN for forecasting precipitation based o statistical reduction has been proposed [11]. Whereas in [12] a neural network based on chaotic short-term oscillations for wind predicting utilizing LIDAR data has been proposed. A long rains prediction form utilizing combined fuzzy logic Wavelet with ANN has been presented in [13]. In [14] the deep learning has been employed in weather prediction. Thus a comparison has been made between the forecast achievement of conditional restricted boltzmann machine (CRBM) with recurrence neural network (RNN) and convolutional network (CN) models. In [15] New machine learning techniques and data processing are presented by the authors (random forest classification) for climate conditions forecasting. They concentrate on the challenges of today's rapid weather change, which renders traditional weather prediction approaches less accurate and time-consuming. To meet these challenges, strong climate forecasting techniques are needed. In [16] a new data-driven lightweight weather-predicting model has been constructed by investigating the performance of temporary forming methods of temporal convolutional networks (TCN) and (LSTM) and comparing it with the current standard methods of machine learning, like support vector regression (SVR), standard regression (SR), random forest (RF) also vector auto regression (VAR), vector error correction model (VECM) and autoregressive integrated moving average (ARIMA), are performed as analytical predicting methods. Moreover, the arbitrage of predicting expert (AFE) is performed as the dynamic ensemble method. They investigate the community LSTM and TCN models, although it is particular model in weather prediction utilizing ANN, using time-series data. The deep learning structure has several layers that add surface weather parameters to weather prediction over a period of time. The proposed deep learning networks including TCN and LSTM layers are tested in multi-input multi-output and multi-input single-output regressions. The challenge of investigating current trends with weather-predicting as a data source covers results over time and space. Particularly, they investigate the potential of the forecast through a heterogeneous discriminant method Predictive models trained in a deep neural network as a standard analytical principle for a collection of weather-related variables. They explain how the fundamental form can develop spatial incorporation that utilizes long-range spatial dependencies [17].

3. NEURAL NETWORK

Neural networks can be utilized to anticipate any procedure as long as adequate information is produced to precisely prepare the model. They can create designs and recognize interface/inclines that are too mind-boggling to be in any way watched through different procedures, in this manner taking into consideration the solid markers of new circumstances of intrigue. ANNs can be utilized to precisely foresee mechanical properties and other significant weld qualities without the requirement for ruinous testing, while exact models will give trust in the welds delivered. The number of hubs in ANNs and the association examples of the hubs can shift. The most broadly utilized association design is the three-layer backpropagation neural system [18]-[19]. The deceptive neurons are managing sections that are made in powerful interconnected layers that get the data (input factors) and spread it towards the yield layer. The data got is weighted by a synaptic weight that picks if the going with neuron will be begun by the approval work, for the most part as far as possible in non-direct issues. System learning happens as the piles are balanced alongside the layers, as per the relationship between the wellsprings of information and the ideal yields. One of the most fundamental models is the multilayer perceptron network (MLP), which is generally utilized in the estimation of non-direct confines that delineate complex relationships among self-controlling and subordinate factors in different applications [20]. Fundamentally, there are 3 distinct layers in a neural system:

a) Input Layer - The Input layer speaks with the outer condition that presents an example to the neural system. Its responsibility is to manage all the information sources as they were.
b) Hidden Layers - The shrouded layer is the assortment of neurons that has actuation work applied on it and it is a middle-of-the-road layer found between the information layer and the yield layer. Its responsibility is to process the information sources acquired by its past layer. So it is the layer that is mindful extricating the necessary highlights from the information.
c) Output Layer - The yield layer of the neural system gathers and transmits the data in like manner in the way it has been intended to give. The data generated at this output layer can be given as input to the first layer as feedback. Nonlinear autoregressive neural network (NAR) empowers the expectation of future estimations of a period arrangement, bolstered by its history foundation, by methods for a re-taking care of component, in which an anticipated worth may fill in as a contribution for new forecasts at further developed focuses in time [21]-[22]. Deep learning techniques have become generally identified since important machine learning mechanisms for regression and classification problems. The most advanced deep learning architectures for subsequent learning responsibilities is LSTM [23]. LSTM designed to use long-range data regions involving data from time-series, which presents the best encouragement and potential for the evolution regarding a suitable arrangement to nonlinear structural challenges. The LSTM model supplants every RNN system under the hidden layer that contains an LSTM block for making a long-term memory [24]. The LSTM block contains four interactive parts: which is an input gate, an output gate, a forgetting gate, and an internal unit. Input is controlled by input gate which passes an initiated data which the next state to the internal part state. The internal block learns the unit status of the previous time step using a self-replicating connection [25]. The forget gate retains and deletes information that passes the internal unit state, which allows the LSTM unit to reset the unit's memory or adaptively forget. Output controlled by output gate which enables a movement within the LSTM block outpu [26]-[27].

