Regularization Activation Function for Extreme Learning Machine

Noraini Ismail[1], Zulaiha Ali Othman[2]
Center for Artificial Intelligence Technology
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Bangi, Selangor, Malaysia

Noor Azah Samsudin[3]
Jabatan Kejuruteraan Perisian
Fakulti Sains Komputer dan Teknologi Maklumat
Universiti Tun Hussein Onn Malaysia
Parit Raja, Johor, Malaysia

Abstract—Extreme Learning Machine (ELM) algorithm based on single hidden layer feedforward neural networks has shown as the best time series prediction technique. Furthermore, the algorithm has a good generalization performance with extremely fast learning speed. However, ELM facing overfitting problem that can affect the model quality due to the implementation using empirical risk minimization scheme. Therefore, this study aims to improve ELM by introducing an Activation Functions Regularization in ELM called RAF-ELM. The experiment has been conducted in two phases. First, investigating the modified RAF-ELM performance using four types of activation functions are: Sigmoid, Sine, Tribas and Hardlim. In this study, input weight and bias for hidden layers are randomly selected, whereas the best neurons number of hidden layer is determined from 5 to 100. This experiment used UCI benchmark datasets. The number of neurons (99) using Sigmoid activation function shown the best performance. The proposed methods has improved the accuracy performance and learning speed up to 0.016205 MAE and processing time 0.007 seconds respectively compared with conventional ELM and has improved up to 0.0354 MSE for accuracy performance compare with state of the art algorithm. The second experiment is to validate the proposed RAF-ELM using 15 regression benchmark dataset. RAF-ELM has been compared with four neural network techniques namely conventional ELM, Back Propagation, Radial Basis Function and Elman. The results show that RAF-ELM technique obtain the best performance compared to other techniques in term of accuracy for various time series data that come from various domain.

Keywords—*Extreme learning machine; prediction; neural networks; regularization; time series*

I. INTRODUCTION

Over the past decade, the use of Artificial Neural Networks (ANN) in time series prediction has been grown and evolved [1]. Although basically this technique is biologically inspired, ANN has been successfully implemented in various fields and performed very well [2], especially for the purpose of forecasting and classification. Compared to statistical-based forecasting techniques, the ANN approach has several unique features [3], such as: 1) non-linear and data driven; 2) not for explicit basic model (non-parametric); and 3) more flexible and universal which allowing the model to work with more complex time series. ANN model does not take the statistical data distribution into account because the suitable model is formed adapively based on the data provided.

Recent discussions on new research in ANN for time series predictions has been presented by [4]. There are various ANN prediction models in literature. The most common and popular ANN are multi-layer perceptron based feedforward hidden layer network. This model has been proposed for non-linear time series forecasting by [5] and the result has overcome traditional statistical methods such as regression and Box-Jenkins approaches. Recurrent neural network based on feedforward hidden layer networks has also been tested for time series forecasting [6]. This model is very dynamic and allows forecasting for non-linear time series from various fields [7], [8]. Although ANN prove predictably well in multiple applications, however, it has some limitations such as black-box based learning [1], over-fitting models and easily trapped in local minimum.

In addition, the Support Vector Machine (SVM) technique also is often used in time series forecasting. It was introduced by [9]. This prediction method based SVM uses the general regression model class, such as Support Vector Regression and Least-Square [10]. SVM can be categorized as Linear, Gaussian or Radial Basis Functions (RBF) [11], polynomials [12], and multi-layer perceptron classifications. For time series prediction [13], linear Support Vector Regression is built by minimizing the risk reduction of the structure (bound to general error) that leading to better predictive performance than conventional techniques.

Due to continuous research, there are many suggestions for improvement on ANN and SVM structures in literature. This is because these techniques play an important role in machine learning and data analysis. However, these two popular techniques face some challenging issues such as slow learning abilities [1]. This is a major constraint in the analysis of time series data. The key factors involved in this problem are the use of slow learning gradient-based algorithm during training process. To address this problem, [14] has introduced a new ANN framework called Extreme Learning Machine (ELM). This technique shows very good performance in predictions and even better than conventional neural network methods, but still exist some shortcomings which need further development and perfection [15]–[17].

Therefore, this research work is an attempt to improve ELM by introducing an Activation Functions Regularization in ELM called RAF-ELM in the contex of this paper. This paper is organized as follows. Section II discussed the existing

application using ELM based on literature review. In Section III discussed the proposed ELM and its implementation. Then, the dataset used and the result obtain are given and discussed in Section IV. Moreover, the performance of the investigated model also was analyzed and compared in Section IV. Finally, the paper is concluded in Section V.

