Echo State Network Optimization using Hybrid-Structure Based Gravitational Search Algorithm with Square Quadratic Programming for Time Series Prediction

Zohaib Ahmad
School of Electronics and Information Engineering,
Beijing University of Technology,
China
Ahmedzohaib03@gmail.com

Muhammad Qasim Memon
Department of Information and Computing,
University of Sufism and Modern Sciences,
Pakistan
memon_kasim@usms.edu.pk

Aasma Memon
School of Management and Economics,
Beijing University of Technology,
China
kaasma.bjut@gmail.com

Parveen Munshi
Faculty of Education,
University of Sufism and Modern Sciences,
Pakistan
vc@usms.edu.pk

Muhammad jaffar Memon
Civil Engineering Department, SZAB Campus,
Mehran University of Engineering and Technology,
Pakistan
jaffar.memon@muetkhp.edu.pk

Abstract: *The Echo-State Network (ESN) is a robust recurrent neural network and a generalized form of classical neural networks in time-series model designs. ESN inherits a simple approach for training and demonstrates the high computational capability to solve non-linear problems. However, input weights and the reservoir's internal weights are pre-defined when optimizing with only the output weight matrix. This paper proposes a Hybrid Gravitational Search Algorithm (HGSA) to compute ESN output weights. In Gravitational Search Algorithm (GSA), Square Quadratic Programming (SQP) is united as a local search strategy to raise the standard GSA algorithm's efficiency. Later, an HGSA-SQP and the validation data set to establish the relation configuration of the ESN output weights. Experimental results indicate that the proposed configuration of HGSA-SQP-ESN is more efficient than the other conventional models of ESN with the minimum generalization error.*

Keywords: *Echo state network, hybrid gravitational search algorithm, network configuration optimization, time series prediction.*

Received April 10, 2022; accepted April 28, 2022
https://doi.org/10.34028/iajit/19/3A/13

1. Introduction

In recent decades there has been enormous research work on Time Series Prediction (TSP) [1]. TSP builds realistic models using previously studied data to define future values. Time-series estimation is also a demonstration of forecasting the future by grasping past data [2]. Prediction is vital in the decision-making method in different domains, such as economics, energy, electromechanical systems, and various engineering fields [3]. It is also a non-linear dynamic problem, and the time series estimation usually has a detailed problem statement. In addition, a superior model of functional prediction competencies could establish a boost in predicting the time series data as reported elsewhere [4]. However, existing literature shows that computational methods enhance prediction competence. Surprisingly, there are no general criteria for solving the most suitable solution to a particular problem [5]-one of the crucial approaches in researching time series prediction using Artificial Neural Networks (ANNs). ANNs' key benefit is Computational simulation competence, a data-driven and a self-adapted process. ANNs are competent approximators [6], suggesting an extensive linear and Non-linear time series for different models. ANNs are reliable and widely used forecasting systems [2].

On the other hand, the same valuable approximation functionality enables ANN model time-series data that complicates the model specification. Therefore, designing a substantial architecture is essential in neural network applications [7]. Developing an internal network state of a system helps the model via Recurrent Neural Networks (RNNs), including Elman, Hopfield, and the Echo State Network (ESN), offering significant memory. RNNs have been implemented effectively in various fields, such as clustering [8], pattern recognition [9], classification [10], and prediction. Usually, such systems based on complex rational behavior have become more valuable insights than conventional Feed-Forward Neural Networks (FFNNs) to model a non-linear dynamic system.

ESN is a specific recursive configuration identified for its ease of construction and good accuracy in forecasting. Two principal characteristics differentiate ESN from other recursive networks: 1. the hidden ESN layer is a sparsely linked large-scale reservoir, and 2. it is necessary to train only the network's weights among the reservoir and the output. The ESN reservoir manipulates low-dimensional input to high-dimensional space via the feedback association concerning the reservoir and the output; thus, it is appropriate for managing dynamic system modeling challenges. Whereas ESN models have a high level of learning competence, they have not been extensively studied for their rich, random weight structure and dynamic features. Hence, designing an optimum reservoir for a particular purpose is a problem that still needs to be addressed. One of the best-known optimization strategies is hyperparameter optimization to address this problem. Intelligent approaches, including the fruit fly algorithm and particle swarm optimization, have been built to optimize important performance-enhancing ESN parameters [11].

