

A comparison of activation functions in multilayer neural network for predicting the production and consumption of electricity power

Abdulwahed Salam¹, Abdelaaziz El Hibaoui², Abdulgabbar Saif³

^{1,2}Faculty of Science, Abdelmalek Essaadi University, Morocco

³Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, University Kebangsaan Malaysia, Malaysia

Article Info

Article history:

Received Aug 4, 2018

Revised Jul 4, 2020

Accepted Jul 29, 2020

Keywords:

Activation functions

Artificial neural networks

multilayer perceptrons

Energy prediction

ABSTRACT

Predicting electricity power is an important task, which helps power utilities in improving their systems' performance in terms of effectiveness, productivity, management and control. Several researches had introduced this task using three main models: engineering, statistical and artificial intelligence. Based on the experiments, which used artificial intelligence models, multilayer neural networks model has proven its success in predicting many evaluation datasets. However, the performance of this model depends mainly on the type of activation function. Therefore, this paper introduces an experimental study for investigating the performance of the multilayer neural networks model with respect to different activation functions and different depths of hidden layers. The experiments in this paper cover the comparison among eleven activation functions using four benchmark electricity datasets. The activation functions under examination are sigmoid, hyperbolic tangent, SoftSign, SoftPlus, ReLU, Leak ReLU, Gaussian, ELU, SELU, Swish and Adjust-Swish. Experimental results show that ReLU and Leak ReLU activation functions outperform their counterparts in all datasets.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Abdulwahed Salam,

Faculty of Science,

Abdelmalek Essaadi University,

Tetouan, Morocco.

Email: sallam.ye@yahoo.com

1. INTRODUCTION

Electrical power is produced from different sources, some of them are renewables and others are non-renewables. Nowadays, consumers can become producers if they have other power resources, especially renewable such as solar or wind power, and this turned to be difficult to control. Due to that, the need for a stable system to manage this new attitude emerged. This system is standing for what we have and what we need. Many organizations are using information system to help their jobs by storing information and use that information to help in making decisions in the future. Planning by forecasting is to analyze historical data to save energy a usage [1]. This data had been collected over the time of the progress. This data is formatted minutely, hourly, daily, weekly, monthly, quarterly, or yearly. Also, it consists of univariate or multivariate depending on the input variables that affect electricity. Due to the complication of prediction, several models have been proposed. These models include engineering, statistical and artificial intelligence [2]. The most

used model is artificial neural network (ANN). Fully connected supervised network with back propagation learning rule is the most often used.

This type of ANN is excellent at prediction. To predict one or more values from the input variables is the aim of supervised learning. Regression that depends on example pairs of data is a form of supervised learning. This type of network is a fully interconnected neurons in different layers, the input layer, the output layer, and the hidden layers which have at least on hidden layer [3]. The hidden layer connect the input and the output layers. The input layer receives data that the neural network learns from and the output layer provides response to the input data. There is no exact role to determine the number of hidden layers and the number of neurons in the hidden layer and there were many studies that tried to figure these numbers. The neural network is learned by algorithms. There are many learning algorithms in the neural network but the most commonly used is back propagation of error. The error in prediction is fed backwards through the network during training to adjust the weights and minimize the error. This step is repeated until achieving the minimum error or reaching the specific number of epochs [4]. In the center of network stands the activation function. In theory, any function can be utilized as an activation function. However, the activation function have a linear and nonlinear character. Nonlinear activation is important in order to be capable to distinguish the complex relationships which exist in the feature space [5]. Whereas linear functions are particularly used in output layers especially for regression, non-linear activation functions can be utilized in hidden layer. Also non-linear can be used in output layers especially for classification [6]. Moreover, derivative of activation function is involved in the calculation for error which effects on weights of neuron connections.

We organized the rest of this paper as follows: Section 2 presents activation functions in comparison. In section 3 we present data sets description and methodology that used in this work in section 4. Experiments and their results covered in Section 5 Finally, Section 6 concludes the paper.