II. RELATED WORK

In recent years, ELM has been used in various applications such as signal processing [18], image processing [19], [20], medical diagnosis [21], market analysis [22], aviation and aerospace [23], forecasting [24] and others [25]. In signal processing, [18] has applied the ELM algorithm to identify two different EEG (electroencephalogram) signals. This computer-based system can be used to determine the intention of a paralyzed patient by analyzing the recorded EEG signals from the patient's scalp. This study also uses non-linear character selection algorithms to eliminate less important features in the data set used. Whereas [26] use multiple kernel ELM to classify EEG signals called MKELM. The MKELM method is developed by integrating two types of kernel so that the algorithm can explore additional information from some non-linear feature space effectively for EEG classification. The experimental results confirm that the advantages of MKELM method for the EEG classification related with motor imaging in BCI applications provide high classification accuracy.

Furthermore, the new human face recognition algorithm based on the extraction of the histogram feature gradient and ELM oriented was proposed by [27]. The results show that the proposed algorithm performance is better than SVM and k-nearest neigbors algorithms. Additionally, the proposed algorithm also shows a significant increase in classification accuracy with short training time, one hundred times better and minimum dependence on the number of prototypes. In the same year, [28] proposed an effective ELM-based local scanning area for object recognition called MM-LRF-ELM. Experimental validation on the Washington RGB-D data set illustrates that the proposed combination method achieves better recognition performance. Whereas [19] proposed ELM based on local recruitment areas with three channels called 3C-LRF-ELM. This suggestion algorithm allows the hepatologic features to be automatically diagnose illness using a set of lungs, kidney and spleen image data sets. The study compares two types of neural network layers i.e. using a single and two hidden layer network. The results showed that 3C-LRF-ELM single layer neural network provided better classification performance.

ELM has also been successfully applied in medical diagnosis. Author in [29] proposed a novel hybrid diagnostic system called LFDA-RKELM which integrates character extraction techniques with ELM algorithms for the diagnosis of thyroid disease. The proposed method consists of three stages namely dimension reduction process based on feature extraction; data modeling using the ELM kernel; and lastly, the best classification model is used to perform thyroid disease diagnosis tasks using the most discriminatory subset of feature and optimum parameters. Experimental results indicate that

LFDA-RKELM overcame the basic method. In contrast to the study, [21] improves the ELM algorithm using the competitive swarm optimization technique. The proposed model has been tested based on 15 medical classification datasets. Experimental results show that the proposed model can achieved better generalized performance with smaller hidden neurons and with greater stability. However, it requires more training time than other ELM-based metaheuristic. While [30] submitted an application to predict the Huntington's disease several years earlier based on data from MRI brain scan. Experimental results show that the predictions are realistic with reasonable accuracy provided that the missing values are dealt with caution.

In addition, ELM has also been applied in market analysis, aviation and aerospace, but not so extensive. Market analysis is a documented investigation of company planning activities, in particular results involving inventory, purchasing, expansion or reduction of work, facility expansion, equipment purchasing capital, promotional activities, and many other aspects involving the company. In this domain, [31] has presented a design for stock index movement analysis using the ELM kernel. The findings suggest that the proposed method provides good results. However, researchers found that the ELM kernel needed more CPU resources than RBFN techniques and SVM that led to longer processing times. Next, [32] using ELM to forecast cash outflow for financial institutions. In this study, the ELM algorithm shows good predictive results in terms of accuracy, error rate and processing time. While in the field of aviation and aerospace, [23] has proposed ELM techniques to improve absolute position accuracy and solve complicated modeling and computing problems for aviation drilling robots. The study was carried out by considering the effect of geometric factor and non-geometric factor in building a predictive positional error model using ELM techniques. The results show that the accuracy of the absolute position and the maximum point of the robot center has increased by 75.89% and 80.93%. Finally, ELM is also used to solve problems in environment domain. Author in [33] used the ELM to predict the air quality index in Delhi, India. In the proposed model, the air pollution quality index and the previous day's meteorological conditions are used for predictions. The performance of this proposed model is compared with the existing forecasting system (SAFAR). The results show that ELM provides higher accuracy results than SAFAR.

Based on the state-of-the-art, it can be concluded that past studies have used various techniques in predicting and analyzing time series data. ANN techniques show satisfactory results, better than regular statistical methods. The popular technique in ANN is ELM, BP, RBFN and Elman. This study was focusing on the performance of ELM techniques in various fields. From the literature, it can be seen that ELM technique is able to overcome other ANN techniques. This technique has also been successfully applied in various fields and obtained very satisfactory results. However, this technique has a potential to generate over-fitting models and leads to a less stable when it comes to a certain condition. Therefore, this study proposed to enhance ELM performance. This paper is an extension of a previous study and further explorations of

the proposed enhancement in proceeding paper of International Conference on Intelligent and Interactive Computing entitled "Improved Ozone Pollution Prediction using Extreme Learning Machine with Tribas Regularization Activation Function".

III. IMPROVED MODEL OF EXTREME LEARNING MACHINE

The ELM algorithm is well known for its fast speed capability and shows good generalization performance. However, ELM has several weaknesses as it tends to produce over-fitting model because ELM is developed based on the Empirical Risk Minimization principles. In addition, ELM also has a weak capacity control because it calculates the min least-square norms directly that may lead to a less robust estimation when dealing with unexpected or outliers. Therefore, this study proposes Activation Functions Regularization in ELM called RAF-ELM based on Structural Risk Minimization scheme that according to statistical learning theory.

