Research on Nonlinear Systems Modeling Methods Based on Neural Networks

Ting Shi1,*, Wu Yang2 and Junfei Qiao3
1,2,3 Faculty of Information Technology, Beijing University of Technology, Beijing

*Corresponding author email: tingshi@bjut.edu.cn

Abstract. Nonlinear systems widely exist in all fields of industrial production and are difficult to model because of complex non-linearity. Neural network is widely used in process prediction, fault detection and fault diagnosis of modern industry because of the nonlinear fitting ability. Due to various structures, there exists diversity in the performance of neural networks. However, only the appropriate network can improve the efficiency and safety in modelling nonlinear industrial process, which requires full consideration of the structure of neural network. In this study, several typical structures of neural networks are compared and analysed, and the performance differences caused by these structures are presented in detail. Finally, performance differences of neural networks with inconsistent structures are verified on several experiments. The results showed that neural networks with inconsistent structures were good at dealing with different types of nonlinear systems. Our work will provide a theoretical basis in accurately modeling the industrial production process, which is beneficial to nonlinear system control.

Keywords: Nonlinear systems; System modeling; System control; Neural networks.

1. Introduction

Industrial production process normally has typical nonlinear and other complex characteristics [1]. It is greatly important to effectively model nonlinear system for process prediction and monitoring of industrial process [2]. However, the traditional multivariate statistical method and mechanism model method are difficult to accurately model the complex industrial process [3], which causes concerns of the efficiency and safety of industrial production.

In recent years, neural networks have been increasingly focused by scholars due to their excellent performance [4]. Neural network can approximate nonlinear system with any precision, so it has made many achievements in nonlinear system modelling [5]. However, the structure of neural network greatly affects the performance of system modelling [6]. Therefore, it is necessary to choose suitable neural network to ensure the precision and application of model. There are only three layers in a typical neural network, including input layer, hidden layer and output layer [7]. Number of input layer nodes are usually determined by the numbers of input features, the hidden layers of neural networks are related with structures of networks, and the design of output layer is related with the function of neural network. Considering the characteristics of middle layers, neural networks can be divided into three categories: shallow neural network, deep neural network and broad learning system network.

Shallow neural network has simple architecture with only one hidden layer [8]. Shallow neural network has been widely applied to process simple and critical dataset. For example, a shallow neural network architecture is presented to recognize hand drawn symbols from different writers [9]. Above shallow neural network had achieved satisfied performance in the corresponding tasks. This is because neural networks with shallow architecture can fully exact features of low dimensional data. However,
most of the data in reality is complex and high-dimensional, which cannot be processed by shallow neural networks [10]. Compared with shallow network, deep neural network has more than two hidden layers. With the ability of extracting features from high-dimensional and complex datasets, deep neural network has been used to solve many tough problems [11]. For example, based on neural network with deep structure, Lee et al. developed a method to evaluate heating energy consumption in old houses [12].

Based on deep neural networks, Aleksandri et al. designed a model to determine the age of author of the text [13]. These methods based on deep neural network can model complicated nonlinear system with high accuracy [14]. However, due to the depth of hidden layers of deep neural network, there exits problem of hyperparameters which makes deep neural network hard to train [15].

To solve the problem of hyperparameters, Chen et al. applied Broad Learning System (BLS) to adjust network by adding enhance nodes in the direction of width of hidden layer [16]. In this way, model based on BLS is trained and updated more efficiently in an incremental manner and achieves better generalization performance [17]. For example, Feng et al. applied BLS to control nonlinear dynamic system and proved the great control performance of BLS [18].

In the other hand, networks can be divided into full-connection network, circular network and partial-connection network according to the connection mode of nodes inside or between hidden layers. Nodes of adjacent hidden layers of full-connection neural network are connected one by one, which means the features of input data are extracted in maximum extent [19]. Multi-Layer Perceptron (MLP) is a commonly used full-connection neural network with flexible number of hidden layers [20]. With the characteristic of structure, the feature vectors are processed in each layer of MLP. So MLP is suitable for tasks of classification and regression prediction in nonlinear system. For example, Mohammad [21] designed a model based on MLP to identify and classify positive and active sonar target. However, due to the frame of full connection, MLP has the same problem of hyperparameter as deep neural network hyperparameter when excessive hidden layers or nodes exit in the neural network.

