A Comprehensive Study and Understanding—A Neurocomputing Prediction Techniques in Renewable Energies

Ghada S. Mohammed[1], Samaher Al-Janabi[2](✉), and Thekra Haider[1]

[1] Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq

[2] Department of Computer Science, Faculty of Science for Women (SCIW), University of Babylon, Hillah, Iraq

samaher@itnet.uobabylon.edu.iq

Abstract. Today's Renewable energy become the best solution to keep the environment from pollution and provide another source of generation energy. Data scientists are expected to be polyglots who understand math, code and can speak the language of generation energy from natural resources. This paper aims to display the main neurocomputing techniques for prediction in huge and complex renewable database to generation the energy form solar plant. Results clearly show that the LSTM improves the predictive accuracy, speed and cost of prediction. In addition, the results prove that LSTM can serve as a promising choice for current prediction techniques.

Keywords: Information Gain · LSTM · GRU · BLSTM · Alexnet · ZFNet · Renewable Energy

1 Introduction

As a result of the development in the world of technology and information, the digital revolution in different fields, this leads to a significant and noticeable increase in the need for energy, which has become an integral part of our lives. Most energy resources have a lot of limitations and drawbacks so shift towards the depended on the sources of renewable energy in solving problems of meeting the increasing demand for energy, reducing environmental impacts is considered one of the most critical challenge facing the world. Forecasting the amount of energy expected to be produced in the near future will help decision-makers to deal with the increasing demand for this energy and work to achieve a balance between energy production and consumption based on various forecasting techniques, but the prediction of the expected energy with high accuracy is considered a critical challenge, so our work aims to make a comparison among different Neurocomputing prediction techniques to find the most efficient techniques.

Intelligent Data Analysis (IDA) is one of the basic and influential tools in the process of decision making due to its importance in identifying new visions and ideas, it combines different strategies and techniques to collect data from multiple sources and use it

A. Abraham et al. (Eds.): HIS 2022, LNNS 647, pp. 152–165, 2023.
https://doi.org/10.1007/978-3-031-27409-1_14

to discover knowledge and interpret it to be accurate and understandable to all. The process of intelligent data analysis begins with defining the problem, determining its data, then defining and using techniques such as artificial intelligence techniques, pattern recognition, and statistics techniques to obtain the required results, and then evaluating, interpreting, and explaining these results and their impact on the decision-making process.

Renewable Energy sources are environmental energy (alternative energy) that reduces the harmful impacts on the environment; This concept of energy is linked to the energy that is obtained from natural sources that produce enormous amounts of energy and is capable of regenerating naturally; the resources of friendly environment energy are driven by the wind, hydropower, ocean waves, biomass from photosynthesis, and direct solar energy; friendly environment energy has many advantages where it considers non-polluting, sustainable, (one-time installation), economic, ubiquitous, safe and it offers a wide variety of options and also there are drawbacks to some of the sources (Maybe more costly or its effects by some environmental influences) [1].

There are many challenges that have curbed renewable energy and didn't help it to expand; the most important of these challenges is: The high-cost challenge compared to the cost of traditional systems for power generation and it constitutes an obstacle to the expansion of this energy. Also, the reliability of environmental and industrial matters that can affect the efficiency of the source used to generate renewable energy is considered an important challenge, It must be taken into account when preparing any feasibility study for a future system for generating renewable energy; the technical innovations, the efficiency development of the methods and techniques that used in renewable energy generation systems are considered an important challenges in this field, these methods have a significant impact on turning some of the challenges into strengths point such as increasing the accuracy will increase the efficiency and reduce the time. Another important challenge in the field of Renewable Energy (RE) is the manpower where the energy generation systems from environmentally friendly sources need more manpower to operate power plants compared to traditional energy generation systems that rely heavily on technology in operation. Our work will try to deal with these challenges from two sides programmable side and the application side to overcome these challenges.