In machine learning, the regularization function is defined as a process of introducing additional information to solve the over-fitting problem. Generally, regularization is a technique used for objective functions to solve optimization problems [34]. This optimization problem is a process for achieving ideal results. In addition, optimization can also be defined as a form of optimizing an existing solution, or designing and creating something optimally. Therefore, this study proposes improvements in ELM techniques by adding this regularization to make the activation function more balance and resilient to random non-uniform distribution, thus can improve ELM performance. A formula of RAF-ELM with an added regularization component $R(f)$ to the activation function $f(x)$ can be summarized in equation 1:

$$\sum_{i=1}^{n} \beta_i [f(z_i) + R(f_i)] = y$$

$$(f) = \lambda \sum_{k=1}^{N} \theta_k^2 \qquad (1)$$

The function of $R(f)$ is added to the activation function $f(z)$. λ is the parameter to impose penalties on the complexity of $f(z)$ to improve the ability of the model to be learned. The parameter value can be adjusted to help the algorithm find the appropriate value for the model. However, if the $R(f)$ value is too small, its function may not do anything and if the value is too large, it can cause under-fit model and loss valuable information. Therefore, the aim of this learning problem is to find the appropriate parameter by predicting the possibility of input (label) and giving a minimum error.

In addition, the number of neurons in hidden layers is varied (usually set at random) and improper value setting will affect the accuracy of results [35]. Therefore, this study will improve the structure of the model by adding a neuron parameter tuning function from [5 to 100] to improve the performance of the ELM algorithm. The resulting models will be evaluated and the number of neurons that perform greatest for the time series data will be identified.

In addition, many recent studies focus only on the use of the Sigmoid activation function, regardless of the performance of other activation functions such as Sin, Hardlim, and Tribas. This is because the Sigmoid activation function has easier math calculations and always gives excellent results [33]. Therefore, this study will compare the four activation functions of Sigmoid, Sin, Hardlim and Tribas to see the effect of different activation functions on RAF-ELM algorithm performance for time series data. The pseudo code for the RAF-ELM algorithm is shown in Fig. 1. The improvements are in line 16-18 where the regularization function is added to the activation function. The pseudo code for the activation functions of Sigmoid, Sin, Hardlim and Tribas can be seen in Fig. 2.

RAF-ELM Algorithm-Pseudocode

```
1 START
2   Input:
3       Sequence of training data, D = (x,t)
4       Type of activation function, f(x) =      5      Sigmoid, Sine, Tribas,
Hardlim
6       NeuronNum, i = [5,100]
7       Regularization function, j = [0.01,1]
8
9   Initialize input weight, wi and bias, bi
10  randomly
11
12  FOR EACH (int i = 0; i < NeuronNum; i++)
13      Compute the hidden output matrix
14      H(w1,...,wL, x1,...,xN, b1,...,bL)
15
16          FOR (int j = 0; j < NeuronNum; j++) 17        Calculate f(x)
18          END FOR
19
20      Compute output weight based on class
21      label
22
23  END FOR
24  repeat process until stoping condition
25  achieved
26 END
```

Fig. 1.   Pseudo-Code of the RAF-ELM Algorithm.

```
1   if (func.startsWith("sig")) {
2       for (int j = 0; j < NumberofHiddenNeurons; j++) {
3           for (int i = 0; i < numTestData; i++) {
4               double temp = tempH_test.get(j, i);
5               temp = 1.0f / (1 + Math.exp(-temp));
6               H_test.set(j, i, temp + norm);
7           }
8       }
9   } else if (func.startsWith("sin")) {
10      for (int j = 0; j < NumberofHiddenNeurons; j++) {
11          for (int i = 0; i < numTestData; i++) {
12              double temp = tempH_test.get(j, i);
13              temp = Math.sin(temp);
14              H_test.set(j, i, temp + norm);
15          }
16      }
17  } else if (func.startsWith("hardlim")) {
18      for (int j = 0; j < NumberofHiddenNeurons; j++) {
19          for (int i = 0; i < numTestData; i++) {
20              Random rndom = new Random();
21              double d = rndom.nextBoolean() ? 1 : -1;
22              double temp = tempH_test.get(j, i) * d;
23              temp = (temp >= 0) ? 1 : 0;
24              H_test.set(j, i, temp + norm);
25          }
25      }
27  } else if (func.startsWith("tribas")) {
28      for (int j = 0; j < NumberofHiddenNeurons; j++) {
29          for (int i = 0; i < numTestData; i++) {
30              Random rndom = new Random();
31              double d = rndom.nextBoolean() ? 1 : -1;
32              double temp = tempH_test.get(j, i) * d;
33              temp = (temp >= -1 && temp < 1) ? 0 : 1 - Math.abs(temp);
34              H_test.set(j, i, temp + norm);
```

Fig. 2.   Pseudo Code of Activation Function.