This paper extends our previous work based on a method of prediction model using an Radial Basis Function (RBF) neural network. The proposed model predicts and identifies a non-linear system using a new hybrid approach of the Hybrid Gravitational Search Algorithm (HGSA) [7]. This paper introduces a hybrid architecture incorporating Gravitational Search Algorithm (GSA) with square quadratic programming Square Quadratic Programming (SQP) to address this drawback. At the beginning of the run, GSA has a greater capacity to explore a vast area, leaving agents free to fly and sit in several places. The optimum value of both agents will be chosen as the starting point for SQP and will be further optimized. Doing so opens the possibility of finding a global optimum for new local optimization issues. In contrast, our proposed study uses a hybrid GSA-SQP approach to adjust the ESN output weight.

The rest of this article is organized as follows. First, the original ESN is defined. Second, the standard GSA, hybrid HGSA-SQP algorithms, and HGSA-SQP-ESN are described. Experiments results are defined in Section 4. Finally, conclusions remarks are given.

2. Preliminaries

2.1. Echo State Networks

Figure 1 demonstrates the layout of ESN, which is presumed to have n nodes of input, N neurons, and one output unit without loss of generality. The inputs are designated using [u (t) = [$u_1(t)$, $u_2(t)$, .., $u_n(t)]^T \in R^n$, and allocated to the training patterns L, $[u(t), y(t)]^L_{t=1}$, where the outputs are indicated *[t]*.The echo states x(K) $\in R^n$ at a time t will be determined as:

$$x(t) = g(Wx(t-1) + W_{in}\,u(t)) \tag{1}$$

g (.) = $[g_1(.), g_2(.),..,g_N(.)]^T$ Represents the activation function of reservoir neurons, $W_{in} \in R^{N\times n}$ indicates the input weight, and $W \in R^{N\times N}$ represents the internal weight of the reservoir. After initialization, these matrices remain unchanged. The output of ESN at a time t is modified as:

$$y(t) = W_{out}(t)\,x(t) \tag{2}$$

Where $W_{out} = (W_1, W_2, \ldots, W_{N+n})^T \in R^{n+N}$ indicates the weight matrix of the output connection adjusted through the training phase. Let's consider that $X = [x(1), x(2), .., x(L)]^T$ it designates the internal state matrix and $T = [t(1), t(2), .., t(L)]^T$ represents the target output matrix. W_{out} is the output weight that can be evaluated by reducing the error as follows:

$$\tilde{E}(W_{out}) = \frac{1}{2} \| XW_{out} - T \|_2^2 \tag{3}$$

Where$\|.\|_2$ indicates the $\|_2$ norm. The solution W_{out} is generally determined by utilizing pseudoinverse [12]:

$$W_{out} = (X^T X)^{-1} X^T T \tag{4}$$

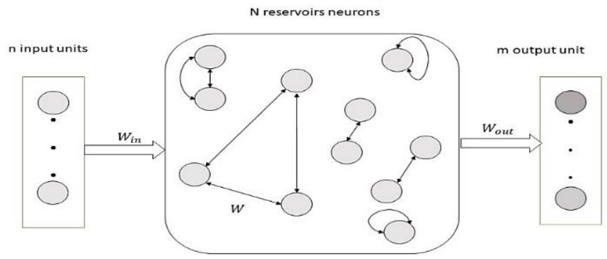

Figure 1. ESN structure.

2.2. Gravitational Search Algorithm

The gravitational search algorithm has its basis in the optimization techniques motivated supposed to be the objects, while to measure performance by the law of gravitation [13]. In this algorithm, agents are, masses are considered. Using the laws of motion and newton's gravity, the agents interact with one another. Assuming a solution comprises N agent's (masses). The position of the agent i can be calculated by:

$$X_i = (X_i^1, \ldots X_i^d \ldots X_i^n) \text{ for } i = 1, 2, \ldots N \tag{5}$$