Prediction techniques can be classified into techniques that are related to data mining such as (Random Forest Regression and Classification (RFRC), Boosted Tree Classifiers and Regression (BTCR), Chi-squared Automatic Interaction Detection (CHAID), Bayesian Neural Networks Classifier (BNNC), Decision Tree (DT), Exchange Chi-squared Automatic Interaction Detection (ECHAID), and Multivariate Adaptive Regression Splines (MARS)) [2], and techniques that are related to Neurocomputing such as (Convolutional Neural Network (CNN), Recurrent Neural Network(RNN), Gated Recurrent Units (GRU), and many others algorithms) [3].

AI is a wide area of the researches that refer to the ability of machines to simulate the human intelligence, the most it is popular branches are ML and DL techniques; The algorithms of ML are trained on a different variety of data, and can improve the algorithm accuracy with more data [4], ML is very popular for making the prediction process due to its high performance in dealing with the data heterogeneity (the data that come from different resources, numerous types and complex characteristics) more than the statistical methods, and handling complex prediction problems; Now with the development of DL techniques, these techniques considered extension of ML techniques and take advantage of the AI capabilities in the predict models) DL consisted of a large number of layers that were capable of learning the characteristics with an excellent level of abstraction [3], These Algorithms operate automatically with elimination of the need for manual operations [4].

2 Related Work

Many researchers have tried to developing prediction models based on deep learning techniques to solve the problem of the increasing demand and the urgent need for electrical energy due to the growing use of electronic devices. There are many different techniques that have been introduced to deal with this problem, through the review of previous works, it has been found that there are number limitation such as the time and the computation complexity and the accuracy problem as shown below.

The authors in [5] proposes an model combine (BLSTM) and the extended scope of wavelet transform to 24-h forecast the solar global horizontal irradiance for the Gujarat, Ahmedabad locations in India, to improve the forecasting accuracy, the input time series statistical features are extracted and decomposes the input into number of finite model functions then reduces it to trained the BLSTM networks. The author based on one year dataset to execution of the proposed model and using different matrices in the evaluation process; the model outperforms in compare with others models but there are some challenges with the design of this model such as: hyper parameters selection and the complexity of simulation time.

The authors in [6] proposed model for forecasting a wind speed based on using deep learning techniques (ConvGRU and 3D CNN)with variational Bayesian inference; historical information for two real-world case studies in the United States is used to apply the model. The results of the model performance evaluation show it outperforms other point forecast models (the Persistence model, Lasso Regression, artificial neural network, LSTM,CNN, GPR and Hidden Markov Model) due to the combination between techniques and using not too wide forecast intervals so the, model need to experiment to wider regions and using advance probabilistic methods to evaluate its performance.

The authors in [7] introduced a two-step wind power forecasting model, the first step is Variational Mode Decomposition (VMD), second step an improved residual-based deep Convolutional Neural Network (CNN). The used dataset was procured from a wind farm in Turkey. The results of the proposed method were compared with the results obtained from deep learning architectures (Squeeze Net, Google Net, ResNet-18, AlexNet, and VGG-16) as well as physical models based on available meteorological forecast data. The proposed method outperformed the other architectures and demonstrated promising results for very short-term wind power forecasting due to its competitive performance.

In [8] the authors Present a model to determine the strategy of real time dynamic energy management of Hybrid Energy Systems (HES) by using a deep reinforcement learning algorithm and training it on numerous data such as water demand, Wind Turbine(WT) output, photovoltaic(PV) output, electricity price, one year of load demand data to obtain optimal energy management policy, the theory of information entropy is used to compute the Weight Factor (WF) and determine the best between different targets. Simulation results of this study show the optimal policy for control and the cost reduce by up to 14.17%. But the model have many limitation in his structure.

The authors in [9] proposed a method for trade-off multi-objective (practical swarm optimization (MOPSO) algorithm and Techniques for order of preference by similarity to ideal solution (TOPSIS)) that used to achieve a strategy for energy management in system optimal configuration; also examining the strategy on the real-world case; The results show that the TPC /COE/ EC set in (grid-connected, off-grid scheme) each one is optimal in different configurations. The method evaluate based on different perspectives (energy, economic, and environmental).