4. METHODOLOGY

In this study, four neural network system designs have been selected to be utilized as prediction models for vitality utilization. The principal model is non-straight autoregressive neural systems (NAR), which are to gauge tests confined in a one-dimensional time arrangement. A non-direct autoregressive system with exogenous data sources (NARX), which grow multidimensional time arrangement utilizing outside data has been selected to improve time arrangement forecast execution, and eventually, a traditional GRNN system is also proposed. Each system model has its advantages and expenses: NAR techniques are less complex than NARX. All things considered, the last model permits the utilization of extra data that may improve expectation precision. In a genuine circumstance, only one out of every odd structure has a similar administration framework nor a similar number of highlights: some lone register utilization while others handle more data, for example, outer and inside temperature. The chosen ANN models permit us to work with the two methodologies. The accompanying subsections portray these models to take care of the issue of vitality utilization time arrangement forecast.

4.1. NAR model

A nonlinear autoregressive neural system applied to time arrangement determining portrays a discrete, non-direct, The following are the characteristics of an autoregressive model.

$$y(t)=h(y(t-1), y(t-2), \dots, y(t-p))+e(t) \quad (1)$$

This equation portrays how a NAR arrangement is utilized to foresee the estimation of an information arrangement y at time t, y(t), utilizing the p past estimations of the arrangement. The capacity h(•) is obscure ahead of time, and the preparation of the neural system plans to estimate the capacity by methods for the enhancement of the system loads and neuron predisposition. At last, the term e(t) represents the blunder of the estimation of the arrangement y at time t. The topology of a NAR arrangement appears in Figure 1. The p highlights y(t−1), y(t−2), . . . , y(t−p), are called input delays.

4.2. ADALINE algorithm

The Adeline algorithm is used among tapped delay line (TDL) for constructing the linear filter which is applied within forecasting weather. Figure 3 shows the illustrated figure of Adaline. The first part of the system is the TDL. The data enters the input from the left and moves within N-1 delays. An N-dimensional vector is a TDL output that is made up of the current input data and the prior input signal [28].

4.3. Generalized regression neural network (GRNN)

GRNN is a radial neural base network variant that, unlike BPNN, has four layers: input, pattern, summation, and output [2]. The average temperature is used in the first layer as the input vector. The pattern units are in the second layer, and the outputs of this layer are fed into the third layer's summation units. The final layer protects the manufacturing units. Its architecture is depicted in Figure 4. GRNN replaces the sigmoid activation function typically used in ANN with a radial base function (RBF), which calculates the probability density function using an estimator [15]. The predicted value is simply a weighted average of target values from training patterns that are similar to the input pattern. The smoothing factor, which represents RBF distance, is the only adjustable parameter. The following are the GRNN specifics:

$$y(x) = \frac{\sum_{k=1}^{N} y_k K(x_1,x_k)}{\sum_{k=1}^{N} K(x_1,x_k)} \tag{2}$$

where,

$$K(x,x_k) = e^{-(x-x_k)^T(x-x_k)/2\sigma^2} \tag{3}$$

where y(x) is the input x prediction value, yk is the activation weight for the pattern layer neuron at k, and K(x, xk) is the RBF as formulated in (3). In the second layer, there are the same number of neurons as there are input/target vectors, but the weights are all set to target T. A column vector of 0.8326/Spread is used to set the bias. To simulate GRNN, use Matlab's built-in function:

net = newgrnn(P, T, Spread) (4)

The newgrnn function takes input vector P and target vector T matrices, as well as a radial base layer spread, and returns a network with weights and biases such that the outputs are exactly T while the inputs are P. The input has 22 inputs, with the first hidden layer consisting of 12 neurons and the second layer consisting of one neuron, and the output has one output.