2. ACTIVATION FUNCTIONS IN COMPARISON

The choice of activation functions strongly influences the performance of neural networks. In this session, we describe the most used activation functions. Sigmoid [7] is a widely used activation function. The output values are between zero and one. Its equation is:

$$f(x) = \sigma(x) = \frac{1}{(1 + e^{-x})} \quad (1)$$

Due to simplicity of sigmoid derivative, Sigmoid is fast to execute. But its problem is that gradient approaches to zero and the learning of network becomes difficult. The hyperbolic tangent function [8] is zero centered and its output between -1 and 1,

$$f(x) = \frac{e^x - e^{-x}}{(e^x + e^{-x})} \quad (2)$$

SoftSign [8] function is closely related to hyperbolic tangent. SoftSign converges polynomially whereas hyperbolic tangent converges exponentially. SoftSign defined as

$$f(x) = \frac{x}{1 + |x|} \quad (3)$$

Rectified linear unit (ReLU) [9] defined as positive part of x:

$$f(x) = x^+ = \max(0, x) \quad (4)$$

It is simple, and fast. In addition, it rectifies vanishing gradient problem. The most aspect of ReLU is dead neurons which mean it is never been activated when x is less than zero. In [10], authors defined a leak rectified linear unit (LReLU) as,

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{otherwise} \end{cases} \quad (5)$$

It is similar to ReLU. The only difference is that LReLU flows when $x < 0$ and that solves the dead neurons of ReLU from the author's point of view.
SoftPlus [11]:

$$f(x) = \log(1 + e^x) \tag{6}$$

SoftPlus is just a smooth approximation to the LReLU but it is positive. Figure 1 shows the difference between among Sigmoid, Tanh and SoftSign, ReLU, Leak ReLU and SoftPlus.

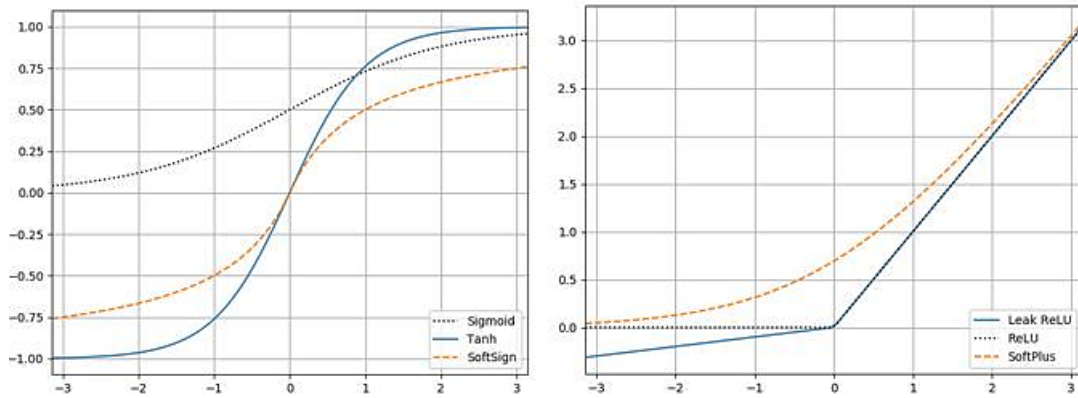


Figure. 1. Sigmoid, Tanh, SoftSign, Leak ReLU, ReLU and SoftPlus activation functions

Gaussian activation function [12], defined as,

$$f(x) = e^{-x^2} \tag{7}$$

Exponential linear unit (ELU) [13] is like ReLU and leaky ReLU for avoiding a vanishing gradient by the identity of positive values,

