Cultural Algorithm Initializes Weights of Neural Network Model for Annual Electricity Consumption Prediction

Gawalee Phatai[1], Sirapat Chiewchanwattana[2*], Khamron Sunat[3]
Department of Computer Science
Faculty of Science, Khon Kaen University
Khon Kaen, Thailand

***Abstract*—The accurate prediction of annual electricity consumption is crucial in managing energy operations. The neural network (NN) has achieved a lot of achievements in yearly electricity consumption prediction due to its universal approximation property. However, the well-known back-propagation (BP) algorithms for training NN has easily got stuck in local optima. In this paper, we study the weights initialization of NN for the prediction of annual electricity consumption using the Cultural algorithm (CA), and the proposed algorithm is named as NN-CA. The NN-CA was compared to the weights initialization using the other six metaheuristic algorithms as well as the BP. The experiments were conducted on the annual electricity consumption datasets taken from 21 countries. The experimental results showed that the proposed NN-CA achieved more productive and better prediction accuracy than other competitors. This result indicates the possible consequences of the proposed NN-CA in the application of annual electricity consumption prediction.**

Keywords*—Neural network; weights initialization; metaheuristic algorithm; cultural algorithm; annual electricity consumption prediction

I. Introduction

Electricity is a major driving force for economic development in many countries. The overall demand for power increases continuously, even more, prominent in the future.

APEC is the acronym of the Asia-Pacific Economic Cooperation that is a cooperative economic group in the Asia-Pacific region. The high growth rates in recent decades of APEC results in a significant increase in electricity consumption. APEC energy data has proved essential in tracking energy consumption, reduction, and in determining the group's renewable energy goals. APEC is committed to improving efficient energy technologies; by setting targets and action plans, thereby creating the necessity to predict future electricity consumption usage accurately.

The artificial neural network (ANN) computation is based on the learning process of human perception and the function of the brain's nervous system, which has been widely applied to various problems in classification, pattern recognition, regression, and prediction. In general, humans have learning processes in which processes are characterized by pattern recognition. The pattern-based learning method is described as follows: People observe unknown objects and perceive their identities as distinct from others, especially when viewed more often and in different ways, which results in learning and memory. The human brain contains numerous processing units linked by several nervous systems that perform rapid analysis and decision making. The artificial neural network represents a simulation of the human brain [1][2]. Many studies regarding ANNs have been conducted for solutions in various disciplines.

A. Background

ANN is a distributed data processing system consisting of several simple calculation elements working together through a weighted connection. This calculation architecture was inspired by the human brain, which can learn intricate data patterns and classify them into general data. ANN can be categorized into several types according to not only instructional and unattended learning methods but also feedback-recall architectures.

ANN's most commonly used architecture is the multilayer neural perceptron (MLP). The weights of MLP can be adjusted using both the gradient-based process and the stochastic-based process. The original gradient-based supervised training algorithm of MLP is the error back-propagation (BP) algorithm [3]. BP and its variants are the most frequently used neural network techniques for classification and prediction [4][5].

However, the gradient-based method has two significant disadvantages: slow convergence speed and being trapped at a local minimum easily because of having a high dependency on the initial parameters (weights) [6][7]. Metaheuristic algorithms can overcome those disadvantages of the gradient-based algorithms. Algorithms of this kind use randomization-based techniques to perform the exploration and exploitation searches [8], which are capable of generating solutions to complex real-life problems that gradient-based methods are unable to solve [9]. The population-based structure is the most efficient and commonly used architecture in metaheuristic algorithms. The two often used categories of metaheuristic algorithms are evolutionary and swarm intelligence algorithms [10][11].

Metaheuristic algorithms were applied as supervised training algorithms of MLPs. For a given problem (input and target values), both the structure and weights of an MLP can be optimized. In this paper, we focus on selecting proper initial values of the connecting weights in an MLP network. A

*Corresponding Author

metaheuristic algorithm will perform the initial weights selection. The existing metaheuristics that used to train MLPs for the annual electricity consumption prediction included the Artificial Bee Colony (ABC) [12][13], Teaching-Learning-Based Optimization (TLBO) [13], Harmony Search (HS) [14] and Jaya Algorithm (JA) [15]. Techniques from prior studies found that the applied ANN model (ANN-TLBO), optimized by the TLBO algorithm to predict electric energy demand outperformed the ANN-BP and ANN-ABC models [13]; In other studies conducted to predict the electricity consumption of the ANN-TLBO in comparison with the ANN-BP, ANN-ABC, ANN-HS, ANN-TLBO, and ANN-JA models; the ANN-TLBO yielded better efficiency than that of the other models [15]. TLBO algorithm itself is two phases algorithm; a teacher phase and a learner phase [16].

Not Free Lunch Theorem (NFL) said that there is no superior optimization algorithm for all optimization problems [17]. Although a variety of evolution-based algorithms have been implemented and examined in the literature for MLP training, recognizing that the question of reaching local minima still exists. The Cultural Algorithm (CA) is very similar to the TLBO because it is also a two-stage algorithm; the population level and the belief space level [18]. This characteristic might lead to a more efficient in the initial weights selection. Therefore, we propose, herein, a new MLP training method based on the CA, in which to develop a single hidden layer neural network for annual electricity consumption prediction.

II. Methodology

A. Multilayer Perceptron for Neural Model Training

MLP is a widely used type of feedforward neural network having a multi-layered structure for complex tasks. There are several layers, namely the input layer, hidden layers, and the output layer. Each layer of MLP comprises of numerous neurons and the connecting weights between the two consecutive layers. The connecting weights are represented by real numbers in [−1, 1]. The input layer is responsible for receiving information for the neural network and sending it to the first hidden layer through the connecting weights. Each hidden layer will contain a layer that is responsible for receiving information for the neural network and sending it to the hidden layer. In an MLP fully interconnected by weights, each neuron of the hidden layer contains summation and activation functions. The weighted summation of input is described in Eq. (1), where I_i is the input variable i, and ω_{ij} is the connection weight between I_i and the hidden neuron j. An activation function is used to trig the output of neurons based on the value of the summation function. The Sigmoid function is most often applied. However, different types of activation functions may be utilized in the MLP.