To solve the problem by changing mode of connection, neural networks with structures of sparse connection are proposed. Unlike fully connected neural network, the nodes between adjacent hidden layers of sparse connection networks are partially connected, which reduce number of parameters of network [22]. Therefore, the training efficiency of network has been greatly improved. For example, Echo State Network (ESN) replace hidden layer of MLP with reservoir pool. The nodes between input layer and reservoir pool are partially connected, and the weights are randomly determined. Moreover, the weights inside the reservoir are also obtained by random. In this way, it is only needed to adjust the connection weight between reservoir and the output layer, which greatly increases the training speed of the network. Adeleke [23] proposed ESN for network traffic prediction and the model achieves faster training speed and better accuracy than other methods. Another frequently used partial connection neural network is Convolutional Neural Network (CNN) with deep structure of convolution kernel and pooling layer. CNN has been widely used to extract deep features from large data sets such as images [24]. For example, Liu [25] designed a deep convolutional neural network for image smoke detection. The result indicated the model achieves higher accuracy while having fewer parameters.

Unfortunately, above networks are forward neural networks without recursive structure, so they are not good at dealing with sequential information. Recursive Neural Network (RNN) has a typical recursive structure with recursive cells of chain link to handle sequential datasets [26]. However, when learning a long sequence, problems of gradient vanishing and gradient explosion will appear in RNN [27]. To improve RNN performance, some scholars developed recursive neural network with gated memory unit named Long Short-Term Memory networks (LSTM). LSTM can control and store information of sequence with three control gates and solve the problem of long time dependence of RNN [28]. For example, by combining LSTM and Fully-Connected Neural Network (FCNN), Zhao [29] designed a model to forecast outdoor PM2.5 concentration in a period of time. Obviously, neural networks with different frame can achieve performance than others in specific tasks, and designing appropriate depth of neural network plays an important role in successfully modelling [30].
2. Detailed Analysis on Structures

The number of neurons in neural networks is determined by the requirements of system. Considering depth of hidden layers, neural networks can be divided into categories: shallow networks, deep networks, and broad learning networks. According to connection style of nodes between hidden layers, neural networks can be divided into categories: fully-connected network, partly-connected network, and recursively-connected network. Furthermore, there exists divergence between neural networks with different composition of hidden layer. The detailed difference of structures is discussed below.

2.1. Shallow and Fully Connected Networks

Neural networks with shallow and fully connected structures contain only three network layers, and neurons between adjacent layers connect with each other while there is no connection in the same layer. Therefore, shallow and fully connected networks are suitable for processing important features of datasets.

Auto-Encoder (AE) is a neural network commonly used in feature extraction. With three layers, AE can be regarded as an encoder and decoder structure [34]. As seen in Figure 1, AE was frequently applied to extract main characteristics of the datasets by reconstructing the input.

\begin{align*}
    h &= f_e(Wx + b) \\
    y &= f_d(W'y + p)
\end{align*}

where \( h \) and \( y \) are the outputs of encoder and decoder. \( f_e \) and \( f_d \) are activation functions of encoder and decoder, \( b \) and \( p \) are the biases of \( f_e \) and \( f_d \), \( W \) and \( W' \) are weights of encoder and decoder.

2.2. Deep and Fully Connected Networks

Deep neural network has more than one hidden layer, which has ability to model systems with complex nonlinear. Deep Belief Network (DBN) is a typical deep neural network with nodes between layers fully connected. And DBN is a probability generation model, which trains the network by maximizing the probability [36]. As presented in Figure 2, DBN is constructed by stacking many Restricted Boltzmann Machines (RBMs). The construction of DBN is unsupervised process, and the adjustment of parameter is a supervised process.

The related mathematical calculation is as following:

\begin{align*}
    E(v, h) &= -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n} \sum_{j=1}^{m} v_i w_{ij} h_j \\
    p(v, h) &= \frac{1}{Z} e^{-E(v, h)} \\
    Z &= \sum_{v, h} e^{-E(v, h)}
\end{align*}

where \( E(v, h) \) is the function of energy probability model, \( v \) and \( h \) represent the state of the visible layer and hidden layer of DBN, \( a_i \) and \( b_j \) are the biases of the visible layer and hidden layer respectively, and \( w_{ij} \) is the weight between visible and hidden layer. \( Z \) is the partition function, and
$p(v, h)$ is the joint probability distribution of $v$ and $h$. In process of constructing DBN, $p$ is needed to maximize.