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha(e^x - 1) & \text{otherwise} \end{cases} \tag{8}$$

where $\alpha=1.0$

Scaled exponential linear unit (SELU) [14] which is a modified type of ELU with two fixed parameters. It is defined as,

$$f(x) = \lambda \begin{cases} x & \text{for } x \geq 0 \\ \alpha(e^x - 1) & \text{for } x < 0 \end{cases} \tag{9}$$

where $\alpha \approx 1.6733$ and $\gamma \approx 1.0507$

Google Brain introduces a new activation function which is called Swish [15]

$$f(x) = \frac{x}{(1 + e^{-\beta x})} \tag{10}$$

Adjusted swish (E-Swish) [16] is a modified type of Swish and is defined as

$$f(x) = \frac{\beta x}{(1 + e^{-x})} \tag{11}$$

The difference between Swish and E-Swish is the position of constant β . The similarities and differences between them shown above in Figure 2. Swish and E-Swish activation functions are equal when $\beta=1$. According to [15, 16] papers, they performed a good result in image classification and never examines before in regression. In this work we set $\beta=1$ for Swish activation function as the best achieved result in the original paper and $\beta=1.5$ in the E-Swish activation function for the same reason.

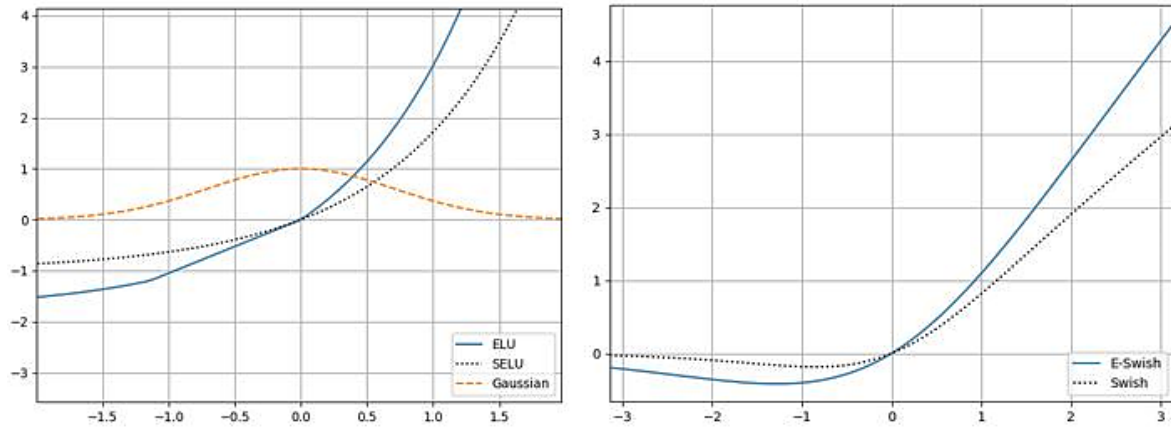


Figure. 2. ELU, SELU, gaussian, swish and E-swish activation functions

2.1. Dataset description

In this paper we try to cover different types of data sets which are related to electricity. The variety in the quality of the data set is used. Consumption, production, multivariate and univariate are considered. Also, some of them depends on date and time. Moreover, we consider the relationships between the input and output variables.

2.2. Combined cycle power plant

Combined cycle power plant data set consists of four input variables temperature, ambient pressure, relative humidity and exhaust vacuum. Those input variables are used to predict electrical energy output. The data set was collected over six years (2006-2011) and contains 9568 data points [17]. There is no missed data in this dataset. Figure 3 below shows the relationship between energy and its factors.

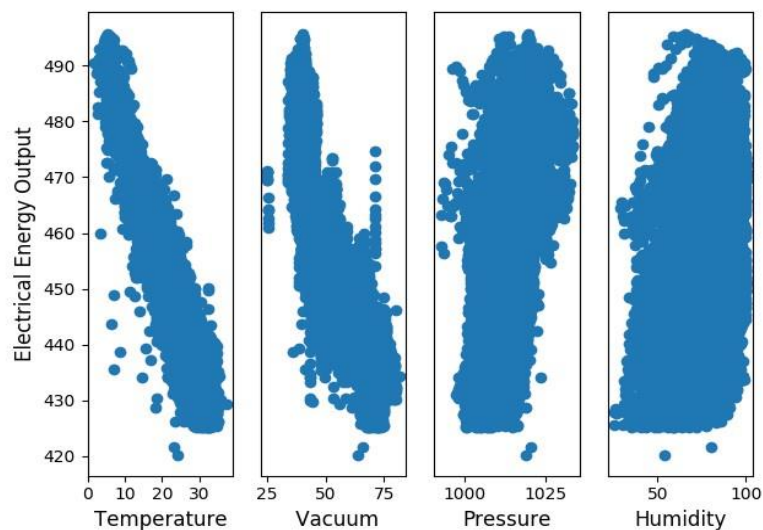


Figure 3. Scatter diagram of electrical energy output vs temperature, pressure, humidity and exhaust vacuum