Where X^i_d designates the agent i position in the dimension d. The fitness evolution is determined by computing the best and worse fitness for all agents at each iteration. For minimization problems:

$$best(t) = \min_{j \in \{1,\ldots N\}} fit_j(t) \tag{6}$$

$$worst(t) = \max_{j \in \{1,\ldots N\}} j \tag{7}$$

In the maximization problems,

$$best(\mathrm{t}) = \max_{j\in\{1,\ldots\mathrm{N}\}} fit_j(\mathrm{t}) \quad (8)$$

$$worst(\mathrm{t}) = \min_{j\in\{1,\ldots\mathrm{N}\}} fit_j(\mathrm{t}) \quad (9)$$

best (t) and *worst (t)* denotes the best and worst fitness value, *fit* $_j$(t) indicates the agent *j* fitness cost at iteration *t*. The constant gravitational *G* is determined at iteration *t*.

$$G(\mathrm{t}) = \mathrm{G}_0 e^{-\beta\frac{t}{T}} \quad (10)$$

To control the search precision, G_0 and β are initialized at the start and decreased over time, where T signifies the maximum iterations. The following equations determine the inertial masses of each agent at iteration:

$$M_{ai}(\mathrm{t}) = \mathrm{M}_{pi}(\mathrm{t}) = \mathrm{M}_{ii}(\mathrm{t}) = \mathrm{M}_i(\mathrm{t}) \quad (11)$$

$$m_{(\mathrm{t})} = \frac{fit_i(\mathrm{t}) - \mathrm{worst(t)}}{best(\mathrm{t}) - \mathrm{worst(t)}} \quad (12)$$

$$M_i(\mathrm{t}) = \frac{m_{(\mathrm{ii})}(\mathrm{t})}{\sum_{j=1}^{N} m_i(\mathrm{t})} \quad (13)$$

Where, M_{ii} (t) represents the inertial mass of the agent *i*. M_{ai} (t) and $M_{pi}t$ indicate the active and passive gravitational masses. The acceleration of the agent *i* at iteration *t* is calculated by:

$$a_i^d(\mathrm{t}) = \frac{F_i^d(\mathrm{t})}{M_{ii}(\mathrm{t})} \quad (14)$$

The gravitational force at iteration *t*, acting on the agent *i* is computed as:

$$F_i^d(\mathrm{t} \sum_{j\in Kbest, j\neq i} rand_j F_{ij}^d(\mathrm{t}) \quad (15)$$

K best Designates the number of first K agents with the most significant mass and the best fitness. *K best* fall linearly over time, and in the end, only one agent will exert force on the others. F^d_{ij} (t) is determined as follows:

$$F_{ij}^d(\mathrm{t}) = G(\mathrm{t})\frac{M_{pi}(\mathrm{t}) \times \mathrm{M}_{aj}(\mathrm{t})}{R_{ij}(\mathrm{t}) + \varepsilon}(\mathrm{x}_j^d(\mathrm{t}) - \mathrm{x}_i^d(\mathrm{t})) \quad (16)$$

F^d_{ij} (t) Indicates the force that acts among two agents *i* and *j* in dimension d at iteration *t*, while ε is a small constant. R_{ij} (t)Represents the Euclidean distance among two agents at iteration *t*, is expressed as:

$$R_{ij}(\mathrm{t}) = \| \mathrm{X}_i(\mathrm{t}), \mathrm{X}_j(\mathrm{t}) \|_2 \quad (17)$$

The agent's velocity at the following rate in dimension *d* equals the fraction of the current speed and acceleration.

$$v_i^d(\mathrm{t+1}) = \mathrm{rand}_i \times v_i^d(\mathrm{t}) + \mathrm{a}_i^d(\mathrm{t}) \quad (18)$$

In addition, the next position of the agent *i* of dimension *d* is determined according to the following equation:

$$x_i^d(\mathrm{t+1}) = \mathrm{x}_i^d(\mathrm{t}) + \mathrm{v}_i^d(t+1) \quad (19)$$