2.3. Deep and Partly Connected Networks

Neural networks with deep and fully connected structure can achieve classification and regression of complex nonlinear systems. However, this structure also leads to problems such as numbers of parameters and heavy calculations. Partial connection can reduce calculate complexity of fully connected networks by sparse parameters.

CNN is a commonly used deep network with sparse connection, which contains many convolution layers and pooling layers as seen in Figure 7. Each convolution layer consists of several convolution kernels to obtain the important feature of input. Pooling layer can further extract features and reduce the dimension of input data.

![Figure 3. Convolution Neural Network.](image)

The concrete calculation of CNN is as following:

$$f_n = \sigma(f_{n-1} \cdot W_n + b_n)$$

$$v_n = pool(v_{n-1})$$

where $f_n$ and $f_{n-1}$ are respectively the output feature map and input feature map of nth convolution layer. $\sigma$, $W_n$ and $b_n$ are the activation function, shared weight and bias of nth convolution. $v_{n-1}$ and $v_n$ are the input and output vectors of nth pooling layer. $pool$ represents pooling operation, which is usually divided into average pooling and maximum pooling.

2.4. Deep and Recursively Connected Networks

With deep and recursive structure, neural networks can process sequential information such as language and time series. For example, RNN is applied to predict traffic in future period with history information. However, RNN has trouble in dealing with long sequence. LSTM can solve the problem of long-time dependence with three control gates. The hidden layer of LSTM consists of many memory units which can control long-term state information by three gates. As seen in Figure 4(b), the forget gate determines how much cell state of the previous remains until the current time, the input gate controls how much input under current time influences the cell state, and the output gate resolves how much cell state is taken as the current output.

![Figure 4. (a). The overall structure of LSTM; (b). The memory unit of LSTM.](image)

The mathematical formulas of LSTM are expressed as below:

$$f_i = \sigma(W_{hi} h_{t-1} + W_{xi} x_t + b_i)$$

$$i_t = \sigma(W_{i} h_{t-1} + W_{xi} x_t + b_i)$$

$$o_t = \sigma(W_{ho} h_{t-1} + W_{xo} x_t + b_o)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot o_t$$

$$h_t = o_t \cdot \tanh(c_t)$$
\[
\begin{align*}
\hat{c}_t &= \tanh (W_{fh} h_{t-1} + W_{fx} x_t + b_f) \\
c_t &= f_t * c_{t-1} + i_t * \hat{c}_t \\
o_t &= \sigma (W_{oh} h_t + W_{ox} x_t + b_o) \\
h_t &= o_t * \tanh (c_t)
\end{align*}
\]  

(10) \hspace{1cm} \text{(11)} \hspace{1cm} \text{(12)} \hspace{1cm} \text{(13)}

where \(W_{fh}, W_{fx}, W_{ox}, W_{oh}, W_{c}, \) and \(W_{oh}, W_{ox}\) are respectively weights between forget gate, input gate, cell state output gate, output gate and input data, output. \(b_f, b_i\) and \(b_o\) are biases of gate activation functions.

2.5. Broad Learning Networks

BLS solves the problems of hyper parameters and hard training of deep networks by constructing the network in width direction instead of adding more layers in depth direction, as shown in Figure 5. The added nodes in width direction are also called “enhancement nodes”. When new samples need to be trained, there is no necessary to readjust the whole network but use incremental learning by adding enhancement nodes.

The feature nodes \(F_i\) obtained by input layer mapping is represented as:

\[
F_i = \sigma_i(XW_{ei} + b_{ei}) \quad i = 1,2,...,n
\]

(14)

where \(\sigma_i\) is the mapping function, \(W_{ei}\) and \(b_{ei}\) are the weight and bias randomly generated of the function, and \(X\) is the input. The enhancement nodes \(E_j\) are described as following:

\[
E_j = \eta_j(F_jW_{ej} + b_{ej})
\]

(15)

where \(\eta_j\) is enhancement mapping function, \(W_{ej}\) and \(b_{ej}\) are the weight and basis of the function. The output of BLS is represented as:

\[
Y = [F_1,...,F_n][\eta(F_aW_{e1} + b_{e1}),...,\eta(F_aW_{en} + b_{en})]W^* = [F_1,...,F_n][E_1,...,E_m]W^* = [F_a][E_m]W^*
\]

(16)

3. Experiments and Results

Several neural networks with typical structures were experimented based on three datasets to compare performance difference. Metrics introduced to assess performance results were time, Mean Absolute Error (MAE), Normalized Absolute Error (NAE), Root Mean Square Error (RMSE), coefficient of determination(R2), accuracy and index of agreement (IA).