2.3. Energy efficiency dataset

the dataset consist of eight input attributes relative compactness (comp), surface area (sa), wall Area (WA), roof area (RA), overall height (OH), orientation (Orie), glazing area (GA), glazing area distribution (GAD) which are the building design parameters. The responses outcomes are heating load (HL) and cooling load (CL) which are power consumption [18].

2.4. Appliances energy prediction dataset

It consists of 19735 instances of 28 attributes. Those attributes are related to temperature, humidity, pressure, wind speed, days of week and day status [19]. It does not containe any missing values.

2.5. Individual household electric power consumption dataset

This dataset is time series which measures an electric power consumption in one household for a single residential customer. It contains the consumption power record that is taken in each minute over a period of four years between December 2006 and November 2010. The original dataset contains nine attributes but we focused on the datetime as a time series and the household global minute-averaged active power. There are missing values of data and we fill them by using the values that come before [20]. Table 1 shows the properties of data set.

Table 1. Individual household electric power consumption properties

Property	Value
Count	2,075,259
Max	11.120
Min	0.076
Mean	1.086
Median	0.598
STD	1.053
Missing values	25,979

3. METHODOLOGY

We use neural network with different depth of hidden layers to examine the most accurate activation function. Many studies tried to define the number of hidden layers and the number of neurons in each hidden layer but we know that no exact role can define that. In this paper, we concentrate on the activation function and we do not focus on the structure of neural network. Therefore, we defined four models of multilayer neural networks with different depth. The first model is sample with one hidden layer and different number of neurons according to the four different datasets. We specified the number of neurons by function [2].

$$N_h = N + 1 \quad (12)$$

where N is the number of input data

The other models are defined with more hidden layers. Model 2 consists of two hidden layers with 30 and 20 neurons in each layer. The third is deeper than previous models with 9 hidden layers with 240, 200, 160, 120, 80, 60, 40, 30, 20 neurons in hidden layer respectively. The last model is designed with 11 hidden layers with 320,280,240, 200, 160, 120, 80, 60, 40, 30, 20 neurons. We tried to go deeper to determine the ability of activation function in deep neural network. We implement the stochastic gradient descent (SGD) [21] as the optimizer. The number of epochs is enough to reach the best solution. The initial learning rate is set to 0.1 and we reduced it by multiplying it by 0.2 at 40, 80, 120, 160 epochs.

The initialization of training is Glorot uniform initialization [8]. We utilized no dropout and 0.9 momentum. The parameters that used in the four models are equal for the all data sets except for the individual household electric power consumption data set because it has huge number of data and needs long time of execution. So, the number of epochs is reduced to 40 with Adam optimizer [22] as it considers faster than others. We use root mean squared error (RMSE) [23] which is one of quantitative approaches of model performace evaluation. It compares predicted values to the actual values

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{model,i})^2}{n}} \quad (13)$$

where X_{obs} is predicted values and X_{model} is actual values

To train the neural network efficiently [24], we perform data preprocessing to transform input data into better form. The transformation was used by Min-Max normalization [25] which is one of the most used techniques and it is accomplished by:

$$X = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (14)$$

4. EXPERIMENTS AND RESULT

In order to evaluate the activation function with the four models and datasets, all datasets divide into train set and test set. Each model trains by using 75% of the data and the rest for testing which are unseen values. The test set is usually used to gage the models but also training set shows the ability of learning. We compare the median of 9 runs in all execution models of datasets except for the Individual household electric power consumption dataset.