2.3. Hybrid GSA-SQP

GSA often has a weakness as a probabilistic multi-point search tool that converges into values that might not be optimal. Moreover, the GSA is looking for an excellent convergence solution that meets the optimum global level. The SQP technique appears to be the most robust Non-Linear Programming (NLP) approach for constrained optimization challenges [14]. For example, NLP techniques, SQP, have the dilemma of being stuck at an optimal local point if the initial selection is close to the optimal focal point. The NLP strategy provides an optimal global solution if the initial selection is correct [14]. In subsequent iterations, GSA implements the SQP workflow to increase GSA convergence as a local search method. GSA method is generally performed initially, and for each iteration, the optimal fitness for each generation is selected. The following optimal wellness is determined as the initial SQP variable values. The SQP method is then utilized to assess the probability of a local search, resulting in an increase in the optimal fitness of GSA in the current (10) . Hybrid GSA-SQP thus offers an optim l solution. The proposed approach is extended and implemented on 2.5-GHz i7 PCs with 16 GB of RAM in the MATLAB 2014b computing environment. SQP simulations are conducted as a local search framework using the MATLAB optimization toolbox.

2.4. HGSA-SQP-ESN

Each agent in HGSA-SQP represents a possible value of the output weights W_{out} matrix. The Root Means Square Error (RMSE) [15] is implemented to compute each agent's effectiveness:

$$RMSE = \sqrt{\sum_{i=1}^{N} y(\mathrm{i}) - \mathrm{y}_o(\mathrm{i}))^2} \quad (20)$$

Where y(i) and yo*(i)* represent the predicted and real output, and N represents the training sample sizes. The HGSA-SQP-ESN design procedure is described as follows:

1. Initialize the HGSA-SQP algorithm parameters containing go, β, the number of searching agents, and the maximum allowable iterations.
2. Randomly create the input weight Win and the reservoir weight W.
3. Run the reservoir through input signals as displayed in Equation (1), select internal states to gain matrix X.
4. Determine each search agent's objective function as seen in equation (12). This article employs the output connection of ESN as the agent's position, and root means square error as the objective function between each predicted and target value.
5. Update best(t), Worst(t) using equations (8) and (9)

6. Determine the gravitational coefficient G *(t)* according to equation (4).
7. Compute each agent's force by formula (14); calculate each agent's acceleration rate according to expression (15).
8. Update the agent position and velocity using Equations (16) and (17).
9. Judge if the current number of iterations approaches T max. If so, the algorithm ends to generate the best connection weight vector. Otherwise, orders t=t+1, return to Step 4, and enters step (11). t=t+1, return to Step 4, and enters step (11).
10. Consider the optimum connection weight gained by GSA as the initial point for SQP.
11. Execute SQP to search for the optimal connection weight that GSA finds.
12. Add the best connection weight to the ESN and assess the optimized ESN competency.

3. Simulations

In this portion, specific experiments are carried out to determine the effectiveness of HGSA-SQP-ESN, containing two benchmarks. In the case of benchmark optimization problems, the performance of the suggested HGSA-SQP-ESN is contrasted with specific existing approaches. The proposed model parameters utilized in the investigation are adjusted as α=10 and Go=100; The Maximum Iteration Numbers (TMAX), and the population size is adjusted to 1000 and 20, respectively. The probability of local search $\propto_{LS}$ by SQP is adjusted as below:

$$\propto_{LS}=1 \quad \text{for } t \le \frac{t_{\max}}{2}, \quad \propto_{LS}=0.95 \quad \text{for } t \rangle \frac{t_{\max}}{2} \tag{21}$$

Where t represents the current iteration.

3.1. Mackay Glass Time Series Prediction

This time series example is based on the differential structure of a time delay as follows [16].

$$\frac{dx(\mathrm{t})}{dt}=\frac{\propto x(\mathrm{t}-\tau)}{1+x^{c}(\mathrm{t}-\tau)}\beta x(\mathrm{t}) \tag{22}$$

The parameters are adjusted as follows: α=0.2, β=0.1 and c=10. Runge-second order Kutta is used to generate 0.1 step-size data sets. There are 2500 samples used for this experiment, 2000 samples are used for the training process, and 500 samples are treated for the testing phase. Since we prefer to get the test output signal close to the target one in supervised learning methods, we have chased these two signals after implementing the proposed strategy. Figure 2 Depicts the training RMSE of HGSA-SQP-ESN. Figure 3 Represents the overlap between the network and target output. They represent a response signal that imitates the desired movement from the network. The error signal of test data is shown in Figure 4 the test results can be seen in Table 1 to illustrate how the testing increased the network's efficiency and the target's output errors.