4.4. LSTM

An LSTM is a multipurpose device that is made up of a recurrent network with feedback connections. Designate replies as training sequences of values shifting with a single time step to predict future values of future time steps for a series [29]. That is, the LSTM network learns to predict the value of the next time step for each time step in the input series. Training sequences without the final time stage are known as predictors [30]. Use the AndUpdatstate prediction function to predict multiple time phase values in the future one at a time, updating the network state in each forecast. Using the previous forecast as a starting point for each forecast [31].

4.5. Performance evaluation

The current dataset is collected from Erbil weather stations. In the LSTM the authors' steps as follow: Launch the data, that focuses on a single time series for step size referring to years and values denotes the number of hot days per month. The result is a cell array with a single time step for each unit. Make a row vector out of the results. After that, split the data into two groups: training and evaluation. Train for the first 85% of the series and then test for the last 15%. Thereafter, standardize the training data to provide a zero mean and unit variance for a better fit and to avoid the training from diverging. At the time of estimation, the test data must be standardized using the same parameters as the training data. Specify the responses to be training sequences with values changed by a one-time phase to predict the values of potential time steps of a series. That is, the LSTM network learns to predict the value of the next time step at each time step of the input series. The predictors are the training sequences that do not include the last time phase. After that, create an LSTM regression network. Set the secret units for the LSTM layer to 100. After that, choose your training choices. Set the solver to 'adam' and practice for 500 iterations. Set the gradient threshold to 1 to keep the gradients from bursting. Set the initial learning rate to 0.0054, and after 125 epochs, reduce the learning rate by a factor of 0.2. The train network is used to train the LSTM network with the defined training options. Use the ‘predictAndUpdateState’ feature to predict the values of several time steps in the future, one at a time, and update the network state at each prediction. Use the previous prediction as input to the function for each prediction. Using the same parameters as the training data, standardize the test data. Predict on the training data XTrain first to initialize the network condition. Then, using the last time phase of the training answer YTrain, make your first prediction (end). Input the previous prediction to ‘predictAndUpdateState’ after looping through the remaining predictions. For large collections of data, long sequences, or large networks, predictions on the GPU are usually faster to compute than predictions on the CPU. Otherwise, predictions on the CPU are usually faster to compute. For single-time step predictions, use the CPU. To use the CPU for prediction, set the 'ExecutionEnvironment' option of ‘predictAndUpdateState’ to 'cpu'. The autoregressive nonlinear adaptive filter model is the second proposed method. In this method, applied the feed-forward neural network with a tapped delay line. It used an adaptive filter to apply the desired response in the time series and the previous data as input to the adaptive filter. The one-step forecasting data is then output by the adaptive filter. The nonlinear prediction scheme was based on the neural network depicted in Figure 1. The most accurate forecast xe for the sequence of 25 values that define the hot degree values for 25 years can be obtained by starting a prediction method from given series data (Xn) at time n and comparing it to a time series. As a result, a predictor filter with an input vector lx is proposed, which is accomplished by applying the delay operator, Z-1, to the series xn. Following that, the filter output will generate xe, which will be equal to the current value xn. As a consequence, the error at time k in this prediction filter can be estimated:

$$e\ (k)=X_n(k)-X_e(k) \quad (5)$$

This is used to change the NN weights by the learning norm. The nonlinear filter coefficients are modified on-the-fly during the learning process using a criterion that changes the number of patterns, iterations, and duration of the tapped delay line at each pass of the time series. The NAR filter model is used to develop the proposed predict method in this paper. The neural network architecture consists of three secret (with 15, 8, and 4 neurons respectively) and one output neuron with lx inputs that reflect the number of input delays. The Levenberg-Marquardt method is the basis for the learning rule used in the procedure. The learning rule in this model, on the other hand, changes the number of patterns and iterations at each time level. The first delay is taken from the tapped-line xn and used as feedback to estimate the series xe one step ahead. As a result, the performance prediction can be written as:

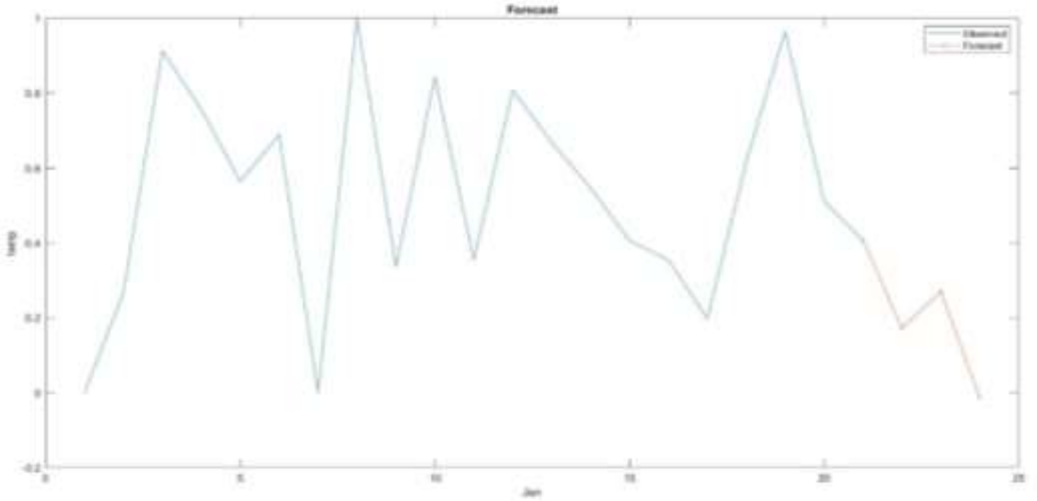

Figure 1. Training time series with the forecasted values

$$X_e(n+1)=F_p(Z\text{-}1I(\{X_n\}))\ (5)$$

xe(n+1) is the performance prediction at n+1, and Fp is the nonlinear predictor filter operator.

Figure 2 provides the performance of NAR based on the number of iterations. From the figure, it is clear that the performance of the NAR is stable when the number of iterations is increasing. From Figure 3 presents the RMSE of the NAR, it's clear that if the delay time till 5 delay time the RMSE is stable.

Figure 4 shows how Adaline was used to compare actual and predicted results. Weather value p(t), which enters from the left into a tapped delay line with the linear activation function in the output layer, is the signal to be predicted here. The previous two values of p (t) are available as outputs from the tapped delay line. The network uses to adapt the change the weights on each time step to minimize the error e (t) on the far right. If this error is 0, the network output a (t) is exactly equal to p (t), and the network has done its prediction properly. Given the autocorrelation function of the stationary random process p (t), you can calculate the error surface, the maximum learning rate, and the optimum values of the weights. After it is initialized and operated, the network adapts at each time step to minimize the error and in a relatively short time can predict the input p(t). Figure 5 present the RMSE of Adaline, Finally, the performance results of GRNN is presented in the Figure 6 which depicted the variation of performance with respect of spreading factor for different number of inputs (number of years). The best performance is attained using 16 years record with a spreading factor of 0.782. The difference between predicted data and real data has been depicted in Figure 7.