Each node of the hidden layer calculates its output by Eq. (2). The production of the node j in the hidden layer is described in Eq. (2.) [19].

$$S_j = \sum_{i=1}^{n} \omega_{ij} I_i + \beta_j \tag{1}$$

$$f_j(x) = \frac{1}{1+e^{-S_j}} \tag{2}$$

The outcomes of the lower hidden layer are fed to the adjacent layer. Once all neurons in the last hidden layer produce results, the production of the network will be obtained by Eq. (3).

$$\hat{y}_k = \sum_{i=1}^{m} W_{kj} f_i + \beta_k \tag{3}$$

The initialization of the weights of a neural network is one of the essential problems, as network initialization can speed up the learning process. Zero initialization [20] and Random initialization [21] are generally practiced techniques used to initialize the parameters. Traditionally, the weights of a neural network are set to small random numbers.

B. Cultural Algorithm

Cultural algorithms (CA) is a kind of evolutionary algorithms; it is first presented by R. G. Reynolds [18]. Their computational models are based on principles of human social Cultural evolution that make practical use of the learning process through various agent-based techniques based on experience and knowledge gained over time. The cultural process allows for improved efficiency in finding the optimal solution within a search space and making it easier to find the optimal global solution. The cultural changes within an optimization problem model represent information transmitted within and between populations. The main principle of the CA is to preserve socially accepted beliefs and discard unacceptable beliefs.

The CA can be divided into two main components as a population space and a belief space. Each member of the former part is evaluated through a performance function and may be carried out by an Evolutionary Algorithm (EA). An acceptance function then determines which individuals are to impact the belief space. At each generation, the knowledge acquired in the population search (e.g., the population’s best solution) will be memorized in the belief space [22]. The interaction and help between the two spaces are similar to the evolution of human culture [23]. The significant components of CA are shown in Fig. 1.

The CA uses a dual evolutionary mechanism, while lower-level populations help periodically enter the top level of beliefs. On the other hand, a high level of belief will evolve these elite people to influence the lower communities [25]. This mechanism results in the improvement of the population diversity and the convergence characteristics, accordingly. The interested reader can see [18] for more details of CA.

C. Cultural Algorithm for Training Neural Network Model

We propose CA as a training algorithm of the Neural Network model. CA will find a proper set of the initial weights for an MLP, and from now on, we call the proposed algorithm as NN-CA. It can be applied not only for a single hidden but also several hidden layers. Two main aspects must be considered when the approach is used: the representation of the weights as the search agent of the CA; and the selection of the fitness function.

The representation is straightforward, as all the weights of an MLP are organized and indexed to be a row vector. This vector is a search agent of CA. The fitness function will be explained after the presentation of the workflow.

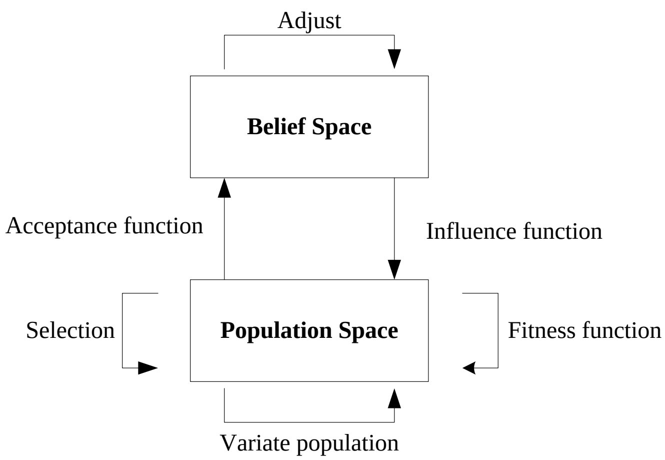


Fig. 1. Flowchart Diagram of Cultural Algorithm [24].

The general steps of the proposed NN-CA are depicted in Fig. 2.

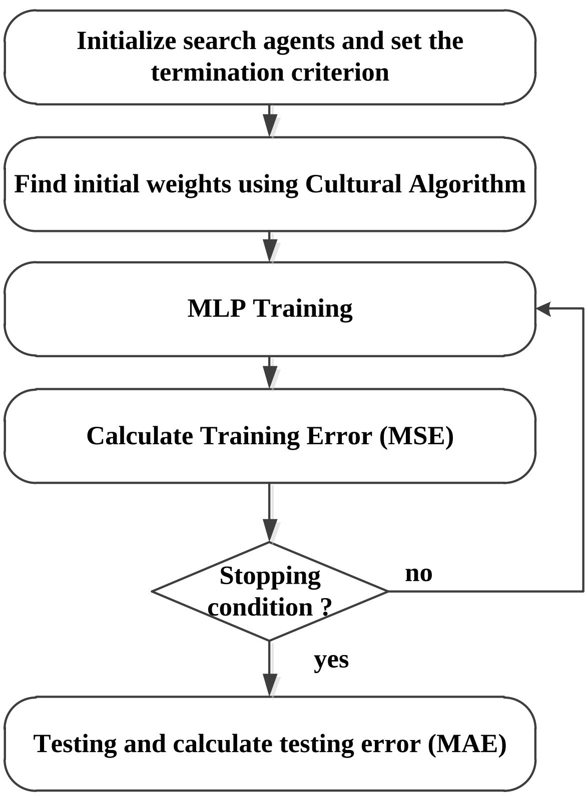


Fig. 2. General Steps of the NN-CA.

The workflow of the CA approach applied to train the neural network model may be described in the following steps:

1) Initialization: the search agents in the population and belief spaces are randomly generated for training. Each search agent in a belief space represents a possible MLP. Each dataset is separated as the training part and the testing part.

2) Fitness evaluation: Each possible MLP is evaluated its quality through a fitness function. All the weights of a search agent of belief space are mapped to an MLP, and then each MLP is assessed by the selected fitness function. Typically, the Mean Squared Error (MSE), which is dependent on the neural network training model and the problem of interest, is selected to perform.

3) Update the accepted population in the belief space.

4) Steps 2 to 3 are repeated until the terminated condition is found.

5) The reliability evaluation of the neural network model that has the lowest MSE value will be conducted on the testing part of the dataset to determine the Mean Absolute Error (MAE).

The MSE, which is the average of the error-squared for all training samples, as shown Eq. (4), acts as the fitness function. It depends on the difference between each actual (or the target) its associated output values of the MLP.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(EC_o - EC_p\right)^2 \quad (4)$$

The Mean Absolute Error (MAE) that evaluates the reliability of each model is shown in Eq. (5).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(EC_o - EC_p\right)^2\right|, \quad (5)$$

where EC_P is annual electricity consumption value produced from MLP and EC_O is the actual annual electricity consumption value.