In addition, the potential of the proposed technique is equated to the other current approaches. Different approaches have been used for this benchmark problem, such as ESN and GSA-ESN. The results illustrate the excellent efficiency of our algorithm against some traditional designs.

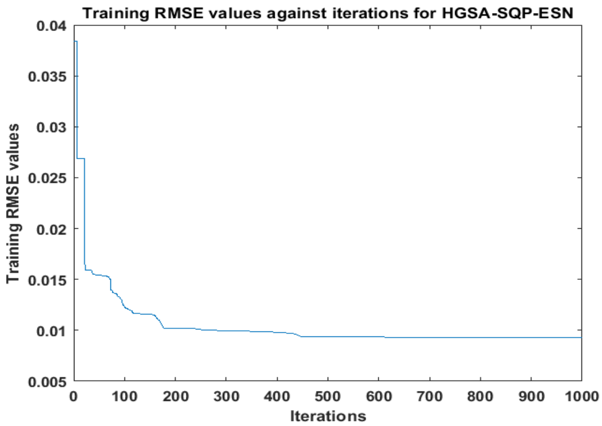

Figure 2. Training RMSE values against iterations.

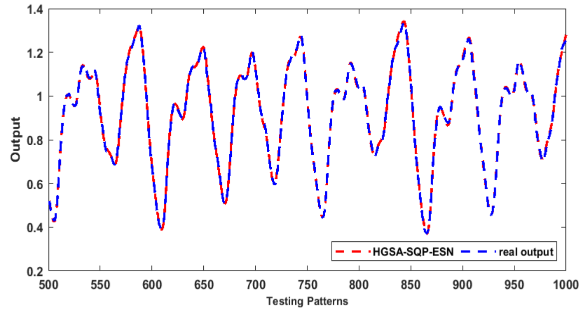

Figure 3. Network outputs vs. target outputs.

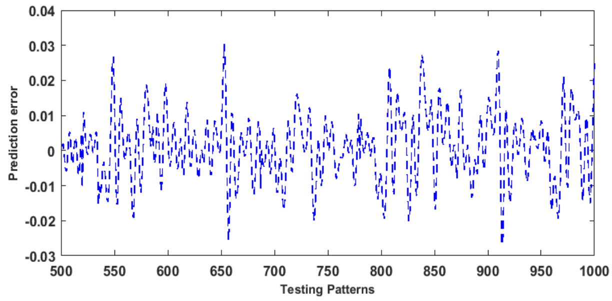

Figure 4. Evolution of the prediction error.

Table 1. Performance of various algorithms (Mackey-Glass).

Methods	RMSE (Training)	RMSE (Testing)
ESN	0.024	0.0208
GSA-ESN	0.018	0.0106
HGSA-SQP-ESN	0.0093	0.0084

3.2. Lorenz Time Series Prediction

Lorenz is a series of differential equations describing the fluid movement between a hot surface and a cold surface, and the following non-linear equations characterize it [17].

$$\begin{aligned}
&\frac{dx}{di} a(-x+y)\\
&\frac{dy}{di} = bx - y - xz\\
&\frac{dz}{di} = xy - cz
\end{aligned} \tag{23}$$

The parameters of the model are adjusted as a= 10, b= 28, and c= 8/3. The kutta-runge method with phase 0.01 generates values for the Lorenz time sequence. y(k_3), y(k_2) and y(k_1) are used to predict *y(k)* for each training pair. The initial size for the reservoir is set at 300. Two thousand five hundred samples have been produced in this analysis, 2 000 samples have been used as the training data set, and the remaining samples have been used as the test data set. The network outputs and their corresponding target outputs have been identified to demonstrate the strength of HGSA-SQP-ESN and show how they are related to the pattern of the Lorenz series. Figure 5 depicts the training RMSE of HGSA-SQP-ESN. Both signals are shown in Figure 6. From the first pattern to the last, we attempted to track the RMSE. The signal diagram for RMSE is shown in Figure 7. RMSE is contrasted with various conventional methods related to the Lorenz attractor, as in the previous benchmarks. According to Table 2, HGSA-SQP-ESN is an outstanding comparative model for the precision result.