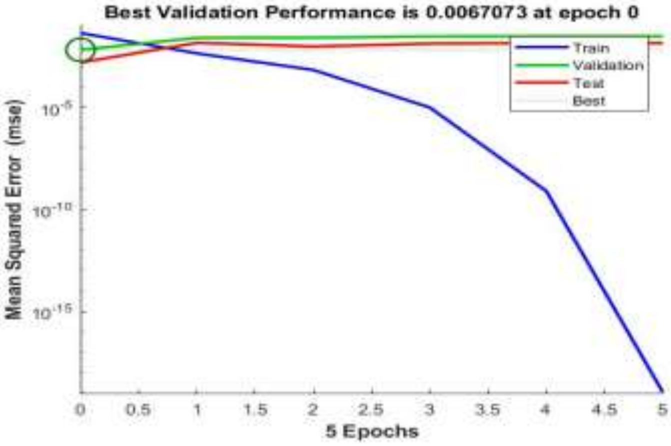

Figure 2. Performance of NAR

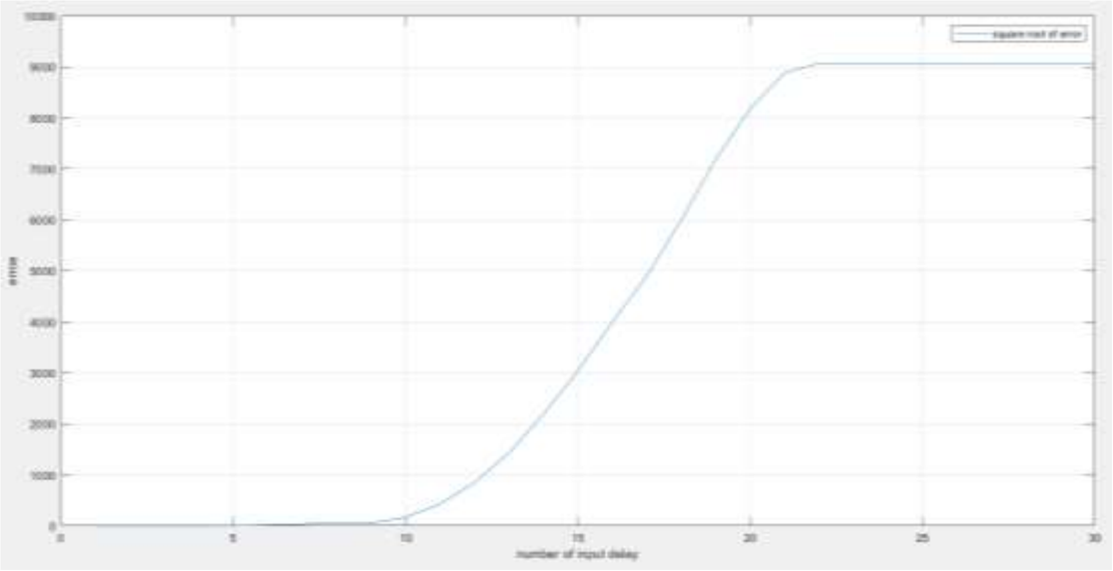

Figure 3. RMSE of NAR

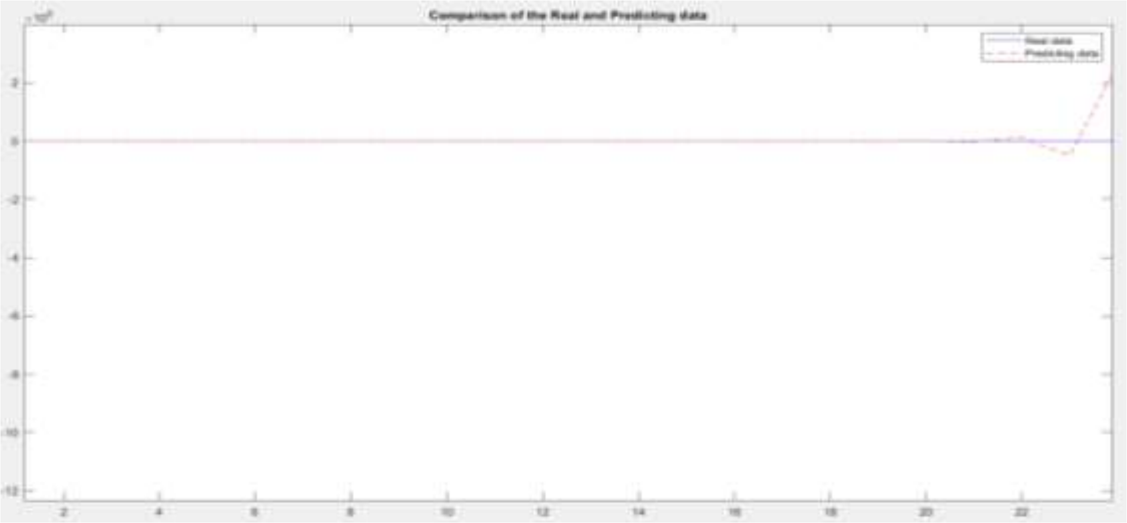

Figure 4. Comparison of the real and predicting data of adaline

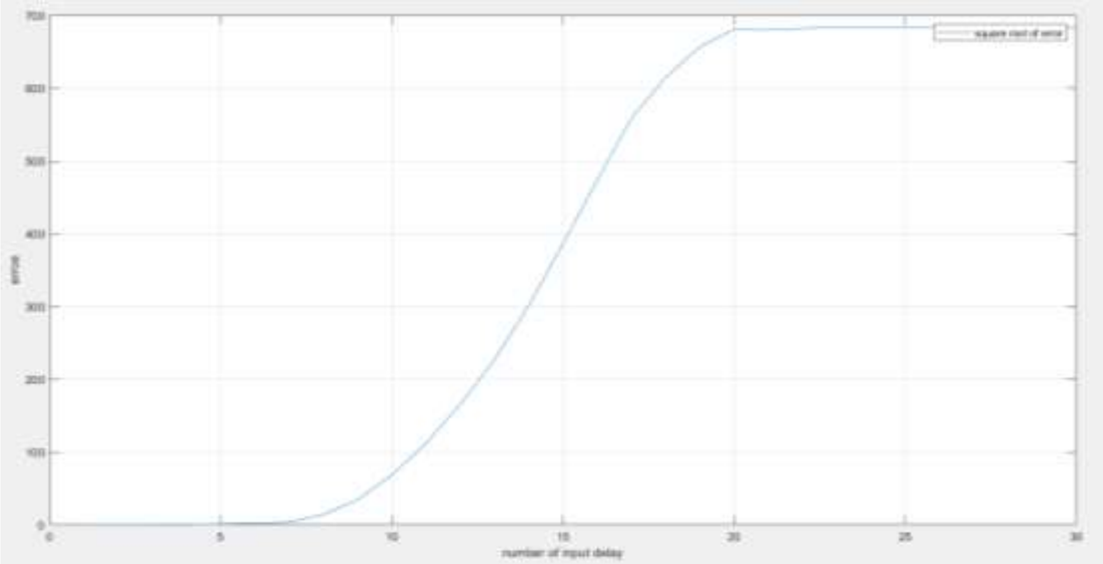

Figure 5. RMSE of adaline

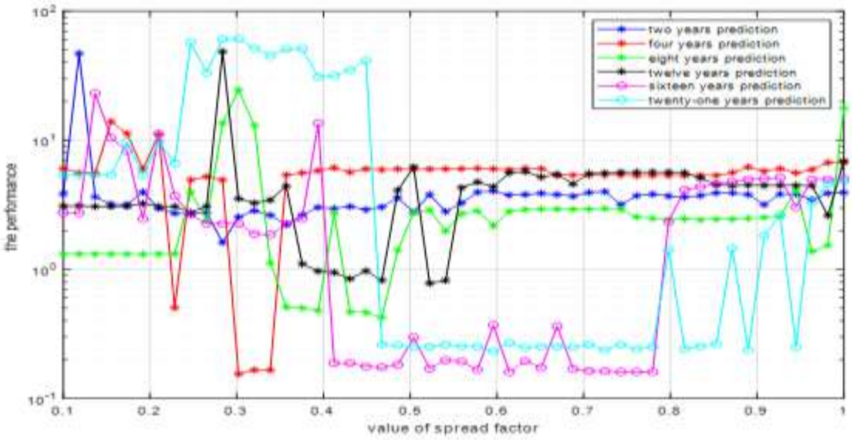

Figure 6. Performance of GRNN

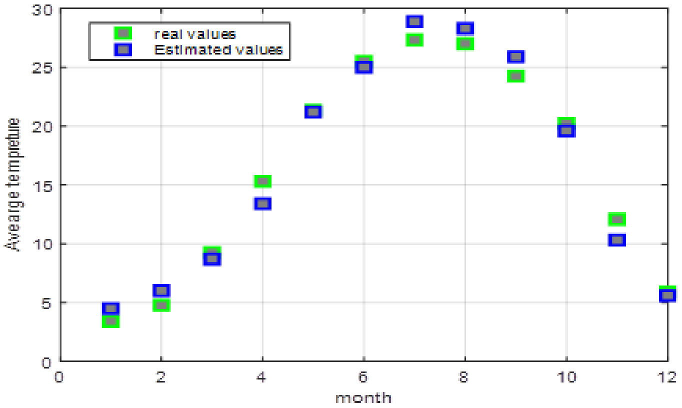

Figure 7. The comparison between real and predicted value

5. CONCLUSION

In this research work, four algorithms are presented. The LSTM model, NARX model, and Adaline neural network with GRNN for predicting the temperature of Erbil city. Various related methods gather data according to specific regions, whereas in this work data has been gathered from 30 different weather stations. The analysis of the data temperature through the neural network and a recurrent neural network model has been developed. In this paper, improved architectures on the neural network are presented with traditional ones (GRNN) such as like ADALINE, NAR neural networks, and LSTM recurrent neural networks, the purpose of which was to compare the effectiveness of various prediction methods. Obtained results illustrate that the proposed neural network algorithms optimize the error better than the statistical methods and the traditional one(GRNN) for the dataset for weather collected from (1992-2016) from KRG meteorological organization and seismology.