3.1. Datasets

In order to verify different performance of neural networks, three open datasets in deep learning with different characteristics and functions were selected in this study. The detail parameters of three datasets were shown in Table 1.

| Datasets                | Dimension | Targets       |
|-------------------------|-----------|---------------|
| Boston house price      | 506*14    | Regression    |
| PM2.5 of Beijing        | 43800*8   | Regression    |
| MNIST                   | 70000*28*28 | Classification |
The dataset of Boston house price contained 506 samples with house price and 13 influencers. Dataset of PM2.5 of Beijing includes hourly data of outdoor PM2.5 and other environment variables collected by Beijing Environmental Protection Testing Centre of China from 2010.1 to 20.14.12. MNIST contains 70000 pictures of handwritten digit from 0 to 9, based on neural networks, models are expected to recognize the digit through the input of picture. To make the results credible, the dataset was divided into two groups, one group was composed of 80% of the original data for training model, the other 20% was used to verify the model.

3.2. Models Based on Neural Networks
Four kinds of neural networks with totally different structure introduced to establish model were MLP, CNN, LSTM and BLS. Information of neural network structures and training method in experiment was seen in Table 2.

Table 2. Parameters of neural networks.

| Neural networks | Hidden layers | Connection way | Optimization |
|-----------------|---------------|----------------|--------------|
| MLP             | 1 layer with 100 nodes | Fully       | Adam         |
| LSTM            | 50 memory units     | Recursively  | Adam         |
| CNN             | 3conv2D 3pooling2D 1dense | Partially   | Adam         |
| BLS             | 1 layer with enhance nodes | Fully     | Adam         |

The second column of the Table.2 represents the information of hidden layers of neural networks on different datasets. Obviously, above four neural networks involve most structures of commonly used neural networks. All experiments in this study were conduct on a PC with python 3.7.0, and models based on neural networks were trained for 1000 epochs on each dataset. Additionally, Adam was used to reduce overfitting during training.

3.3. Results
Figure 6. showed that prediction value of MLP was closer to test data of Boston house price than that of other three models, while prediction value of CNN deviates from true value. This is because CNN with sparse connection structure has discarded the detailed information of low-dimensional data set, and neural networks with deep structure overfit the important and simple features. MLP with shallow and full connection structure fully extracted the information of Boston house price. The NAE, MAE and RMSE of MLP were smaller than other models seen in Table 3, R2 and IA of MLP are larger, indicating that MLP achieve better performance than other models on data set of Boston house price.

Figure 6. Test results of models on Boston house price.
With structure of convolution and pooling, CNN has faster training speed than other models on low-dimensional. Due the deep hidden layers, the training time of LSTM is longer than CNN with sparse connection and MLP with shallow structure. Specially, the R2 of CNN on Boston house price was
smaller than zero, which implied the model was not as effective as the mean model, indicating CNN cannot be used to model low-dimension dataset with simple characteristics.

As seen in Table 4, the NAE, MAE and RMSE indexes of LSTM on dataset of PM2.5 of Beijing were smaller than that of other three models, indexes of R2 and IA of LSTM were larger than that of other models. With deep and recursive layers, LSTM has advantage in process sequence information and achieves smaller error and better prediction performance than other three models. However, MLP with shallow structure is unable to fit the complex relationship between PM2.5 and other environment variables in dataset. So, model based on MLP made worst result in fitting complex and high-dimensional system.