III. Experimental Results

The experiments aimed to examine the effectiveness of the proposed method for the annual electricity consumption prediction. The neuron network model used was a single hidden layer MLP. All the experiments are programmed in MATLAB, and ran on Intel 2.9 GHz, 8 GB memory. The operating system is Windows 10.

Our study utilized the Asia-Pacific Cooperation (APEC) energy database, which contained the annual electricity statuses of 21 countries in the Asia-Pacific region. There are four input variables: Population (million person), GDP (billion US$), imports (billion US$), and exports (billion US$) were independent variables in model annual electricity consumption (TWh).

Data were divided into two parts: training data (1990 to 2008) and testing data (2009 to 2017); which consisted of population, GDP, imports, and exports data from the World Bank [26]; and annual electricity consumption data, from the Expert Group on Energy Data and Analysis (EGEDA) [27]. Annual Electricity consumption target data are shown in Fig. 3.

The Pearson correlation coefficient (R) was applied to examine the dependency between each input variable and annual electricity consumption. All related R-values are shown in Table I.

From Table I, the GDP of Russia has a relatively low R-value (R < 0.5). That means the annual electricity consumption in Russia does not maintain a linear relationship with its GDP parameters.

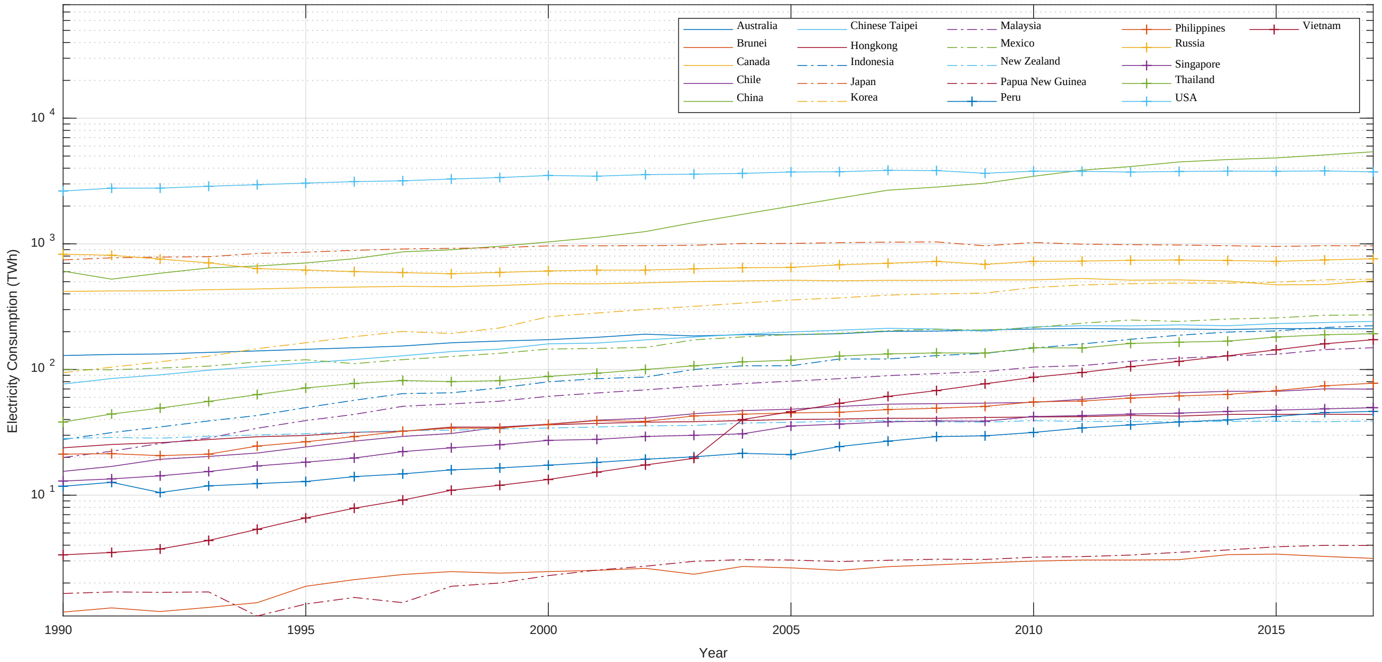


Fig. 3. Variation of annual Electricity Consumption in various Datasets.

TABLE I. EXAMINATION OF THE RELATIONSHIPS (*R*) BETWEEN INPUT PARAMETERS AND OUTPUT OF NEURAL NETWORK MODEL

Country	Population	GDP	Import	Export
Australia	0.9379	0.8367	0.8591	0.8451
Brunei	0.9587	0.7768	0.7094	0.7292
Canada	0.7980	0.7971	0.8242	0.8836
Chile	0.9962	0.9196	0.9033	0.9092
China	0.8984	0.9853	0.9893	0.9919
Chinese Taipei	0.9936	0.9435	0.9702	0.9724
Hongkong	0.9886	0.8971	0.8615	0.8751
Indonesia	0.9892	0.9391	0.9174	0.9310
Japan	0.9839	0.5534	0.6855	0.7611
Korea	0.9880	0.9618	0.9385	0.9512
Malaysia	0.9922	0.9443	0.9423	0.9346
Mexico	0.9890	0.9601	0.9889	0.9878
New Zealand	0.9392	0.8402	0.8756	0.8671
Papua New Guinea	0.9462	0.7558	0.6475	0.7930
Peru	0.9288	0.9722	0.9610	0.9439
Philippines	0.9902	0.9414	0.9788	0.9759
Russia	-0.2056	0.5540	0.5528	0.5117
Singapore	0.9867	0.9327	0.9437	0.9487
Thailand	0.9507	0.9287	0.9208	0.9585
USA	0.9458	0.9179	0.9205	0.8443
Vietnam	0.9088	0.9932	0.9939	0.9881

Based on the electric consumption data we studied, the size of MLP is 4:*h*:1, where *h* is the number of neurons in the hidden layer. Because the prediction accuracy depends on the MLP size or *h*, we compare two strategies to study the effect *h*. The first strategy *h* was assigned as 2×N +1, where N is the dimension of dataset features or the dataset features [19]. The second strategy appointed the number of hidden neurons to be 5, 10, 15, and 20 [13]. There are some predefined settings: all BP experiments were executed with 5,000 iterations, each metaheuristic algorithm evolved 5,000 iterations, the population size of CA is 50, the MLPs weights must be in the interval of [-1, 1].