3 Theoretical Background

3.1 Multi Variant Analysis

The high dimensionality of the dataset that use to build the predictor model is a very important issue because the high dimension of the dataset can include input features that are irrelevant to the target feature so that, this will increase the time complexity of the model, also, the process of training will be slow and the required system memory will be a large amount, all which will reduce the model performance and overall effectiveness; so must select the only important features that have an impact and it useful in prediction the target feature and removing the excessive and non-informative feature [10]; the feature selection technique contribute in cost reduction and performance(accuracy)improvement; in this work information gain, entropy and correlation methods used to perform feature selection.

Information Gain (IG) is a popular filter (entropy-based) technique proposed by Quinlantat, it can be applied to categorical features [11] it represents how much information would be gained by each attribute (The attribute with the highest information gain, is selected), the Entropy(H) is the average amount of information (change in uncertainty)that needed to identify the attribute [12], the interval of the Entropy is [0, 1]. The (IG) measure is biased toward the attributes with many outcomes(values).

$$H = -\sum_{i-1}^{D} p_i * \log_2 p_i \tag{1}$$

$$IG = H - \sum \frac{DO}{DS} * H_J \tag{2}$$

where DO, DS is the dataset and sub-dataset, H_J is entropy of sub-dataset.

3.2 Long Short-Term Memory Algorithm (LSTMA)

It is one of Recurrent Neural networks (RNNs) that demonstrated clear superiority [15], the default behaviour is remembering the context for long intervals so it is capable of facilitating detection the long term dependencies. In LSTM, the memory cell is operated instead of the activation function of the hidden state in the RNN; LSTM consists of cell with three gates (Memory cell, input, forget and output gates), The three gates regulate the preceding information (the flow of information to next step) while the cell used to remember the values (maintain the state) over different intervals [13]. Each gate has its special parameters that need to be trained. Also, there are hyper-parameters that need to be selected and optimized (hidden neurons number and batch size). Due to its impact on the performance of LSTM architecture [14, 15]; The LSTM architecture was presented by Hochreiter and Jürgen [16–18], there are many of modifications that performed on the classical LSTM architecture to decrease the design complexity and time complexity.

Algorithm 1: LSTM Algorithm

Input: Dataset.
Output: Predicted of target
Initialization the net parameters(#epochs , batch_size,input_size ,LR, # units)
//Control net complexity and input the data blocks to the allocated layers

1. **While** not meeting the max _epoch do
2. **for** each batch in max _iterations
3. **for** each x(t) in the batch
4. Compute(c(t) ,i(t) ,fg(t),cs(t),o(t),y(t))
   $c(t) = tanh(Wc(x(t), y(t-1)) + bc)$
   $i(t) = \sigma(Wi(x(t), y(t-1)) + bi)$
   $fg(t) = \sigma(Wf(x(t), y(t-1)) + bf)$
   cs(t)=f(t) ∗c(t-1) *(i(t)*c(t))
   $o(t) = \sigma(Wo(x(t), y(t-1)) + bo)$
   $y(t) = o(t) * tanh(c(t))$
5. **End for**
6. **End for**
7. **End while**
8. **End**

Where c(t) is the Memory cell, i(t) is the Input Gate, fg(t)is the Forget Gate, cs(t) is the New cell state, and o(t) is the output Gate; Wc, Wi, Wf and, Wo are the metrics of weights respectively; bc, bi, bf, andbo are represent the biases; x(t) is the input, σ is the logistic Sigmoid Function; for each batch needed W, y, b to be trained and updated input of the model.

4 Methodology

The proposed model consists from multi step as shown in Fig. 1.