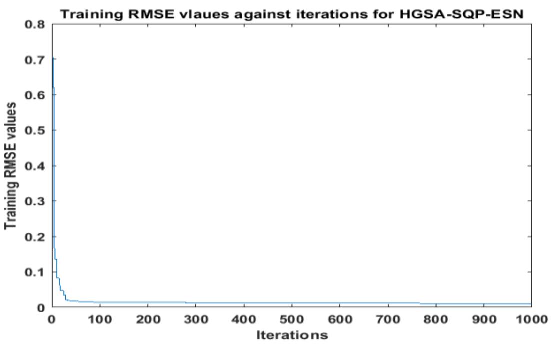

Figure 5. Training RMSE values against iterations.

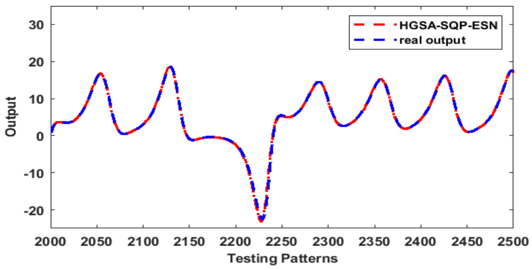

Figure 6. Network outputs vs. target outputs in the testing phase Lorenz.

Table 2. Performance of various algorithms (Lorenz)

Methods	Training Phase	Testing Phase
	RMSE	
ESN	0.056	0.012
GSA-ESN	0.018	0.012
HGSA-SQP-ESN	0.0107	0.0145

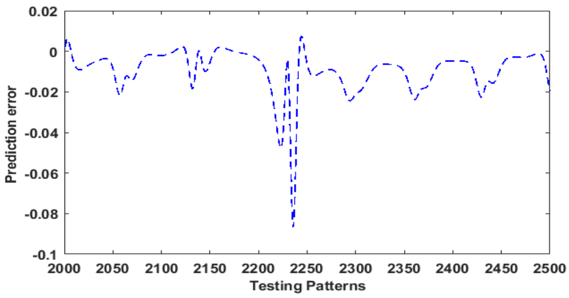

Figure 7. Evolution of the prediction error.

4. Conclusions

In this study, the GSA's hybrid configuration with SQP describes a novel method to calculate the output weights of ESN. The SQP method is introduced in GSA as a local search method to advance the standard GSA algorithm's performance. The results of the simulations indicated that the proposed HGSA-SQP-ESN performs superior to the existing ESN model. In future work, we could evaluate more effective intelligent algorithms to optimize the output weights of ESN. Further, combining the Differential Evolution (DE) to replace the GSA and DE with PSO significantly improves solid exploration.

References

[1] Ahmad Z., Memon M., Memon A, Munshi P., and Memon M., "A New Hybrid Approach of Gravitational Search Algorithm with Spiral-Shaped Mechanism-Based RBF Neural Network," *in Proceeding of the 22nd International Arab Conference on Information Technology*, Jordan, pp. 1-6, 2021.

[2] Bedekar P. and Bhide S., "Optimum Coordination of Directional Overcurrent Relays Using the Hybrid GA-NLP approach," *IEEE Transactions on Power Delivery*, vol. 26, no. 1, pp. 109-119, 2010.

[3] Fu T., "A Review on Time Series Data Mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164-181, 2011.

[4] Hornik K., Stinchcombe M., and White H., "Multilayer Feedforward Networks Are Universal Approximators,"*Neural Networks*, vol. 2, no. 5, pp. 359-366, 1989.

[5] Jaeger H. and Haas H., "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication," *Science*, vol. 403, no. 5667, pp. 78-80, 2004.

[6] Kobialka H. and Kayani U., "Echo State Networks With Sparse Output Connections," *in Proceeding of the International Conference on Artificial Neural Networks*, Berlin, pp. 356-361, 2010.