A. Comparing the Results of Neural Network Models

The proposed NN-CA was compared with MLP trained by the error back-propagation algorithm (which is label as BP), as well as other metaheuristic algorithm trainers, based on the MSE evaluation measures. Input, hidden layer neurons, and output variables were assigned before starting the experiment. From Table I, there are four input variables: population, GDP, imports, and exports. To determine the suitable network architecture, the BPs were trained with a single hidden layer, incorporating nine hidden nodes that specified by the first strategy, and 5, 10, 15, 20 hidden nodes as determined by the second strategy.

Table II presents the average ranks, Friedman test [28], produced by each competitor, where a lower score is better. The significant differences do exist between the six algorithms. As seen from Table II, NN-CA with a 4-20-1 architecture produced the best overall ranking in comparison with other algorithms, which shows the merits of the proposed NN-CA.

The annual electricity consumption variable was provided in the output data. The overall results that the 4-20-1 architecture of the neural network model was the most superior.

We can see that NN-CA outperforms the other algorithms in all the results of the Friedman rank test with an even lower number of hidden neurons, such as the 4-5-1 neural network model architecture.

The overall results, the 4-20-1 architecture of the neural network model and the 4-5-1 neural network model architecture presented in Tables III and IV.

As demonstrated in each table above, the proposed NN-CA outperformed all other training optimizers and BP. CA can select a proper search agent to be the initial weights of an MLP. Each algorithm within each dataset was statistically compared via the Friedman test. This comparison confirmed the significance, and contrast, of the NN-CA's ability with that of the other trainers.

Fig. 4 shows the MSE convergence curves of the 4-20-1 neural network model architecture utilized in the prediction datasets and trained by ABC, CA, HS, JA, and TLBO.

TABLE II. RESULTS OF THE FRIEDMAN RANK TEST

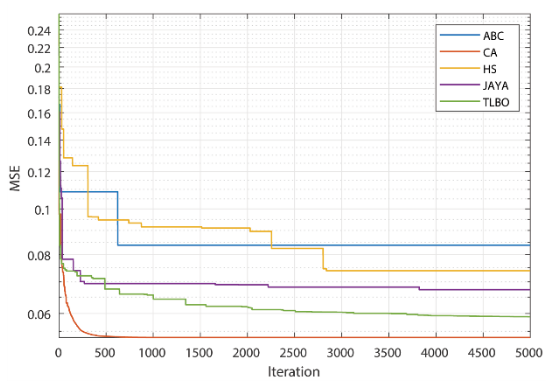
Algorithm	Ranking				
	4-5-1	4-10-1	4-15-1	4-20-1	4-9-1
BP	6.00	6.00	6.00	5.95	6.00
ABC	4.24	4.24	4.24	4.71	4.19
CA	1.31	1.31	1.31	**1.29**	1.43
HS	4.48	4.48	4.48	4.24	4.48
JA	3.24	3.24	3.24	3.05	3.24
TLBO	1.74	1.74	1.74	1.76	1.67

TABLE III. MSE AND MAE VALUES OF DIFFERENT METAHEURISTIC ALGORITHMS WITHIN THE 4-20-1 NEURAL NETWORK MODEL ARCHITECTURE

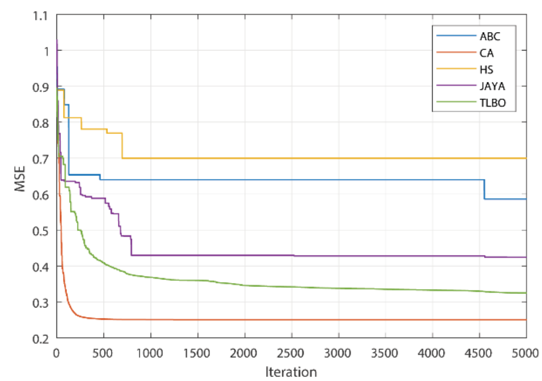
Country	**MSE from Training**						**MAE from Testing**					
	BP	**ABC**	**CA**	**HS**	**JAYA**	**TLBO**	**BP**	**ABC**	**CA**	**HS**	**JAYA**	**TLBO**
Australia	1.8558	0.0838	**0.0533**	0.0739	0.0674	0.0590	10.5534	7.1501	5.1081	6.3893	5.8884	5.2598
Brunei	1.0547	0.5863	**0.2504**	0.6996	0.4240	0.3247	0.4232	0.3971	0.2488	0.4317	0.3233	0.2947
Canada	1.4451	0.3669	**0.0847**	0.4876	0.4062	0.2450	23.3678	16.4532	7.0591	20.9438	18.6190	14.0389
Chile	0.6715	0.1843	**0.0508**	0.1698	0.1135	0.0825	5.5987	5.6646	2.8245	5.4038	4.4637	3.6630
China	4.8680	0.1330	0.0113	0.0398	0.0380	**0.0068**	28.8633	48.9373	10.6728	27.9172	21.3205	6.6458
Chinese Taipei	1.1856	0.0676	0.0185	0.0354	0.0272	**0.0142**	9.4288	11.2984	6.0540	8.0929	7.4067	5.2358
Hongkong	0.7192	0.2669	**0.1195**	0.2321	0.2053	0.1332	2.7506	2.4894	1.4485	2.3302	2.2148	1.4924
Indonesia	1.5123	0.3734	0.0372	0.1236	0.0871	**0.0260**	30.3337	24.8620	6.8910	14.2380	13.2884	6.8579
Japan	2.5163	0.2062	**0.0900**	0.2037	0.1085	0.0941	37.5703	29.8761	17.5061	30.1001	21.8478	19.2708
Korea	2.3068	0.0762	**0.0232**	0.0808	0.0564	0.0358	52.9777	28.9537	16.4857	28.4239	25.5435	18.3189
Malaysia	1.0742	0.3935	0.1245	0.4744	0.2740	**0.1068**	19.8689	18.1442	8.5110	18.9106	14.0960	8.3304
Mexico	2.1067	0.1150	**0.0240**	0.0446	0.0271	0.0249	26.3023	13.8243	5.2876	8.9366	6.4515	5.5591
New Zealand	0.5732	0.2191	**0.1490**	0.2559	0.2225	0.1897	1.7637	1.5507	0.9873	1.4720	1.2715	1.1473
Papua New Guinea	0.4207	0.3025	0.0661	0.2260	0.1488	**0.0640**	0.4476	0.4624	0.1658	0.3498	0.2745	0.1551
Peru	0.7525	0.2076	**0.0241**	0.1279	0.0577	0.0356	7.2074	3.3716	1.3156	3.1291	1.8440	1.1969
Philippines	1.0347	0.2377	**0.0513**	0.1272	0.0981	0.0541	9.4401	5.7589	2.4693	4.5334	3.8628	2.5846
Russia	2.8502	0.4656	**0.1939**	0.4462	0.3205	0.2265	34.3029	33.5951	17.9250	34.1849	27.1204	20.6213
Singapore	0.7203	0.2771	**0.0572**	0.2237	0.1497	0.1131	5.8382	4.9481	2.0600	4.3414	3.4922	2.9320
Thailand	1.3158	0.2126	**0.0756**	0.1322	0.0815	0.0834	19.6150	15.2629	8.3153	11.7838	9.3112	8.0198
USA	5.2565	0.0643	**0.0137**	0.0394	0.0281	0.0161	20.6635	14.4222	6.9762	10.8600	10.8860	7.4154
Vietnam	1.3056	0.1194	0.0062	0.0431	0.0243	**0.0034**	33.5102	13.7530	2.6370	8.0160	5.9483	2.0628