4.1 Description of Dataset

In this work, there are two datasets (solar and weather dataset) each one of these datasets has different features; the solar dataset consists of 68778 samples while the weather dataset consists of 3182 samples. The solar panel dataset contains seven features

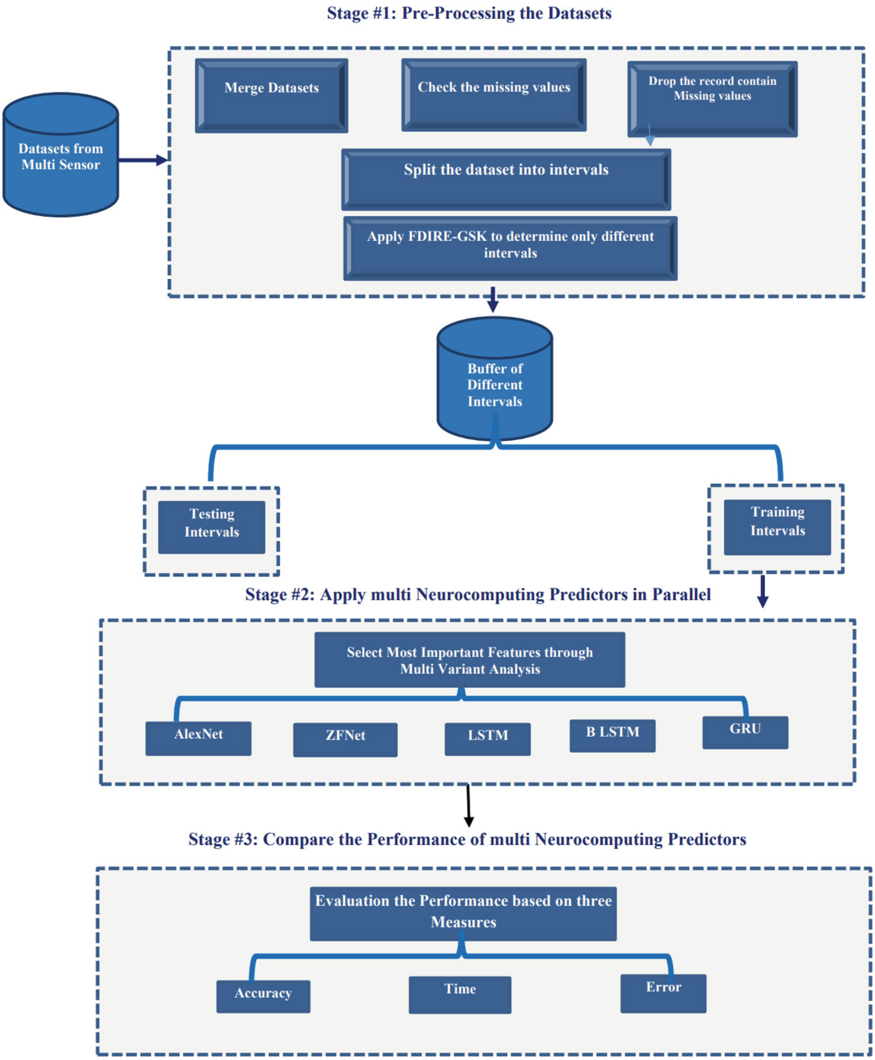

Fig. 1. The proposed model

(Date_time, plant_id, source_key, Dc_Power, Ac_Power, Daily_Yield, Total_Yield), whereas the weather dataset includes six features (Date_time, plant_id, source_key, Ambient_temperature, Module_temperature, irradiation).

4.2 Preprocessing the Data

This step involves handling the data sets, where:

1. The real time data capturing from multi sensors (solar plant sensor, weather sensor)
2. The datasets merging in one dataset based shared features (Source_key, and Date_time) this will causes to reduce the number of shared features and compressed the data in vertical manner.
3. Checking the new merged data set for the missing values, if there are any missing value the record will dropped this will causes to compressed the datasets in horizontal manner and this precise data compression will caused to reduce the time of computation.
4. Now the dataset will be cleaned, to increase the accuracy of predictor, the most important features in the dataset must determine, in our work using the, Information gain (that based on the computing the entropy) and the correlation methods to determine the importance of each features and its relation to the targets as shown in the Table 1.