IV. EXPERIMENT AND RESULT

In ELM, there are four important variables including the number of neurons, the random ranges between the input layer and the hidden layer, the random range of hidden nodes, and the type of activation function. However, this study focuses on two main variables, namely the activation function using RAF and the number of neurons whereas the range between the input layer and the hidden layer, and the range of hidden nodes is determined randomly to maintain the fast learning concept of ELM technique. This is because if all variables are controlled by parameter tuning, the algorithm's learning speed will slow down [36], [37].

To test the effects of activation functions and neuron number on RAF-ELM performance, several experimental sets have been carried out. In this experiment, the parameter tuning of the neuron number in hidden layers is set to [5-100], a random range of values for hidden weights [-1, 1], and a random range of hidden node thresholds [0, 1]. The test variable (i.e., activation function of RAF) is set to Sigmoid, Sin, Hardlim, or Tribas.

For this experiment, the result of the error value obtained was calculated based on MAE. The small MAE values show the best model. This study will also take the processing time into consideration as an efficiency measurement of the model.

A. Dataset

This study is used the air quality data from UCI Machine Learning repository, where data sources are collected by Saverio De Vito from ENEA (National Agency for New Technologies, Energy and Sustainable Economic Development) in Italy. This original UCI data set comprises 9358 records which contain a gas response from multi-sensor devices that used to measure air quality within the city. This multi-sensor device is placed in a highly exposed area with high air pollution. The datasets are recorded from March 2004 to February 2005 (one year) which contains 15 attributes. The air quality estimation problem cannot be well solved with to the lost precision over time due to the emergence of sensors drift and concept drifts caused by seasonal influence, human behavior and sensor aging. In our experiment, we mainly solved Carbon Oxide (CO) concentration estimation problem. The result will be compare with the study in [38] that improve Elman network to solve silimilar problem using same dataset.

B. The Effect of different Activation Functions on the Performance of RAF-ELM

Selecting the activation function that corresponds to the algorithm as well as the type of data is very important. Therefore, this study will compare 4 types of activation functions: Sigmoid, Sin, Hardlim and Tribas. Fig. 3 shows the average of the results for the activation function variable where the neuron number parameter setting is set to [5-100], the random range of the hidden weights [-1, 1], and random ranges of hidden node [0, 1].

Average results indicate that Sigmoid gives a smaller error value as a whole with a value of 0.042687 MAE, followed by Tribas (0.043940 MAE), Sin (0.045456 MAE) and Hardlim

(0.102183 MAE). The overall error value for Sigmoid and Tribas is approximately, with a difference of 0.001254 MAE. While Hardlim gives the highest average result with error value of 0.102183 MAE. For processing time, the activation function of Sin shows a shorter time duration than other activation functions with a value of 0.030715 s, Sigmoid with 0.040186 s and followed by Tribas (0.047390 s) and Hardlim (0.049683 s).

C. The Effect of different Number of Neuron on the Performance of RAF-ELM

In this phase, experiments were performed to see the best neurons for the proposed RAF-ELM method. According to [15], the number of neurons has a significant impact on the performance of ELM techniques. Therefore, the appropriate number of neurons should be identified because the determination of arbitrary number of neurons will affect the ELM's performance. The automatic tuning [39] should be used by setting neuron's initial value using additional constructive techniques. This experiment set the number of neurons to [5-100].

Fig. 4 illustrates optimal results for each number of hidden neurons. Based on this figure, the graph shows that the number of neurons that are too small gives poor performance for RAF-ELM forecasting model, while the RAF-ELM model stabilizes when the number of neurons is 35 and above for the activation function of Sigmoid, Sin and Tribas. RAF-ELM forecasting model for the number of neuron 99 for Sigmoid achieves best performance with error value 0.037955137 MAE. In addition, the Tribas activation function is seen to be performing well for the number of neurons 5 to 30 compared to other activation functions and thus providing similar results with the activation function of Sigmoid and Sin for the number of neurons of 31 to 100. This finding is similar in the study by proceeding paper that is mentioned before. However, Tribas activation function does not give an optimal result. While Hardlim provides a stable result for the number of neurons 5 to 100 with a small amount of difference, but unable to outperform other activation functions.

Table I is the summary of optimal error values for all activation fuvtion. Based on the result, it can be seen that the Sigmoid activation function provides optimal results with 0.037955 MAE error value, followed by Sin (0.038087 MAE), Tribas (0.03963 MAE) and Hardlim (0.101762 MAE). The optimal result gives the same result as the overall average value in Fig. 3 where sigmoid activation function is better than other activation functions.