As shown in Table 2 that ReLU activation gets best result in model II, III and IV for training and testing set of combined cycle power plant data set. In addition, SELU gets best performance for training and testing set in model I. Table 3 shows that ReLU and Leak ReLU get the best performance in all models for energy efficiency dataset. The minimum train errors in model II, III, IV and the minimum test error in model II is gotten by ReLU. Leak ReLU achieves the minimum errors in the rest.

Table 2. RMSE comparison of activation function in 4 models for combined cycle power plant dataset

Activation	MODEL I		MODEL II		MODEL III		MODEL IV	
	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>
Sigmoid	4.421	4.619	4.501	4.707	4.182	4.365	4.125	4.310
Tanh	4.251	4.442	4.130	4.326	3.978	4.228	3.927	4.204
SoftSign	4.251	4.451	4.132	4.325	4.039	4.265	3.742	4.102
SoftPlus	4.303	4.404	4.236	4.419	4.150	4.323	4.155	4.340
ReLU	4.218	4.413	3.856	4.141	3.214	3.845	3.126	3.860
LReLU	4.212	4.409	3.915	4.173	3.438	3.939	3.383	3.943
Gaussian	4.223	4.394	4.153	4.328	17.07	17.04	17.07	17.04
ELU	4.383	4.451	4.158	4.337	4.055	4.273	3.949	4.200
SELU	4.187	4.383	4.138	4.334	3.753	4.091	3.809	4.200
Swish	4.239	4.424	4.233	4.413	4.071	4.273	3.945	4.187
E-Swish	4.239	4.426	4.214	4.397	3.991	4.207	3.936	4.186

Table 3. RMSE comparison of activation function in 4 models for energy efficiency (heating load) dataset

Activation	MODEL I		MODEL II		MODEL III		MODEL IV	
	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>
Sigmoid	2.927	3.140	2.928	3.127	2.675	3.530	2.357	3.977
Tanh	2.921	3.137	2.969	3.227	2.636	2.773	1.021	1.467
SoftSign	2.614	2.464	3.031	3.171	0.724	2.195	0.906	1.276
SoftPlus	2.860	2.984	2.724	2.599	1.990	2.863	1.021	1.531
ReLU	2.130	2.189	0.446	0.679	0.385	1.002	0.398	1.248
LReLU	2.021	2.057	0.652	0.783	0.462	0.918	0.451	1.044
Gaussian	2.457	3.235	2.797	3.425	2.476	2.717	2.665	3.506
ELU	2.752	2.669	2.766	2.809	1.797	2.035	1.612	2.190
SELU	2.265	2.400	2.857	2.991	0.477	1.165	0.439	1.062
Swish	2.734	2.651	2.656	3.085	0.885	1.548	0.808	1.074
E-Swish	2.725	2.639	2.704	2.729	1.370	1.759	0.833	1.135

The performance of ReLU and Leak ReLU also outperform the other activation functions. Table 3 and 4 present the prediction of the same dataset. The only difference is that Table 3 presents the prediction of heating load response variable and Table 4 shows the prediction of cooling load response variable.

The correlation among variables in appliances energy is obviously clear that the relationships between input and output of the appliances energy prediction dataset are weak. The error is large compared to other datasets. In the original paper [19], multiple linear regression model gets (RMSE=93.21) for training set and (RMSE=93.18) for testing set. It is seen in Table 5 below that SELU achieves best result in Model I. In addition, ReLU and Leak ReLU still outperform other activation functions in Model II, Model III and Model IV.

Table 4. RMSE comparison of activation function in 4 models for energy efficiency (cooling load) dataset