Table 1. Information gain and correlation of the dataset features

Feature	Information Gain	Correlation
Dc_Power	1	1
Ac_Power	0.98746418	1
Daily_Yield	0.963139904	0.082
Total_Yield	0.996712596	0.039
Ambient_ Temperature	0.734895299	0.72
Module_ Temperature	0.734895299	0.95
Irradiation	0.734895299	0.99
Date	0.290856799	−0.037
Time	0.433488671	0.024
Hours	0.314651802	0.024
Minutes	0.110210969	0.0012
Minutes_ Pass	0.433488671	0.024
Date_Str	0.290856799	−0.037

From the Table 1 can show the Dc_Power feature has maximum information gain (1) and correlation to target feature (Dc_Power),also the Ac_Power has high correlation (1) to the target where the Date feature and data_str features have lowest correlation (−0.037). The Total_Yield features has highest information gain value () also the

Ac_Power has high information gain (0.98746418) where the Minutes feature has lowest information gain (0.110210969), these method determine the most related feature to target features and the most feature that have effect on the generation of Dc_Power

1. Now the datasets will contain the most important features only and based on the time and source key the data set will split into intervals and (each intervals for 15 min)
2. Based on the FDIRE-GSK algorithm, different intervals only will determine and saved in buffer to using them in the implementation of predictors; this determination to different intervals will increase the speed of performance of multi predictors model.
3. The data split into (Train_X (80%) of the data to train the model and Test_X(20%) to evaluate the model).

4.3 Built in Parallel Multi Predictor

In our work perform multi predictor in parallel in order to comparison between them and find the most accrues one, these predictors build based on some of Neurocomputing techniques (AlexNet, ZFNet, LSTM, BLSTM, GRU).

4.4 Performance Evaluation

The quality of the multi predictor model was evaluated using the error and accuracy measures. The results showed that a comparison between the predictors performance in terms of the error (mean square error) for each techniques and the accuracy. See algorithm 2 that show the main steps of the proposed model.

5 Results and Discussion

The aim of this work is predictions of maximum DC_Power that generated by solar plant, the prediction process based on different Neurocomputing techniques to compare between them and find the most efficient one., the merging process will reduce the number of processed features (reduce the number of columns from 13 feature to 11 features) then the process of cleaning the dataset from the missing values will reduce the number of processed samples (reduce the number of samples), this will increase the speed of predictor spatially the data collected in real time (Figs. 2, 3, 4 and 5).

```
Algorithm 2: Main Algorithm                                                    9
Input:        Data sets contain features capture from multi sensors
Output:       Prediction maximum DC power
   // Handel dataset based on FDIRE-GSK
   for each i in NRK                                   // k:# datasets
     for each j in NCK                  // NCK:#column
            Call Merging
     End for
   End for
   for  each i in NR
     for r each j in NC
     Checking the missing values
     Compute MIF                                       //MIF: Most important features
      End for
   End for
   for  each i in NR
     for each j in NMIF
      W[i]=stream o  data                              //split based on SOURCE _KE
                                                       &each 5 minutes
     End for
   End for
   for  each  l  in TNi                                //TNi: total number of intervals
     for r each j in NR
      Call FDIRE-GSK
     End for
   End for
   //Build in parallel multi predictor
   for each i in NR
     for each j in NC
     Split Data (Train_X, Train_Y, Test_X, Test_Y)
     End for
   End for
   for each i in Train_X
     for each j in Train_X
       Call (AlexNet , ZFNet , LSTM, BLSTM  ,GRU)
     End for
   End for
   //Evaluation the performance
    Compute the error
    Compute the accuracy
   End
```