TABLE I.   SUMMARY OF OPTIMAL ERROR VALUES FOR ALL ACTIVATION FUNCTION

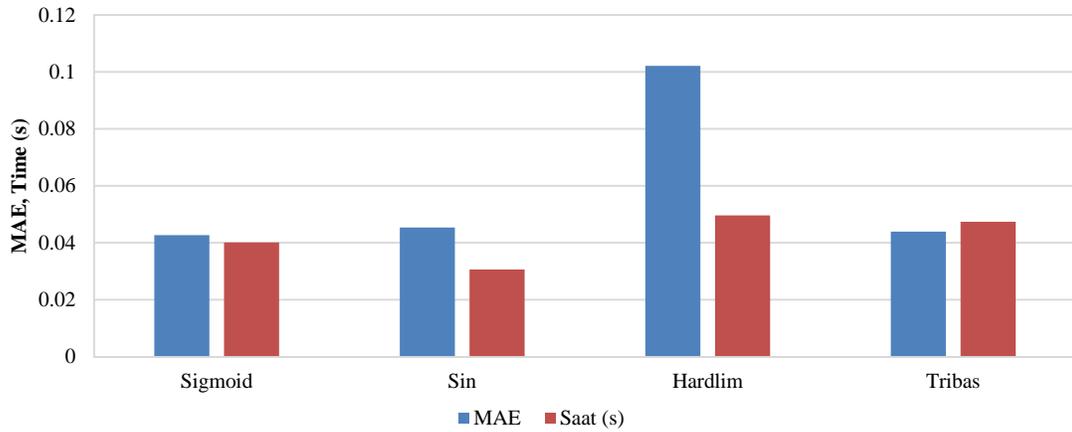| Activation Function	Num. of Neuron	MAE		Time (s)	
		Train	*Test*	*Train*	*Test*
Sig	**99**	**0.037955**	**0.037955**	**15.312**	**0.014**
Sin	97	0.038087	0.038082	14.953	0.014
Hardlim	20	0.101762	0.101285	3.14	0.016
Tribas	75	0.039630	0.039627	11.625	0.015

Fig. 3. Comparison of Average Results for four Types of Activation Functions based on MAE Error Rate.
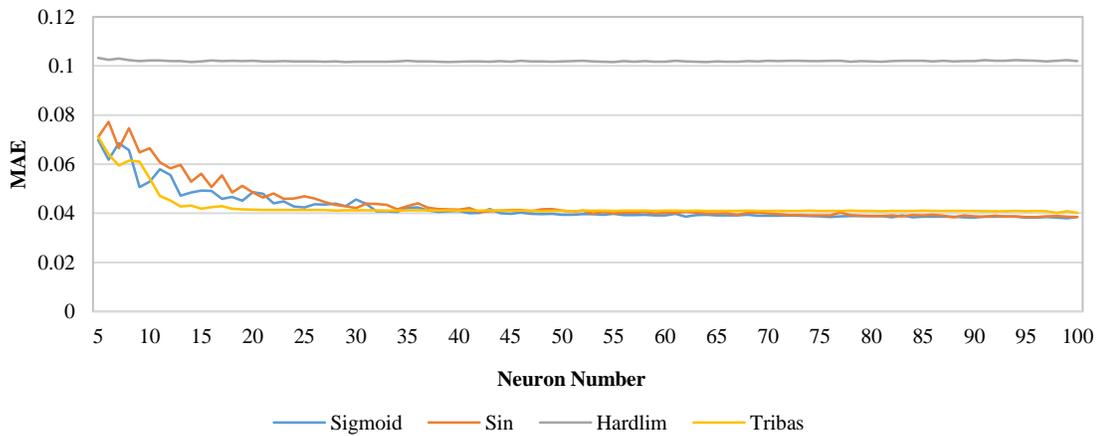

Fig. 4. Comparison of different Neuron Number for four Types of Activation Functions based on MAE Error Rate.
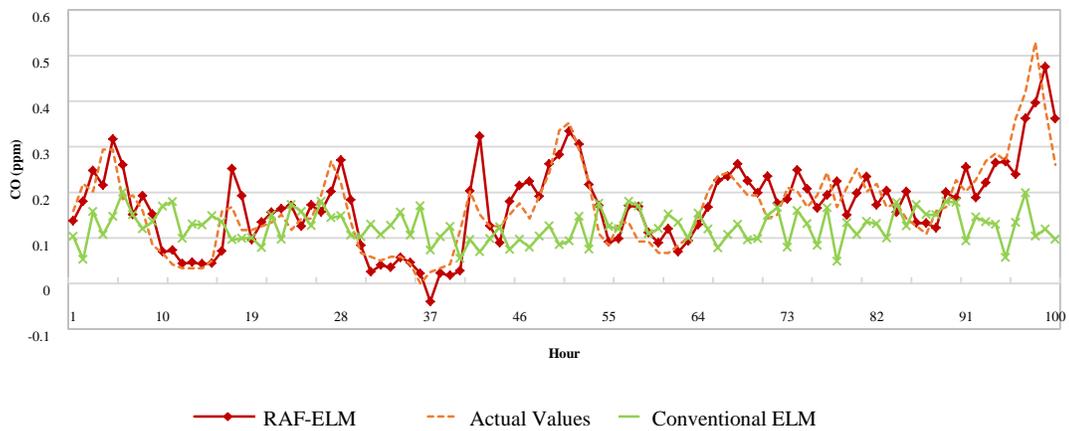

Fig. 5. Comparison of Actual Values with Predictive Values for the Optimal RAF-ELM and Conventional ELM within 100 Hours.