REFERENCES

[1] Shahab Wahhab Kareem and Mehmet Cudi Okur, “Evaluation of Bayesian Network Structure Learning Using Elephant Swarm Water Search Algorithm,” em *Handbook of Research on Advancements of Swarm Intelligence Algorithms for Solving Real-World Problems*, Chapter 8, IGI Global, pp. 139-159, 2020, doi: 10.4018/978-1-7998-3222-5.ch008.

[2] Nazmul Siddique and Hojjat Adeli, “Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks, and Evolutionary Computing,” First Edition, West Sussex, PO19 8SQ, United Kingdom: John Wiley & Sons, Ltd, 2013.

[3] Peter J. Brockwell Richard A. Davis, “Introduction to Time Series and Forecasting,” Second Edition, New York, Inc: Springer-Verlag., 2002.

[4] Shahab Wahhab Kareem and Mehmet Cudi Okur, “Pigeon Inspired Optimization of Bayesian Network Structure Learning and a Comparative Evaluation,” *Journal of Cognitive Science*, pp. 535-552, 4 20 2019, doi: 10.17791/jcs.2019.20.4.535.

[5] Shahab Wahhab Kareem, Mehmet Cudi Okur, “Structure Learning of Bayesian Networks Using Elephant Swarm Water Search Algorithm,” *International Journal of Swarm Intelligence Research*, pp. 19-30, 2 11 2020, doi: 10.4018/IJSIR.2020040102.

[6] Man Galih Salman, Bayu Kanigoro, Yaya Heryadi, “Weather Forecasting using Deep Learning Techniques,” em *ICACSIS 2015*, Indonesia, 2015, doi: 10.1109/ICACSIS.2015.7415154.

[7] Maqsood, I., M. R. Khan, and A. Abraham, “An ensemble of neural networks for weather forecasting,” *Neural Computing & Applications*, pp. 112-122, 2 13 2004, doi: 10.1007/s00521-004-0413-4.

[8] Shahab Wahhab Kareem, Raghad Zuhair Yousif, Shadan Mohammed Jihad Abdalwahid, "An approach for enhancing data confidentiality in Hadoop," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, pp. 1547-1555, 3 20 2020, doi: 10.11591/ijeecs.v20.i3.pp1547-1555.

[9] Ranzato, M., Y., Boureau, Chopra, S., and LeCun, Y, "A unified energy-based framework for unsupervised learning," em *AI-Stats*, 2009.

[10] Chen, S.M., and J.R. Hwang., “Temperature prediction using fuzzy time series,” em *Systems, Man, and Cybernetics, Part B: Cybernetics IEEE Transactions on*, vol. 30, no. 2, pp. 263-275, 2000, doi: 10.1109/3477.836375.

[11] N. P, “Artificial Forecasting in Statistical Downscaling," Thesis, Indonesia: Institut Pertanian Bogor, 2004.

[12] Kwong, K. M., Liu, J. N. K., Chan, P. W., and Lee, R, "Using LIDAR Doppler velocity data and chaotic oscillatory-based neural network for the forecast of the mesoscale wind field," em *2008 IEEE Congress on*

Evolutionary Computation (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, doi: 10.1109/CEC.2008.4631064.

[13] Belayneh, "Standard Precipitation Index Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Support Vector Regression," *Hindawi Publishing Corporation Applied Computational Intelligence and Soft Computing Volume*, 2012, doi: 10.1155/2012/794061.

[14] Man Galih Salman, Bayu Kanigoro, and Yaya Heryadi, "Weather Forecasting using Deep Learning Techniques," em *ICACSIS*, IEEE, 2015, doi: 10.1109/ICACSIS.2015.7415154.

[15] Nitin Singh, Saurabh Chaturvedi, Shamim Akhter, "Weather Forecasting Using Machine Learning Algorithm," em *Signal Processing and Communication (ICSC), International Conference on, IEEE*, India, 2019, doi: 10.1109/ICSC45622.2019.8938211.

[16] Pradeep Hewage, Marcello Trovati, Ella Pereira, and Ardhendu Behera, "Deep learning-based effective fine-grained weather forecasting model," *Pattern Analysis and Applications*, p. Springer, 2020, doi: 10.1007/s10044-020-00898-1.