The process of feature selection and determine the irrelevant feature in data set will effect on the accuracy of the predictor because it operate on most important features to the target, the selection based on the information gain value (that based on the entropy) and on the correlation between the datasets features and the target. In our work using the, Information gain (that based on the computing the entropy) and the correlation methods

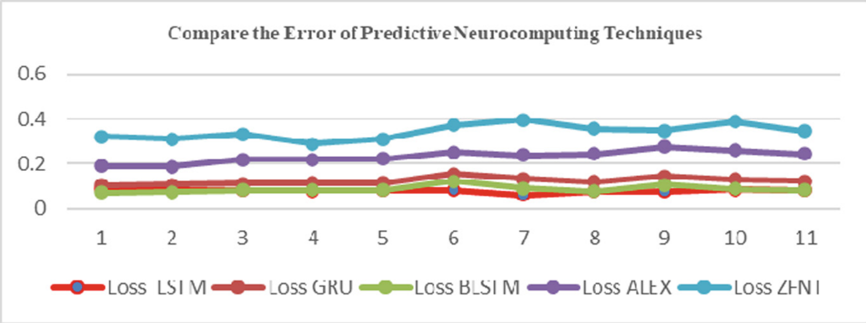

Fig. 2. Compare the Neurocomputing techniques based on Error

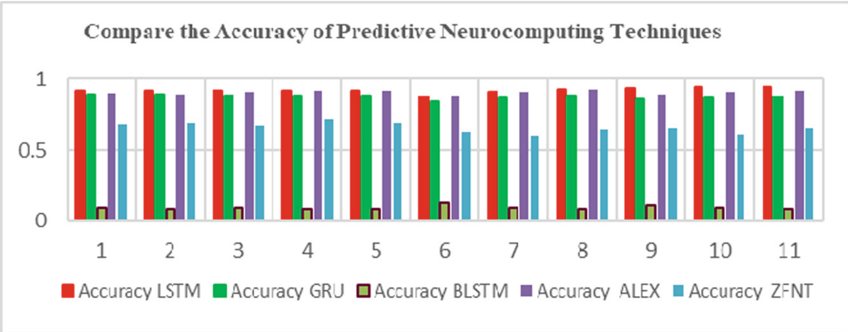

Fig. 3. Compare the Neurocomputing techniques based on Accuracy

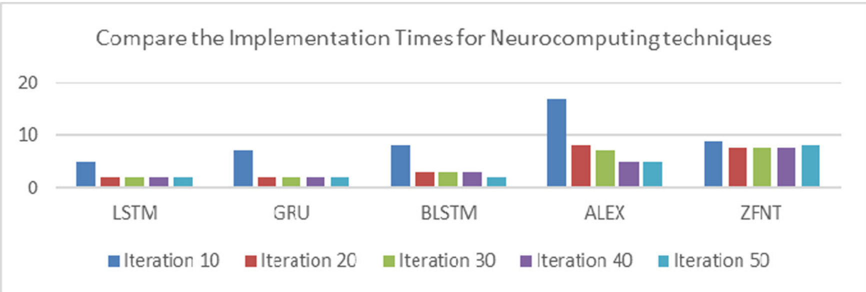

Fig. 4. Compare the Neurocomputing techniques based on Implementation Time

to determine the importance of each feature and its relation to the targets as shown in the Table 1. Building the multi predictor in parallel manner and compare between them based on the error of each predictor as shown in the Table 2, and the accuracy of the model as shown in the Table 3 while Table 4 shown the time required to execution each predictor. Finally; Total Time of each predictor shown in Table 5.