TABLE IV. MSE AND MAE VALUES OF DIFFERENT METAHEURISTIC ALGORITHMS WITHIN THE 4-5-1 NEURAL NETWORK MODEL ARCHITECTURE

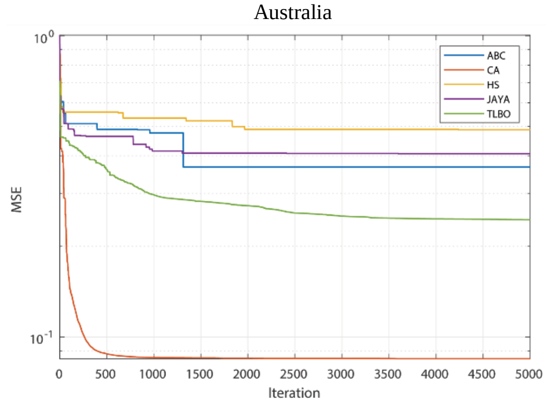
Country	MSE from Training						MAE from Testing					
	BP	**ABC**	**CA**	**HS**	**JAYA**	**TLBO**	**BP**	**ABC**	**CA**	**HS**	**JAYA**	**TLBO**
Australia	1.5729	0.0658	**0.0461**	0.0672	0.0669	0.0582	8.5458	7.0993	4.9810	6.2592	5.9593	5.3299
Brunei	0.9042	0.5360	**0.3201**	0.6360	0.5469	0.5052	0.4258	0.4001	0.3009	0.4179	0.3795	0.3654
Canada	1.3952	0.3125	**0.0959**	0.4858	0.4712	0.4367	20.4912	13.9783	8.2325	20.4358	20.0953	19.7515
Chile	0.6546	0.1033	**0.0539**	0.1341	0.1226	0.0933	4.9831	4.5533	2.9904	4.7539	4.4267	4.0753
China	4.6984	0.0359	0.0135	0.0170	0.0212	**0.0061**	26.1438	22.6626	9.9075	12.0142	16.7640	6.7927
Chinese Taipei	1.1471	0.0446	0.0211	0.0349	0.0294	**0.0175**	9.7913	7.8809	6.3223	8.3315	7.3382	5.8382
Hongkong	0.6731	0.1734	**0.1276**	0.2009	0.1944	0.1583	2.7399	2.1919	1.6193	2.0849	1.9471	1.4592
Indonesia	1.3943	0.1177	0.0393	0.0575	0.0612	**0.0351**	36.7804	16.6837	8.8297	9.4206	10.1301	7.4799
Japan	2.5308	0.1097	**0.0951**	0.1701	0.1334	0.1087	37.7423	24.2206	20.1802	26.4450	24.2393	21.5002
Korea	1.9595	0.0784	**0.0398**	0.0706	0.0674	0.0483	58.3515	29.5178	20.0582	25.8154	26.5983	22.0616
Malaysia	0.9710	0.2567	**0.1556**	0.3388	0.2905	0.2466	19.5628	14.4917	9.6247	15.1668	15.3101	12.4294
Mexico	1.8606	0.0411	0.0252	0.0280	0.0273	**0.0224**	26.0858	8.2147	5.8130	6.3860	6.3732	5.0313
New Zealand	0.5570	0.1702	**0.1597**	0.2420	0.2261	0.2147	1.5785	1.5426	1.1897	1.4519	1.3562	1.2779
Papua New Guinea	0.2874	0.1446	**0.0891**	0.1318	0.1358	0.0911	0.3320	0.2438	0.2016	0.2531	0.2533	0.1908
Peru	0.7471	0.0977	**0.0392**	0.0851	0.0585	0.0435	6.6411	2.5816	1.3081	2.4125	1.7299	1.4332
Philippines	0.9596	0.1122	0.0627	0.0957	0.0750	**0.0572**	10.0227	4.2566	2.5691	3.7190	3.1442	2.5663
Russia	2.7264	0.2626	**0.1577**	0.3593	0.3216	0.2544	30.1140	26.9641	17.0105	26.5047	24.4000	21.4613
Singapore	0.6790	0.1926	**0.0818**	0.1602	0.1598	0.1235	6.3694	4.0392	2.6527	3.6022	3.6631	3.1494
Thailand	1.2956	0.1195	**0.0841**	0.1008	0.0975	0.0866	19.6608	10.7466	8.3404	9.4941	9.2509	8.2743
USA	5.0420	0.0343	**0.0198**	0.0261	0.0306	0.0201	19.6188	11.7549	8.5884	10.1170	10.1933	8.0254
Vietnam	1.3203	0.0279	0.0054	0.0146	0.0132	**0.0035**	47.5357	6.5294	2.3461	4.5924	4.5872	2.1222