Table 3. Result of Boston house price.

| Models | NAE   | MAE   | MAEP | RMSE  | R2    | IA    | Time(s) |
|--------|-------|-------|------|-------|-------|-------|---------|
| MLP    | 0.134 | 2.979 | 14.459% | 4.478 | 0.773 | 0.934 | 24.359  |
| LSTM   | 0.124 | 2.759 | 12.784% | 4.584 | 0.763 | 0.930 | 42.475  |
| CNN    | 0.363 | 7.273 | 32.339% | 10.970 | -0.357 | 0.558 | 3.932  |
| BLS    | 0.163 | 3.964 | 17.828% | 5.253 | 0.688 | 0.919 | 0.095  |

Table 4. Result of PM2.5 of Beijing.

| Models | NAE   | MAE   | RMSE  | R2    | IA    | Time(s) |
|--------|-------|-------|-------|-------|-------|---------|
| MLP    | 0.759 | 73.327 | 93.621 | -0.001 | 0.473 | 58.384  |
| LSTM   | 0.142 | 13.699 | 25.393 | 0.926 | 0.981 | 116.364 |
| CNN    | 0.522 | 50.414 | 78.499 | 0.296 | 0.621 | 33.759  |
| BLS    | 0.483 | 30.731 | 50.238 | 0.437 | 0.772 | 41.717  |

CNN discarded redundant pixel information and extract the key features of MNIST, so CNN achieved higher accuracy of classification than other models seen in Table 5. Due to the relatively simple characteristics and sufficient samples, models based on four neural networks in this study all made generally high accuracy. BLS with increment structure in width direction has much faster training speed than other models on the three datasets. Therefore, BLS has advantage in modelling systems required for quick responses such as fault monitoring and adaptive control.

Table 5. Result of MNIST.

| Models | Accuracy | Time(s) |
|--------|----------|---------|
| MLP    | 0.956    | 30.251  |
| LSTM   | 0.965    | 350.834 |
| CNN    | 0.993    | 250.104 |
| BLS    | 0.962    | 13.752  |

4. Discussion

(1) The structure of shallow network is simple and easy to train, which is suitable for dataset with simple characteristics or low dimension. Deep network has more network layers to model complex system, but it has more parameters and heavy computation burden, which is applicable to systems with complex characteristics and strong computing power. Broad learning system constructs the network in the direction of width, it has faster training speed due to less parameters and computation.

(2) The forward network is able to extract the relationship characteristics of different variables. The recursive network can keep the current information until next time, which has ability to deal with sequential datasets such as time series and natural language processing.

(3) Fully-connected network can extract features of date with high accuracy, but there is a risk of hyperparameters. Sparse connection networks have an advantage in processing large amount of data.
and locally related data with less parameter and computation, but it may lose some characteristics of data.

5. Conclusion
In this paper, several structures of neural networks were analysed and performance difference of the neural networks is given. Based on three datasets, performance differences of neural networks with inconsistent structures were verified. Shallow neural networks have simple structure, so that they are easy to train with less parameters. However, they are most likely to fall into the problem of local minimum value. Deep neural networks can model more complex system but there are problems such as hyperparameters and long training time. Therefore, optimized algorithms are needed to improve performance during network training. BLS converges quickly by constructing the network in the direction of width. Neural networks with recursive structure such as RNN, LSTM can process the sequential information of input data, which are able to predict and monitor time series systems. Sparse connection networks can solve the problem of hyperparameter of fully connected problem and speed up network training. The differences in the structure of neural networks mainly result from the depth of networks, the composition of hidden layers and the way of connection. It is of great significance to select the appropriate neural networks for the production efficiency and safety of industrial systems. In the future, more structure of neural network will be analysed and the specific methods on neural network selection will be given.