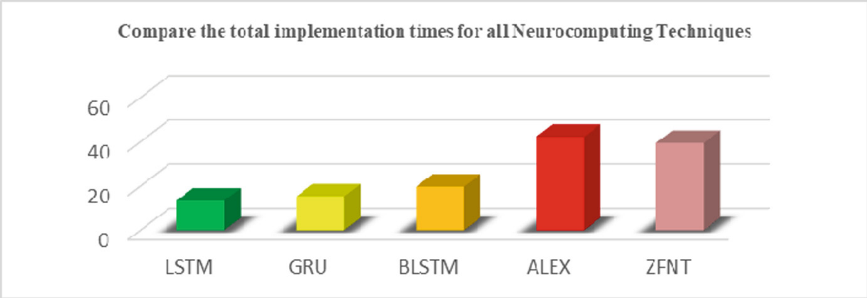

Fig. 5. Compare the Neurocomputing techniques based on Total Implementation Time

Table 2. Loss value of each predictor

LSTM	GRU	BLSTM	ALEX	ZFNT
0.091702	0.10561	0.073293	0.189727	0.320713
0.083836	0.10643	0.075757	0.188772	0.311013
0.081828	0.113829	0.0839	0.220824	0.33184
0.081664	0.114509	0.083522	0.220975	0.290288
0.081894	0.115322	0.083605	0.223171	0.312138
0.082543	0.156428	0.124139	0.251438	0.372194
0.062385	0.133367	0.092274	0.239495	0.395843
0.07846	0.119534	0.078328	0.243314	0.356411
0.076677	0.145451	0.107084	0.275609	0.349333
0.085221	0.130142	0.087893	0.260646	0.387477
0.083479	0.124513	0.083356	0.243735	0.346402

6 Conclusion and Future Works

This paper implemented neurocomputing techniques for predicting the DC_Power in renewable energy. In addition to that, it analyzed and compared some of the existing prediction neurocomputing techniques in an attempt to determine the main parameters that have the most important effects on their predictor. From the analysis, we found the techniques that are not dependent on randomization provided better results, while the ones using mathematical basis offered more powerful and faster solutions. In the light of this, mathematical basis is used in the proposed model.

The results show that the LSTM performs better than other prediction techniques in prediction in renewable domain. Also; it achieves an improvement in accuracy, speed of prediction and less cost. Therefore, the LSTM is promising choice compared to other prediction techniques. The experimental results also show that the LSTM employed in this work overcomes some of the shortcomings in other prediction techniques.

Table 3. Accuracy of each predictor

Accuracy LSTM	Accuracy GRU	Accuracy BLSTM	Accuracy ALEX	Accuracy ZFNT
0.918298	0.89439	0.094293	0.895707	0.679287
0.916164	0.89357	0.085757	0.89424	0.688987
0.918872	0.886171	0.0939	0.9061	0.66816
0.918936	0.885491	0.083522	0.916478	0.709712
0.919106	0.884678	0.083605	0.916395	0.687862
0.877957	0.843572	0.124139	0.875861	0.627806
0.907615	0.866633	0.092274	0.907726	0.604157
0.92154	0.880466	0.078328	0.921672	0.643589
0.933323	0.854549	0.107084	0.892916	0.650667
0.941779	0.869858	0.087893	0.912107	0.612523
0.946521	0.875487	0.083356	0.916644	0.653598

Table 4. The Time of each predictor model

Iteration	**LSTM** (s)	**GRU** (s)	**BLSTM** (s)	**ALEX** (s)	**ZFNT** (s)
10	5	7	8	17	9
20	2	2	3	8	7.5
30	2	2	3	7	7.5
40	2	2	3	5	7.5
50	2	2	2	5	8

Table 5. Total Time of each predictor model

Iteration	LSTM	GRU	BLSTM	ALEX	ZFNT
50	13.441	15.339	19.737	42.197	39.458

The results also show that some of predictors give very close results to each other such as (ALEX and ZFNT) while some of them are similar in both the work structure and the results such as (GRU and LSTM). As future work, we planning to develop LSTM by use optimization Algorithm (i.e., GSK). Using one of optimization algorithms such as swarm optimization, Ant Colony Optimization (ACO) and Genetic Algorithm (GA) to determine and select the most important features in order to reduce the time used in the predictor.