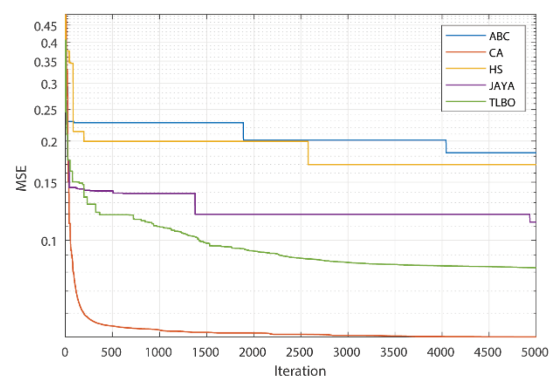
Australia



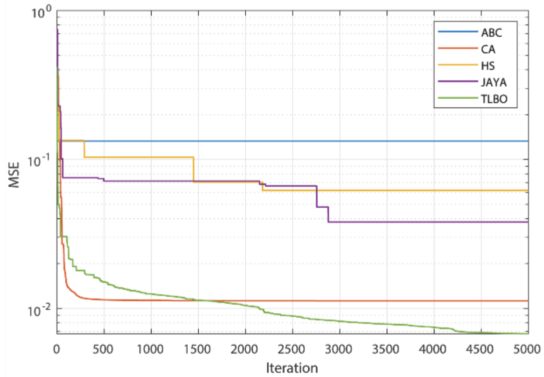
Brunei



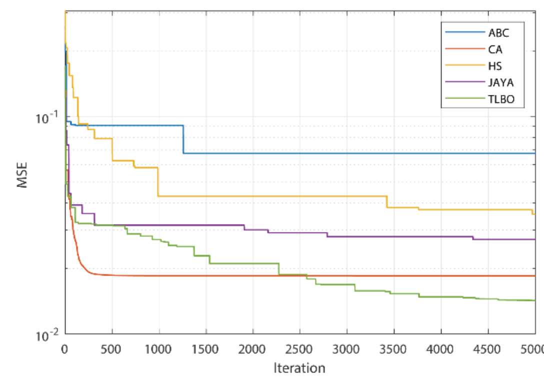
Canada



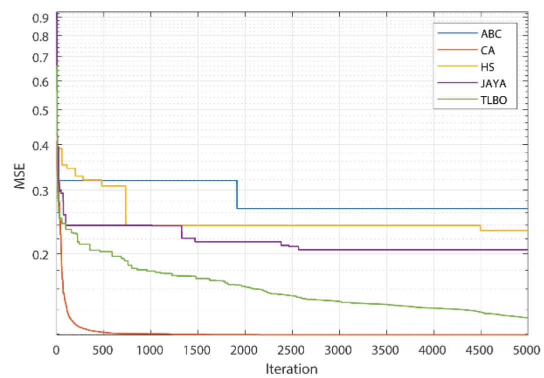
Chile



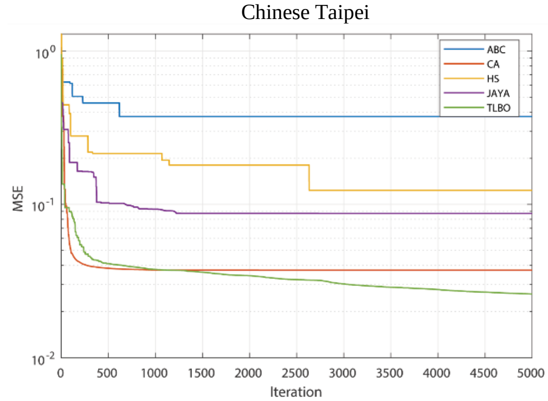
China



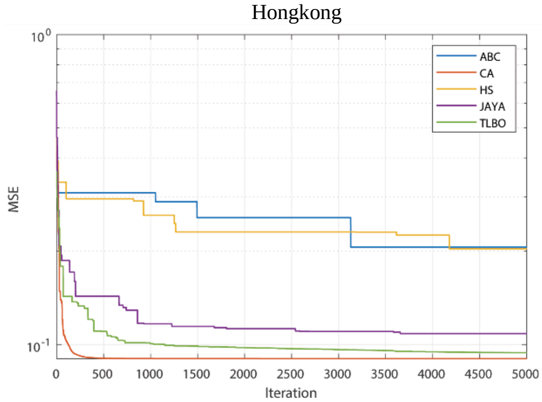
Chinese Taipei



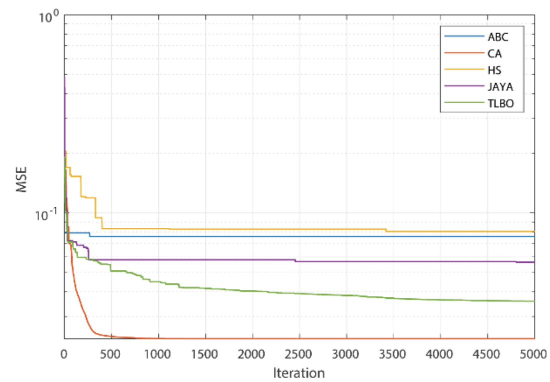
Hongkong



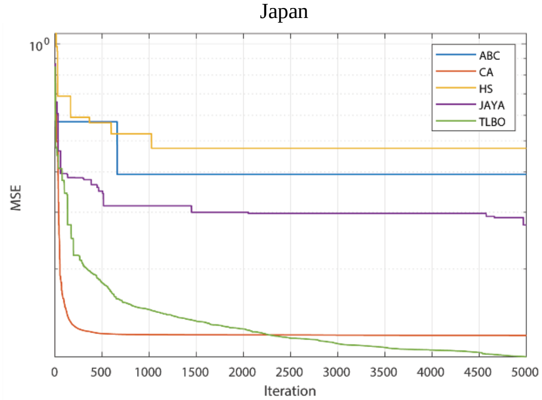
Indonesia



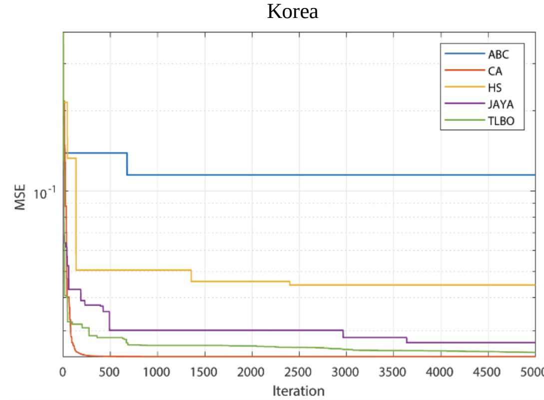
Japan



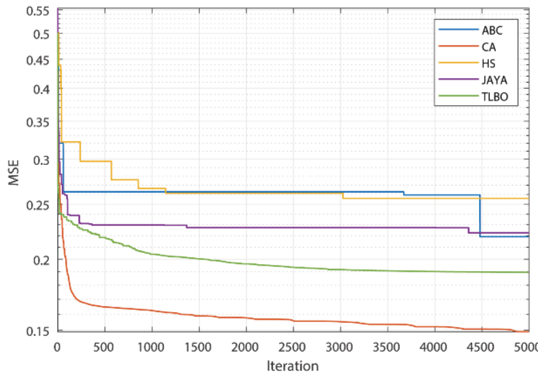
Korea



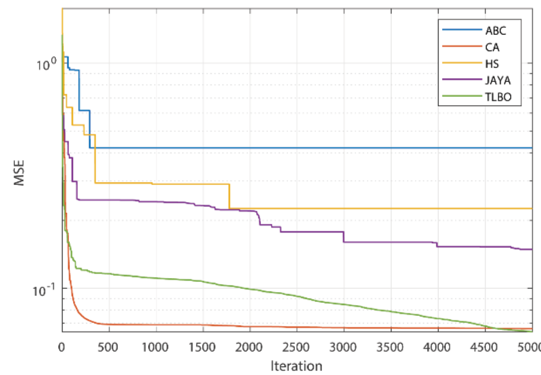