Fig. 5 shows the difference of actual value and predictive value between the optimal RAF-ELM techniques and the conventional ELM techniques. Based on the figure, RAF-ELM technique performs better than the conventional ELM with a small distinction between the non-linear lines of the actual value and the predicted value. Then, t-tests were conducted to see the significant differences between the conventional ELM and RAF-ELM neural network techniques. Table II shows, there was significant difference in the accuracy level of the conventional ELM and RAF-ELM performance (t = 12.45; p <0.05).

Compared to previous studies using the same UCI data in forecasting, [38] has improved Elman's neural network technique and obtained better performance from BP techniques and original Elman with 0.0769 MSE results. Whereas, the optimal error rate for RAF-ELM is 0.0415 MSE and 0.03796 MAE. This shows that RAF-ELM technique was able to outperform the improved Elman technique in [38] with a difference of 0.0354 MSE, thus far better than the BP and the traditional Elman technique.

This study has suggested the best predictive techniques of ELM neural network techniques. The discussions on improvements and the selection of variables are described in detail. Proper parameter determination also contributes to good modeling. Therefore, a comprehensive study on the relationship between parameter determinations and performance of the proposed RAF-ELM algorithm is based on two test variables including type of activation function and number of hidden neurons. The results show that the activation function of Sigmoid achieves the best performance in most cases while the number of neurons is found to provide more stable prediction results at values of 35 to 100.

Overall, the forecasting model that provides optimal performance results is the proposed RAF-ELM technique where the activation function is Sigmoid and the number of neurons is 99.

TABLE II. T-TEST RESULT FOR RAF-ELM OVER CONVENTIONAL ELM

	T-test	Significat Level (p)	Significant Status (p<0.05)
RAF-ELM	12.45	<.0001	Yes
ELM			

V. VALIDATION

The validation phase was conducted to verify the effectiveness of proposed RAF-ELM techniques in various data environments.

As previously mentioned, the comparison of technical performance is based on the error result of MAE and processing time. The result of proposed RAF-ELM will be compare with other ANN techniques namely conventional ELM, BP, RBFN and Elman using 15 set of benchmark data have been selected from the OpenML repository (https://www.openml.org). Table III shows the list of names and data set information to be used. Based on Table IV, it can be concluded that the RAF-ELM technique provides better results for all sets of benchmark data. In addition, processing time also indicates that the RAF-ELM technique provides better results than BP, RBFN and Elman. This shows that the proposed RAF-ELM technique provides good performance on various time series data that come from various domains.

TABLE III. SPECIFICATION OF REAL-WORLD REGRESSION CASES

No.	Data Sets	Attributes	Total Record
1	SWD	11	1000
2	LEV	5	1000
3	USCrime	14	47
4	Sleuth	7	93
5	Vineyard	3	52
6	Elusage	3	55
7	Machine_cpu	7	209
8	Satellite_image	37	6435
9	GeographicalOriginalofMusic	118	1059
10	Analcatdata_vehicle	5	48
11	Wind	15	6574
12	Pol	49	15000
13	Bodyfat	15	252
14	Pollen	5	3848
15	Houses	9	20640

TABLE IV.   PERFORMANCE COMPARISON RESULTS OF RAF-ELM WITH CONVENTIONAL ELM, BP, RBFN AND ELMAN OVER 15 REGRESSION BENCHMARK DATASETS

Data Sets	RAF-ELM		ELM		BP		RBFN		Elman	
	MAE	*Time (s)*	*MAE*	*Time (s)*	*MAE*	*Time (s)*	*MAE*	*Time (s)*	*MAE*	*Time (s)*
SWD	**0.250285**	**0.015**	0.328954	0.019	0.5408	1.11	0.5378	0.03	0.5029	0.75
LEV	**0.211866**	**0**	0.245896	0.013	0.544	0.31	0.7264	0.02	0.5221	0.47
USCrime	**0. 443344**	**0**	0.761179	0	20.1735	0.09	23.5787	0	21.7713	0.06
Sleuth	**1.052529**	**0**	5.402610	0	57.1175	0.06	65.0199	0.02	84.7826	0.06
Vineyard	**0.174844**	**0**	3.017522	0	2.2966	0.03	1.5853	0.8408	2.6887	0.03
Elusage	**6.491505**	**0**	6.784657	0	8.4883	0.02	19.3721	0	16.3226	0.02
Machine cpu	**2.982449**	**0.016**	0.352996	0	26.4298	0.11	36.1974	0.02	21.1478	0.13
Satellite_image	**0.20239**	**0.031**	0.384473	0.07	0.9936	63.91	1.6722	0.75	0.7105	11.7
Geographical Original of Music	**0.123212**	**0**	0.176078	0.01	0.5124	99.54	0.7292	0.75	0.4824	5.25
Analcatdata vehicle	**0.221790**	**0**	0.534427	0	147.1069	0.02	176.611	0	185.5561	0.03
Wind	**0. 027079**	**0.04**	0. 06120	0.06	2.6852	11.97	3.9969	0.35	2.6666	5.83
Pol	**0.373177**	0.049	0.776129	**0.012**	13.9565	251.5	29.2806	3.29	11.6061	33.7
Bodyfat	**0.298404**	**0**	0.357808	0.015	0.407	0.47	5.7504	0.4101	0.6772	0.23
Pollen	**0.113975**	0.032	1.093838	**0.019**	1.2649	1.31	2.3506	0.17	1.2548	1.8
Houses	**0.090348**	0.091	0.508891	**0.013**	46560.93	16.29	91801.7	0.44	53696.89	12.9