References

1. Al-Janabi, S., Alkaim, A.: A novel optimization algorithm (Lion-AYAD) to find optimal DNA protein synthesis. Egypt. Informatics J. **23**(2), 271–290 (2022). https://doi.org/10.1016/j.eij.2022.01.004
2. Al-Janabi, S., Alkaim, A.F., Adel, Z.: An Innovative synthesis of deep learning techniques (DCapsNet & DCOM) for generation electrical renewable energy from wind energy. Soft. Comput. **24**(14), 10943–10962 (2020). https://doi.org/10.1007/s00500-020-04905-9
3. Baydyk, T., Kussul, E., Wunsch II, D.C.: Intelligent Automation in Renewable Energy. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-02236-5
4. Al-Janabi, S., Mahdi, M.A.: Evaluation prediction techniques to achievement an optimal biomedical analysis. Int. J. Grid Util. Comput. **10**(5), 512–527 (2019)
5. Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., Sierra, J.E.: Urban traffic flow prediction techniques: a review. Sustain. Comput. Informatics Syst. **100739**,(2022). https://doi.org/10.1016/j.suscom.2022.100739
6. Sony, S., Dunphy, K., Sadhu, A., Capretz, M.: A systematic review of convolutional neural network-based structural condition assessment techniques. Eng. Struct. **226**, 111347 (2021). https://doi.org/10.1016/j.engstruct.2020.111347
7. Singla, P., Duhan, M., Saroha, S.: An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. Earth Sci. Inf. 1–16 (2021). https://doi.org/10.1007/s12145-021-00723-1
8. Liu, Y., et al.: Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model. Appl. Energy **260**, 114259 (2020). https://doi.org/10.1016/j.apenergy.2019.114259
9. Yildiz, C., Acikgoz, H., Korkmaz, D., Budak, U.: An improved residual-based convolutional neural network for very short-term wind power forecasting. Energy Convers. Manage. **228**, 113731 (2021). https://doi.org/10.1016/j.enconman.2020.113731
10. Zhang, G., et al.: Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach. Energy Convers. Manage. **227**, 113608 (2021). https://doi.org/10.1016/j.enconman.2020.113608
11. Zhao, P., Gou, F., Xu, W., Wang, J., Dai, Y.: Multi-objective optimization of a renewable power supply system with underwater compressed air energy storage for seawater reverse osmosis under two different operation schemes. Renew. Energy **181**, 71–90 (2022). https://doi.org/10.1016/j.renene.2021.09.041
12. Al-Janabi, S., Alkaim, A.F.: A nifty collaborative analysis to predicting a novel tool (DRFLLS) for missing values estimation. Soft. Comput. **24**(1), 555–569 (2019). https://doi.org/10.1007/s00500-019-03972-x
13. Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S.: A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. **53**(8), 5455–5516 (2020). https://doi.org/10.1007/s10462-020-09825-6
14. Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Asari, V.K., et al.: The history began from alexnet: a comprehensive survey on deep learning approaches (2018). arXiv:1803.01164. https://doi.org/10.48550/arXiv.1803.01164
15. Mirzaei, S., Kang, J.L., Chu, K.Y.: A comparative study on long short-term memory and gated recurrent unit neural networks in fault diagnosis for chemical processes using visualization. J. Taiwan Inst. Chem. Eng. **130**, 104028 (2022). https://doi.org/10.1016/j.jtice.2021.08.016
16. Nakisa, B., Rastgoo, M.N., Rakotonirainy, A., Maire, F., Chandran, V.: Long short term memory hyperparameter optimization for a neural network based emotion recognition framework. IEEE Access **6**, 49325–49338 (2018). https://doi.org/10.1109/ACCESS.2018.2868361

17. Darmawahyuni, A., Nurmaini, S., Caesarendra, W., Bhayyu, V., Rachmatullah, M.N.: Deep learning with a recurrent network structure in the sequence modeling of imbalanced data for ECG-rhythm classifier. Algorithms **12**(6), 118 (2019). https://doi.org/10.3390/a12060118
18. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)