Malaysia



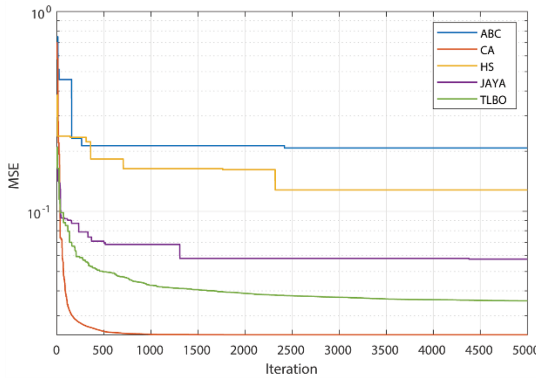
Mexico



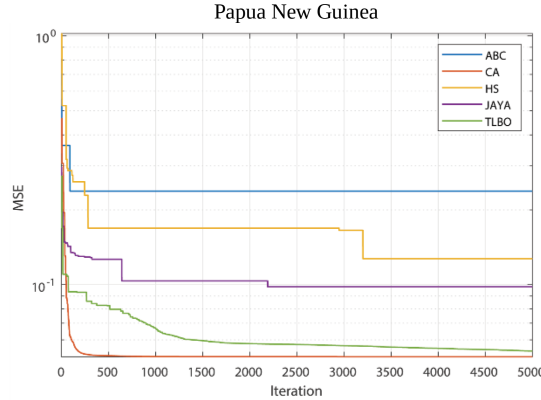
New Zealand



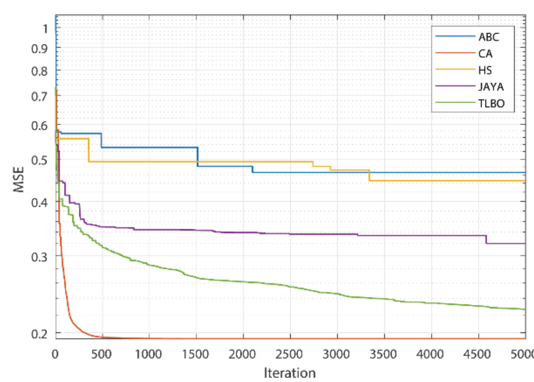
Papua New Guinea



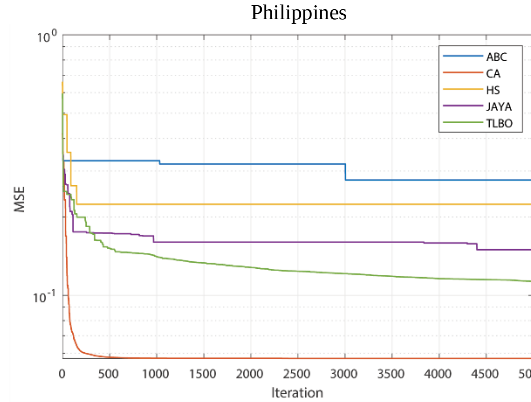
Peru



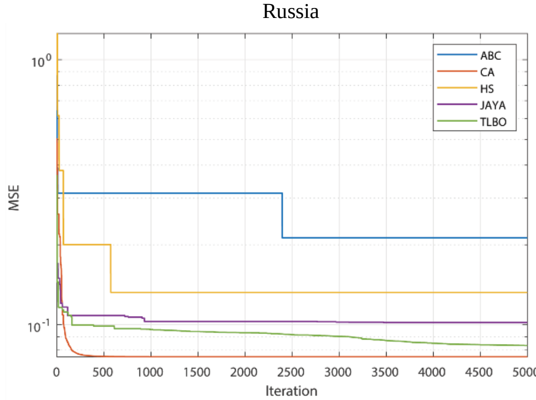
Philippines



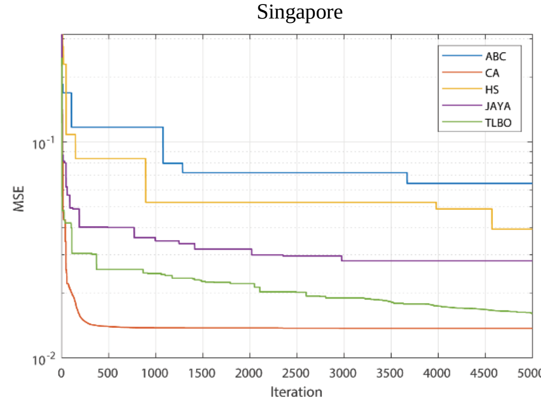
Russia



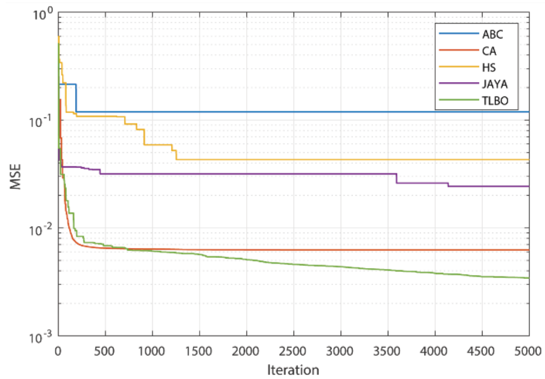
Singapore



Thailand



USA



Vietnam

Fig. 4. MSE Convergence Curves of different Metaheuristics Algorithm with 4-20-1 Neural Network Model Architecture.