VI. CONCLUSION

This study was conducted to suggest the best predictive model for time series data using neural network techniques. In achieving this objective, several sets of experiments have been implemented. The main phase of the experiment is to anticipate improvements to ELM techniques to improve predictive performance. In assessing the performance of a proposed ELM called RAF-ELM, a set of experiments was conducted based on several key variables ie the activation function type, the number of neurons and the value of the controlling function. Based on this experiment, the best forecasting method has been selected. The proposed RAF-ELM method provides an optimal predictive decision on Sigmoid activation function, the number of neurons 99 and the value of escort 0.7. This optimal RAF-ELM method has also been validated through validation experiments on 15 sets of regression data. The results show that the proposed RAF-ELM model provides excellent overall performance results. This proves that the proposed ELM model not only provides good performance on time series data, but also for many other types of data. However, there are several limitations in this study that require further research. Among the suggestions for future studies is integration of RAF-ELM algorithms with other algorithms or technologies. In recent years, researchers often combine the best features in certain algorithms to improve the efficiency of an algorithm. This method is called hybridization method. Based on previous studies, genetic algorithms and Support Vector Machines often provide good research results and solve problems in various fields. Therefore, the combination of genetic algorithms, Support Vector Machines and RAF-ELM are seen to be able to produce better training models and this study is worth the effort to explore. The second suggestion is expanding the application of RAF-ELM algorithm. Although the ELM technique has the advantage in theory, the use in real-world applications is limited. This is because ELM is a new algorithm compared to other neural network techniques. Therefore, how to effectively implement ELM in daily life is an important aspect of future research.

REFERENCES

[1] N. da Silva, D. Hernane Spatti, R. Andrade Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, Artificial Neural Networks, vol. 26, no. 9. 2017.

[2] S. Walczak, "Artificial neural networks," in Encyclopedia of Information Science and Technology, Fourth Edition, IGI Global, 2018, pp. 120–131.

[3] Z. Zhang, "Artificial neural network," in Multivariate Time Series Analysis in Climate and Environmental Research, Springer, 2018, pp. 1–35.

[4] A. S. Weigend, Time series prediction: forecasting the future and understanding the past. Routledge, 2018.

[5] A. S. Lapedes, E. W. Steeg, and R. M. Farber, "Use of Adaptive Networks to Define Highly Predictable Protein Secondary–Structure Classes," Mach. Learn., vol. 21, no. 1, pp. 103–124, 1995.

[6] C. de Groot and D. Würtz, "Analysis of univariate time series with connectionist nets: A case study of two classical examples," Neurocomputing, vol. 3, no. 4, pp. 177–192, 1991.

[7] G. Grudnitski and L. Osburn, "Forecasting S&P and gold futures prices: An application of neural networks," J. Futur. Mark., vol. 13, no. 6, pp. 631–643, 1993.

[8] C. Kuan and T. Liu, "Forecasting exchange rates using feedforward and recurrent neural networks," J. Appl. Econom., vol. 10, no. 4, pp. 347–364, 1995.

[9] V. Vapnik, "Support vector machine," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[10] R. M. Balabin and E. I. Lomakina, "Support vector machine regression (SVR/LS-SVM) - An alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data," Analyst, vol. 136, no. 8, pp. 1703–1712, 2011.

[11] Caroline Kleist, "Time Series Data Mining Methods: A Review," Eng. Appl. Artif. Intell., vol. 24, no. 533039, p. 61, 2015.

[12] C. Chatfield, "Time-series forecasting," Significance, vol. 2, no. 3, pp. 131–133, 2005.

[13] L. J. C. L. J. Cao, K. S. C. K. S. Chua, and L. K. G. L. K. Guan, "Combining KPCA with support vector machine for time series forecasting," 2003 IEEE Int. Conf. Comput. Intell. Financ. Eng. 2003. Proceedings., no. 1, pp. 325–329, 2003.

[14] G. Huang, Q. Zhu, and C. Siew, "Extreme Learning Machine : A New Learning Scheme of Feedforward Neural Networks," IEEE Int. Jt. Conf. Neural Networks, vol. 2, pp. 985–990, 2004.