References
[1] Zhang Z, Zhu J, Liu, Y and Ge Z 2020. Industrial Process Modeling and Fault Detection with Recurrent Kalman Variational Autoencoder. 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS). IEEE.
[2] Pang H J, Geng Y S and Jiang X S 2018. Complex network model of process industry based on big data. Iop Conference, 382, 052004-.
[3] Li Z, Tong X, Man J, Ho W and Yim S 2020. A practical framework for predicting residential indoor PM2.5 concentration using land-use regression and machine learning methods. Chemosphere, 265(2), 129140.
[4] Li M, Yan Y, Wang Q, Du M and Wang C 2020. Secure prediction of neural network in the cloud. IEEE Network, pp (99), 1-7.
[5] Hu X, Li G, Niu P, Wang J and Zha L 2021. A generative adversarial neural network model for industrial boiler data repair. Applied Soft Computing, 104(4), 107214.
[6] Kaviani S and Sohn I 2021. Application of complex systems topologies in artificial neural networks optimization: an overview. Expert Systems with Applications.
[7] Lim H I 2020. A Study on Layers of Deep Neural Networks. 2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS).
[8] Zhou X H, Zhang M X, Xu Z G, Cai C Y and Zheng Y J 2019. Shallow and deep neural network training by water wave optimization. Swarm and Evolutionary Computation, 50, 100561.
[9] Dey S, Dutta A, Llados J, Fornes A and Pal U 2017. Shallow Neural Network Model for Hand-Drawn Symbol Recognition in Multi-Writer Scenario, 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE.
[10] Yang P, Zhou H, Zhu Y, Liu L and Zhang L 2020. Malware classification based on shallow neural network. Future Internet, 12(12), 219.
[11] Pan Y 2020. Network security and user abnormal behavior detection by using deep neural network. Internet Technology Letters.
[12] Lee S, Cho S, Kim S H, Kim J, Chae S and Jeong H 2020. Deep neural network approach for prediction of heating energy consumption in old houses. Energies, 14(1), 122.
[13] Romanov A S, Kurtukova A V, Sobolev A A, Shelupanov A A and Fedotova A M2020. Determining the age of the author of the text based on deep neural network models. Information (Switzerland), 11(12), 589.
[14] Eivazi H, Veisi H, Naderi M H and Esfahanian V 2020. Deep neural networks for nonlinear model order reduction of unsteady flows.

[15] Ding W, Tian Y, Han Sand and Yuan Y 2021. Greedy broad learning system. IEEE Access, pp (99), 1-1.

[16] Chen C and Liu Z 2018. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. IEEE Transactions on Neural Networks & Learning Systems, 29(99), 10-24.

[17] Liu Y, Wang Y, Chen L, Zhao J, Wang W and Liu Q 2020. Incremental Bayesian broad learning system and its industrial application. Artificial Intelligence Review, 1-21.

[18] Shuang F and Chen C 2018. Broad Learning System for Control of Nonlinear Dynamic Systems. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE.

[19] Li J, Li B, Xu J, Xiong R and Gao W 2018. Fully connected network-based intra prediction for image coding. IEEE Trans Image Process, 3236-3247.

[20] Wang L, Qin Y and Tao T 2019. Data modeling of calibration parameter measurement based on MLP model. 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI). IEEE.

[21] Khishe M and Safari A 2019. Classification of sonar targets using an MLP neural network trained by dragonfly algorithm. Wireless Personal Communications, 108(4), 2241-2260.

[22] Alford S, Robinett R, Milechin L and Kepner J 2018. Training behavior of sparse neural network topologies.

[23] Adeleke O A 2019. Echo-State Networks for Network Traffic Prediction. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE.

[24] Roy A, Anju P, Tomy L and Rajeswari M 2021. Recent Study on Image Denoising using Deep CNN Techniques. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).

[25] Liu M, K Gu, Wu L, X Xu and Qiao J 2019. Learn a Deep Convolutional Neural Network for Image Smoke Detection. Digital TV and Multimedia Communication. 2019.

[26] Xiao J and Zhou Z 2020. Research Progress of RNN Language Model. 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE.

[27] Tao L, Teng, W B and Zhao L A. Nonlinear unsteady bridge aerodynamics: reduced-order modeling based on deep LSTM networks-ScienceDirect. Journal of Wind Engineering and Industrial Aerodynamics, 198.

[28] Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena, 404.

[29] Zhao J, Deng F, Cai Y and Chen J 2019. Long short-term memory - fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere, 220(APR.), 486-492.

[30] Kim D E and Gofman M 2018. Comparison of shallow and deep neural networks for network intrusion detection. 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). IEEE.

[31] Yang Z Y and Jing, H 2017. A deep learning method based on hybrid auto-encoder model. IEEE Information Technology, Networking, Electronic and Automation Control Conference (pp.1100-1104). IEEE.

[32] X Chen, Liu J and X Sun 2019. Abnormal Gait Recognition Based on RBF Neural Network. 2019 Chinese Control and Decision Conference (CCDC).

[33] Yang Y 2020. Medical multimedia big data analysis modeling based on DBN algorithm. IEEE Access, pp (99), 1-1.

[34] Han H, Wu X, Liu Z and Qiao J 2019. Data-knowledge-based fuzzy neural network for nonlinear system identification. IEEE Transactions on Fuzzy Systems, pp (99), 1-1.