[17] Aditya Grover, Ashish Kapoor, and Eric Horvitz, "A Deep Hybrid Model for Weather Forecasting," em *KDD'15, August 10-13*, Sydney, 2015, doi: 10.1145/2783258.2783275.

[18] S. Haykin, "Neural Networks: A Comprehensive Foundation," 2nd edition, Canada: Pearson Prentice Hall: Hamilton, 1999.

[19] Sami H. Ismael, Shahab Wahhab Kareem, and Firas H. Almukhtar, "Medical Image Classification Using Different Machine Learning Algorithms," *AL-Rafidain Journal of Computer Sciences and Mathematics*, pp. 135-147, 1 14 2020, doi: 10.33899/csmj.2020.164682.

[20] Amin Salih Mohammed, Shahab Wahhab Kareem, Ahmed khazal al azzawi, and M. Sivaram, "Time Series Prediction Using SRE- NAR and SRE- ADALINE," *Jour of Adv Research in Dynamical & Control Systems*, 12 10 2018.

[21] M. A. Rizkina, D. Adytia and N. Subasita, M. A. Rizkina, D. Adytia, and N. Subasita, "Nonlinear Autoregressive Neural Network Models for Sea Level Prediction, Study Case: in Semarang, Indonesia," 2019, em *7th International Conference on Information and Communication Technology (ICoICT), Kuala Lumpur*, Indonesia, 2019, doi: 10.1109/ICoICT.2019.8835307.

[22] S. Netsanet, J. Zhang, and D. Zheng, "Short term load forecasting using wavelet augmented non-linear autoregressive neural networks: A single customer level perspective," em *IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, Shanghai, 2018, doi: 10.1109/ICBDA.2018.8367717.

[23] Chowdhury, K. Nguyen, C. Fookes, and S. Sridharan, "A cascaded long short-term memory (LSTM) driven generic visual question answering (VQA)," em *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, pp 1842-1846, 2017, doi: 10.1109/ICIP.2017.8296600.

[24] Lu, Q. Zhang, Z. Yang, and M. Tu, "A hybrid model based on convolutional neural network and long short-term memory for short-term load forecasting," em *2019 IEEE Power & Energy Society General Meeting (PESGM)*, Atlanta, GA, USA, pp. 1-5, 2019, doi: 10.1109/PESGM40551.2019.8973549.

[25] Li, S. Dai, and Z. Cao, "Deep Long Short-term Memory (LSTM) Network with Sliding-window Approach in Urban Thermal Analysis," em *IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops)*, Changchun, China, 2019, doi: 10.1109/ICCChinaW.2019.8849965.

[26] K. Verma and V. Kuppili, "Data-oriented Neural Time Series with Long Short-Term Memories (LSTM) for Malaria Incidence Prediction in Goa, India," em *10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, 2019.

[27] Tongta and K. Chooruang, "Long Short-Term Memory (LSTM) Neural Networks Applied to Energy Disaggregation," em *8th International Electrical Engineering Congress (iEECON)*, Chiang Mai, Thailand, 2020, doi: 10.1109/ICCCNT45670.2019.8944749.

[28] Ghlib, Y. Messlem, and Z. Chedjara, "ADALINE-Based Speed Control For Induction Motor Drive," em *international Conference on Advanced Electrical Engineering (ICAEE)*, Algeria, 2019, doi: 10.1109/ICAEE47123.2019.9015162.

[29] G. N. Kouziokas, "Long Short-Term Memory (LSTM) Deep Neural Networks in Energy Appliances Prediction," em *Panhellenic Conference on Electronics & Telecommunications (PACET), Volos*, Greece, 2019, doi: 10.1109/PACET48583.2019.8956252.

[30] Qu Xiaoyun, Kang Xiaoning, Zhang Chao, Jiang Shuai, and Ma Xiuda, "Short-term prediction of wind power based on deep Long Short-Term Memory," em *IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, Xi'an, pp. 1148-1152, 2016, doi: 10.1109/APPEEC.2016.7779672.

[31] Li, H. Chen, T. Zhou, and X. Li, "Tailings Pond Risk Prediction Using Long Short-Term Memory Networks," *IEEE Access*, pp. 182527-182537, 7 2019, doi: 10.1109/ACCESS.2019.2959820.