[7] Kumar Y., Verma S., and Sharma S., "Multi-Pose Facial Expression Recognition Using Hybrid Deep Learning Model With Improved Variant of Gravitational Search Algorithm," *The*

International Arab Journal of Information Technology, vol. 19, no. 2, pp. 281-287, 2022.
[8] Liu J., Sun T., Luo Y., Fu Q., Cao Y., Zhai J, and Ding X., "Financial Data Forecasting Using Optimized Echo State Network," *in Proceeding of the International Conference on Neural Information Processing*, Bangkok, pp. 138-149, 2018.
[9] Lorenz E., "Deterministic Nonperiodic Flow," *Journal of Atmospheric Sciences*, vol. 20, no. 2, pp. 130-141, 1963.
[10] Lv S., Peng L, and Wang L., "Stacked Autoencoder With Echo-State Regression for Tourism Demand Forecasting Using Search Query Data," *Applied Soft Computing*, vol. 73, pp. 119-133, 2018.
[11] Memon M., He J, Lu Y., Zhu N., and Memon A., "An Improvised Sub-Document Based Framework for Efficient Document Clustering," *Journal of Internet Technology*, vol. 20, no. 4, pp. 1191-1203, 2019.
[12] Memon M., Qu S., Lu Y., Memon A., and Memon A., "An Ensemble Classification Approach Using Improvised Attribute Selection," *in Proceeding of the 22nd International Arab Conference on Information Technology*, Jordan, pp. 1-5, 2021.
[13] Memon M., Lu Y., Chen P., Memon A., Pathan M., and Zardari Z., "An Ensemble Clustering Approach for Topic Discovery Using Implicit Text Segmentation," *Journal of Information Science*, vol. 47, no. 4, pp. 431-457, 2021.
[14] Peng H., Wu S., Wei C., and Lee S., "Time Series Forecasting With A Neuro-Fuzzy Modeling Scheme," *Applied Soft Computing*, vol. 32, pp. 481-493, 2015.
[15] Wang L., Hu H., Liu R., and Zhou X., "An Improved Differential Harmony Search Algorithm for Function Optimization Problems." *Soft Computing*, vol. 23, no. 13, pp. 4827-4852, 2019.
[16] Wang L., Lv S., and Zeng Y., "Effective Sparse Adaboost Method With ESN and FOA for Industrial Electricity Consumption Forecasting in China," *Energy*, vol. 155, pp. 1013-1031, 2018.
[17] Zhou H. and Qiao J., "Soft Sensing of Effluent Ammonia Nitrogen Using Rule Automatic Formation-Based Adaptive Fuzzy Neural Network," *Desalination and Water Treatment*, vol. 140, pp. 132-142, 2019.

Zohaib Ahmad has completed his Ph.D. from the school of Electronics & Information Engineering, Beijing University of Technology, Beijing. Zohaib Ahmed's Interests include Artificial Intelligence, Pattern Recognition, and Machine Learning.

Muhammad Qasim Memon is currently working as an Assistant Professor in the Department of Computer Science, University of Sufism and Modern Sciences, Bhitshah. Dr. Memon is also a Post-doctorate fellow at the Advanced Innovation Center for Future Education (AICFE), Faculty of Education, Beijing Normal University, China. He received his Ph.D. degree from the School of Software Engineering at Beijing University of Technology, China, in 2018. He received his Bachelor of Engineering and Master of Engineering from Mehran University of Engineering & Technology Jamshoro (MUET) in 2009 and 2014. Dr. Qasim has published several papers in international conferences and research journals indexed in SCI., EI., and Scopus. Dr. Memon's research interests include Educational Data Mining, Text Analytics, Information Extraction, and Technology Education.

Aasma Memon completed her Ph.D. from the School of Economics and Management at Beijing University of Technology, China. She received her Bachelor in Arts and Masters in Public Administration from the University of Sindh, Jamshoro, Pakistan, in 2008 and 2012. Her research interests include firm performance and corporate sustainability, human resource management, and data mining.

Parveen Munshi is currently working as a professor of Education and vice-chancellor in Sufism and Modern Sciences, Bhitshah. She completed her Ph.D. from Hamdard University, Karachi, in 2003. She completed her Master's and bachelor's in Education from the University of Sindh in 1986 and 1983. Dr. Munshi's research interests include Education Analytics, Psychological testing, Distance Education, and Computer Education.

Muhammad jaffar Memon is currently working at the SZAB campus, Mehran University of Engineering & Technology, Khairpur. He completed his Ph.D. from the College of Environmental Engineering at Beijing University of Chemical Technology, Beijing. Dr. Memon's research Include Biomass, Bioenergy, Air Pollution, and Water Pollution.