Activation	MODEL I		MODEL II		MODEL III		MODEL IV	
	Train	Test	Train	Test	Train	Test	Train	Test
Sigmoid	3.195	3.284	3.193	3.272	3.069	3.347	2.597	3.092
Tanh	3.189	3.276	3.243	3.460	2.492	2.945	1.197	2.152
SoftSign	3.037	3.006	3.036	3.077	2.794	3.058	0.758	2.060
SoftPlus	3.169	3.222	3.133	3.113	2.155	2.545	1.686	2.003
ReLU	2.698	2.789	1.009	1.362	0.404	1.149	0.396	1.429
LReLU	2.693	2.746	1.404	1.859	0.477	1.225	0.460	1.203
Gaussian	2.864	2.927	2.404	2.440	2.376	3.264	2.665	3.924
ELU	3.133	3.150	3.118	3.036	2.989	3.032	2.420	2.521
SELU	2.855	2.906	3.081	3.171	0.755	1.429	0.632	1.469
Swish	3.135	3.134	3.084	2.040	1.225	1.924	1.171	1.619
E-Swish	3.126	3.138	2.929	3.004	1.300	1.788	1.102	1.464

Table 5. RMSE comparison of activation function in 4 models for appliances energy prediction dataset

Activation	MODEL I		MODEL II		MODEL III		MODEL IV	
	Train	Test	Train	Test	Train	Test	Train	Test
Sigmoid	93.47	93.71	93.46	93.70	81.64	85.10	73.99	81.82
Tanh	93.46	93.69	86.92	88.81	59.60	76.52	54.75	74.60
SoftSign	89.27	90.11	81.66	85.20	50.54	73.43	44.36	72.71
SoftPlus	93.43	93.67	92.20	92.60	69.92	81.43	63.81	78.63
ReLU	84.77	86.82	77.44	82.42	37.77	72.86	36.56	72.26
LReLU	84.57	86.62	74.93	81.66	43.47	72.27	39.27	71.31
Gaussian	85.37	87.20	78.71	83.09	57.41	77.08	70.97	80.88
ELU	92.68	92.99	86.53	87.80	63.91	76.92	58.28	76.34
SELU	84.50	86.50	79.85	83.82	52.39	74.68	48.03	73.99
Swish	91.10	91.67	88.01	88.95	57.53	75.69	50.22	73.82
E-Swish	91.51	91.97	86.79	88.33	57.31	77.42	50.03	74.51

Individual household electric power consumption dataset needs long time to execute due to its huge size. Therefore, we reduced the epoch number to 40 instead of 200. In addition, the number of implementations is reduced to three instead of nine. Moreover, we just tested Model I and Model II due to the same reason. Table 6 shows that ReLU Activation function outperforms its counterparts. Table 7 summarizes the results of all experiments. It shows that ReLU and LReLU outperform the other activation functions. In addition, we can observe that ReLU learn in train data better than others do.

Table 6. RMSE comparison of activation function in 4 models for individual household electric power consumption dataset

Activation	MODEL I		MODEL II	
	Train	Test	Train	Test
Sigmoid	0.277	0.228	0.544	0.517
Tanh	0.298	0.250	0.275	0.225
SoftSign	0.283	0.234	0.623	0.606
SoftPlus	0.278	0.228	0.346	0.306
ReLU	0.276	0.225	0.274	0.224
LReLU	0.278	0.227	0.278	0.227
Gaussian	0.285	0.237	0.285	0.237
ELU	0.279	0.230	0.279	0.231
SELU	0.287	0.239	0.285	0.236
Swish	0.292	0.244	0.282	0.234
E-Swish	0.321	0.278	0.334	0.296

Table 7. Best performance of activation functions for all data sets

Dataset	Train	Test
Combined cycle power plant	ReLU	ReLU
Energy efficiency (Heating Load)	ReLU	ReLU
Energy efficiency (Cooling Load)	ReLU	ReLU
Appliances energy prediction	ReLU	Leak ReLU
Individual household electric power consumption	ReLU	ReLU

5. CONCLUSION

Prediction of power consumption and production is an important step to manage and control power utilities. This prediction helps the utilities to figure the quantity of consumed and produced electricity. There are many models designed for that purpose. Nowadays, artificial intelligence models such as neural networks (Feedforward neural network) proved its efficiency. Building strong model for prediction using multilayer neural network needs several important things one of them is choosing suitable activation function. We compare the most used activation functions on data sets related to electrical power. We present that ReLU and Leak ReLU outperform other activation function. ReLU activation achieves 13 over 18 of train set and 8 over 18 of testing sets and Leak ReLU achieves 3 over 18 in training sets and 8 over 18 in testing sets. ReLU achieves the smallest error in training set in all datasets. ReLU in combined cycle power plant, energy efficiency (Heating Load) and energy efficiency (Cooling Load) and individual household electric power consumption datasets performs the minimum RMSE in test set. In the other hand, Leak ReLU achieves the minimum test set error in appliances energy prediction dataset.