[15] G. Huang, G. Bin Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," Neural Networks, vol. 61. pp. 32–48, 2015.

[16] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, "Extreme learning machine: algorithm, theory and applications," Artif. Intell. Rev., vol. 44, no. 1, pp. 103–115, 2015.

[17] X. Liu, S. Lin, J. Fang, and Z. Xu, "Is extreme learning machine feasible? A theoretical assessment (Part I)," IEEE Trans. Neural Networks Learn. Syst., vol. 26, no. 1, pp. 7–20, 2015.

[18] H. Zhao, X. Guo, M. Wang, T. Li, C. Pang, and D. Georgakopoulos, "Analyze EEG signals with extreme learning machine based on PMIS feature selection," Int. J. Mach. Learn. Cybern., vol. 9, no. 2, pp. 243–249, 2018.

[19] J. Fang, X. Xu, H. Liu, and F. Sun, "Local receptive field based extreme learning machine with three channels for histopathological image classification," Int. J. Mach. Learn. Cybern., pp. 1–11, 2018.

[20] J. X. Tang, C. W. Deng, G. B. Huang, and B. J. Zhao, "Compressed-Domain Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine," Ieee Trans. Geosci. Remote Sens., vol. 53, no. 3, pp. 1174–1185, 2015.

[21] M. Eshtay, H. Faris, and N. Obeid, "Improving Extreme Learning Machine by Competitive Swarm Optimization and its application for medical diagnosis problems," Expert Syst. Appl., vol. 104, pp. 134–152, 2018.

[22] P. P. Das, R. Bisoi, and P. K. Dash, "Data decomposition based fast reduced kernel extreme learning machine for currency exchange rate forecasting and trend analysis," Expert Syst. Appl., vol. 96, pp. 427–449, 2018.

[23] P. Yuan, D. Chen, T. Wang, S. Cao, Y. Cai, and L. Xue, "A compensation method based on extreme learning machine to enhance absolute position accuracy for aviation drilling robot," Adv. Mech. Eng., vol. 10, no. 3, p. 1687814018763411, 2018.

[24] Z. M. Yaseen et al., "Predicting compressive strength of lightweight foamed concrete using extreme learning machine model," Adv. Eng. Softw., vol. 115, pp. 112–125, 2018.

[25] M. Alizamir, O. Kisi, and M. Zounemat-Kermani, "Modelling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data," Hydrol. Sci. J., vol. 63, no. 1, pp. 63–73, 2018.

[26] Y. Zhang et al., "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces," Expert Syst. Appl., vol. 96, pp. 302–310, 2018.

[27] S. Naik and R. P. K. Jagannath, "GCV-Based Regularized Extreme Learning Machine for Facial Expression Recognition," in Advances in Machine Learning and Data Science, Springer, 2018, pp. 129–138.

[28] H. Liu, F. Li, X. Xu, and F. Sun, "Multi-modal local receptive field extreme learning machine for object recognition," Neurocomputing, vol. 277, pp. 4–11, 2018.

[29] C. Ma, J. Guan, W. Zhao, and C. Wang, "An Efficient Diagnosis System for Thyroid Disease Based on Enhanced Kernelized Extreme Learning Machine Approach," in International Conference on Cognitive Computing, 2018, pp. 86–101.

[30] E. Eirola, A. Akusok, K.-M. Björk, H. Johnson, and A. Lendasse, "Predicting Huntington's Disease: Extreme Learning Machine with Missing Values," in Proceedings of ELM-2016, Springer, 2018, pp. 195–206.

[31] X. Li et al., "Empirical analysis: stock market prediction via extreme learning machine," Neural Comput. Appl., 2016.

[32] J. Xue, S. H. Zhou, Q. Liu, X. Liu, and J. Yin, "Financial time series prediction using ℓ2,1RF-ELM," Neurocomputing, 2017.

[33] M. Bisht and K. R. Seeja, Air pollution prediction using extreme learning machine: A case study on Delhi (India), vol. 79. 2018.

[34] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent Neural Network Regularization," Iclr, no. 2013, pp. 1–8, 2014.

[35] W. Cao, J. Gao, Z. Ming, and S. Cai, "Some Tricks in Parameter Selection for Extreme Learning Machine," in IOP Conference Series: Materials Science and Engineering, 2017, vol. 261, no. 1.

[36] J. Bergstra JAMESBERGSTRA and U. Yoshua Bengio YOSHUABENGIO, "Random Search for Hyper-Parameter Optimization," J. Mach. Learn. Res., 2012.

[37] M. Williams et al., Numerical Optimization. 2013.

[38] G. Mei, H. Zhang, and B. Zhang, "Improving Elman neural network model via fusion of new feedback mechanism and Genetic Algorithm," in Progress in Informatics and Computing (PIC), 2016 International Conference on, 2016, pp. 69–73.

[39] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification.," IEEE Trans. Syst. Man. Cybern. B. Cybern., 2012.