The results indicated that the NN-CA algorithm was the fastest convergence speeds in Australia, Brunei, Canada, Chile, Hong Kong, Japan, Korea, Mexico, New Zealand, Peru, Philippines, Russia, Singapore, Thailand, and USA datasets. Within other datasets, the NN-CA was not deemed best; its results remained very competitive in each case.

IV. Conclusion

In this paper, MLP predicted consumer electricity usage, a method based on metaheuristic algorithms for the weights initialization of an MLP was implemented, as well as to analyze annual electricity consumption. The goals of the training problem were to avoid high local optima with convergence to the best solution in the predefined time. The result of the proposed technique was an MLP that has the lowest MSE. The proposed NN-CA outperformed all competitive algorithms, found the best-initialized weight values that the error back-propagation algorithm did not stick at a local minimum and can reduce the MSE effectively. The proposed method was proved to be suitable for the annual electricity consumption prediction, which will accurately support the power network infrastructure plan.

Because the MLP in this paper is a fully connected network, there are a lot of unnecessary weights or links. An MLP having only the necessary weights is not only more compact but also more accurate than the MLP with whole weights. Therefore, determining those weights and removing them from the final model is a necessity. However, this problem is very time-consuming. That is our future work.

Acknowledgment

This research was partially supported by the Advance Smart Computing Lab of the Department of Computer Science, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand.

References

[1] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," J. Microbiol. Methods, vol. 43, no. 1, pp. 3–31, 2000.

[2] G. Panchal, A. Ganatra, Y. P. Kosta, and D. Panchal, "Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers," Int. J. Comput. Theory Eng., vol. 3, no. 2, pp. 332–337, 2011.

[3] D. E. Rumelhart, R. Durbin, R. Golden, and Y. Chauvin, "Back-propagation: The basic theory," Backpropagation Theory, Archit. Appl., pp. 1–34, 1995.

[4] M. S. Kıran, E. Özceylan, M. Gündüz, and T. Paksoy, "Swarm intelligence approaches to estimate electricity energy demand in Turkey," Knowledge-Based Syst., vol. 36, pp. 93–103, 2012.

[5] M. E. Günay, "Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey," Energy Policy, vol. 90, pp. 92–101, 2016.

[6] M. Črepinšek, S.-H. Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms: A survey," ACM Comput. Surv., vol. 45, no. 3, pp. 1–33, 2013.

[7] S. Mirjalili, S. Z. M. Hashim, and H. M. Sardroudi, "Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm," Appl. Math. Comput., vol. 218, no. 22, pp. 11125–11137, 2012.

[8] X.-S. Yang, Engineering optimization: an introduction with metaheuristic applications. John Wiley & Sons, 2010.

[9] F. W. Glover and G. A. Kochenberger, Handbook of metaheuristics, vol. 57. Springer Science & Business Media, 2006.

[10] T. Bäck, D. B. Fogel, and Z. Michalewicz, Handbook of evolutionary computation. CRC Press, 1997.

[11] E. Bonabeau, M. Dorigo, D. de R. D. F. Marco, G. Theraulaz, and G. Théraulaz, Swarm intelligence: from natural to artificial systems, no. 1. Oxford university press, 1999.

[12] F. Gürbüz, C. Öztürk, and P. Pardalos, "Prediction of electricity energy consumption of Turkey via artificial bee colony: a case study," Energy Syst., vol. 4, no. 3, pp. 289–300, 2013.

[13] M. Kankal and E. Uzlu, "Neural network approach with teaching–learning-based optimization for modeling and forecasting long-term electric energy demand in Turkey," Neural Comput. Appl., vol. 28, no. 1, pp. 737–747, 2017.

[14] H. Ceylan, H. Ceylan, S. Haldenbilen, and O. Baskan, "Transport energy modeling with meta-heuristic harmony search algorithm, an application to Turkey," Energy Policy, vol. 36, no. 7, pp. 2527–2535, Jul. 2008.

[15] G. Phatai, S. Chiewchanwattana, and K. Sunat, "A Comparative of Neural Network with Metaheuristics for Electricity Consumption Forecast Modelling," in 2018 22nd International Computer Science and Engineering Conference (ICSEC), 2018, pp. 1–4.

[16] R. V Rao, V. J. Savsani, and D. P. Vakharia, "Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems," Comput. Des., vol. 43, no. 3, pp. 303–315, 2011.

[17] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," IEEE Trans. Evol. Comput., vol. 1, no. 1, pp. 67–82, 1997.

[18] R. Reynolds, An Introduction to Cultural Algorithms. 1994.

[19] I. Aljarah, H. Faris, and S. Mirjalili, “Optimizing connection weights in neural networks using the whale optimization algorithm,” Soft Comput., vol. 22, no. 1, pp. 1–15, 2018.

[20] S. Masood and P. Chandra, Training neural network with zero weight initialization. 2012.

[21] W. Cao, X. Wang, Z. Ming, and J. Gao, “A review on neural networks with random weights,” Neurocomputing, vol. 275, pp. 278–287, 2018.

[22] E. Talbi, “Population-based metaheuristics,” Metaheuristics from Des. to implementation, John Wiley Sons, Inc., Hoboken, New Jersey, pp. 190–200, 2009.

[23] R. G. Reynolds and B. Peng, “Cultural algorithms: modeling of how cultures learn to solve problems,” in 16th IEEE International Conference on Tools with Artificial Intelligence, 2004, pp. 166–172.

[24] S. khan, I. Qureshi, F. Zaman, B. Shoaib, A. Naveed, and A. Basit, “Correction of Faulty Sensors in Phased Array Radars Using Symmetrical Sensor Failure Technique and Cultural Algorithm with Differential Evolution,” ScientificWorldJournal., vol. 2014, p. 852539, Jan. 2014.

[25] H. Ma and Y. Wang, “Cultural Algorithm Based on Particle Swarm Optimization for Function Optimization,” in 2009 Fifth International Conference on Natural Computation, 2009, vol. 3, pp. 224–228.

[26] The World Bank, “World Development Indicators.” [Online]. Available: https://databank.worldbank.org/source/world-development-indicators.

[27] Expert Group on Energy Data and Analysis (EGEDA), “the APEC Energy Database.” [Online]. Available: https://www.egeda.ewg.apec.org/egeda/database_info/index.html.

[28] D. W. Zimmerman and B. D. Zumbo, “Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks,” J. Exp. Educ., vol. 62, no. 1, pp. 75–86, 1993.