REFERENCES

- [1] P. Chujai, N. Kerdprasop, and K. Kerdprasop, "Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models," *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, pp. 295-300, 2013.
- [2] K. G. Sheela and S. N. Deepa, "Review on Methods to Fix Number of Hidden Neurons in Neural Networks," *Mathematical Problems in Engineering*, vol. 2013, pp. 1-11, 2013.
- [3] A. Salam and A. El Hibaoui, "Comparison of machine learning algorithms for the power consumption prediction - Case study of tetouan city," *6th International Renewable and Sustainable Energy Conference*, pp. 1-5, 2018.
- [4] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *J. Pharm. Biomed. Anal.*, vol. 22, no. 5, pp. 717-727, 2000.
- [5] B. Ding, H. Qian, and J. Zhou, "Activation functions and their characteristics in deep neural networks," *Proceedings of the 30th Chinese Control and Decision Conference*, pp. 1836-1841, 2018.
- [6] C. Özkan and F. Sunar Erbek, "A Comparison of Activation Functions for Multispectral Landsat TM Image Classification," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 11, pp. 1225-1234, 2003.
- [7] A. A. Minai and R. D. Williams, "On the derivatives of the sigmoid," *Neural Networks*, vol. 6, no. 6, pp. 845-853, 1993.
- [8] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249-256, 2010.
- [9] R. H. R. Hahnloser, et al., "Erratum: Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature*, vol. 405, no. 6789, pp. 947-951, 2000.
- [10] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," *arXiv preprint arXiv:1505.00853*, 2015.
- [11] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *ICML*, 2010.
- [12] S. Lee and R. M. Kil, "A Gaussian potential function network with hierarchically self-organizing learning," *Neural Networks*, vol. 4, no. 2, pp. 207-224, 1991.
- [13] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUS)," *arXiv preprint arXiv:1511.07289*, 2015.
- [14] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks," *Advances in neural information processing systems*, pp. 971-980, 2017.
- [15] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," *arXiv preprint arXiv:1710.05941*, 2017.
- [16] E. Alcaide and Eric, "E-swish: Adjusting Activations to Different Network Depths," *arXiv preprint arXiv:1801.07145*, 2018.
- [17] L. X. Niu and X. J. Liu, "Multivariable generalized predictive scheme for gas turbine control in combined cycle power plant," *IEEE Conference on Cybernetics and Intelligent Systems*, pp. 791-796, 2008.
- [18] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools," *Energy Build.*, vol. 49, pp. 560-567, 2012.
- [19] L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," *Energy Build.*, vol. 140, pp. 81-97, 2017.
- [20] N. Farag Abed and M. MMilosavljevic, "Single Home Electricity Power Consumption Forecast Using Neural Networks Model," *IJSET-International J. Innov. Sci. Eng. Technol.*, vol. 3, no. 1, pp. 100-106, 2016.
- [21] T. Zhang and Tong, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," *Twenty-first international conference on Machine learning - ICML '04*, 2004.
- [22] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [23] T. Chai, T. Chai, and R. R. Draxler, "Ozone health and ecosystem impacts View project Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature," *Geosci. Model Dev*, vol. 7, pp. 1247-1250, 2014.
- [24] A. Famili, W. M. Shen, R. Weber, and E. Simoudis, "Data preprocessing and intelligent data analysis," *Intell. Data Anal.*, vol. 1, no. 1, pp. 3-23, Jan. 1997.
- [25] .T. Jayalakshmi and A. Santhakumaran, "Statistical Normalization and Back Propagation for Classification," *International Journal of Computer Theory and Engineering*, vol. 3, no. 1, pp. 1793-8201, 2011