

An Expression-Driven Approach for Long-Term Electric Power Consumption Forecasting

S. Hr. Aghay Kaboli¹, Alireza Fallahpour², Nima Kazemi², Jeyraj Selvaraj¹, N. A. Rahim¹

¹University Malaya Power Energy Dedicated Advanced Centre (UMPEDAC), Level 4, Wisma R&D, UM, Kuala Lumpur, Malaysia

²Department of Mechanical Engineering, Faculty of Engineering, University Malaya, Kuala Lumpur, Malaysia

Email address:

kaboli@siswa.um.edu.my (S. Hr. A. Kaboli), nasrudin@um.edu.my (N. A. Rahim)

To cite this article:

S. Hr. Aghay Kaboli, Alireza Fallahpour, Nima Kazemi, Jeyraj Selvaraj, N. A. Rahim. An Expression-Driven Approach for Long-Term Electric Power Consumption Forecasting. *American Journal of Data Mining and Knowledge Discovery*. Vol. 1, No. 1, 2016, pp. 16-28.

doi: 10.11648/j.ajdmkd.20160101.13

Received: November 24, 2016; **Accepted:** December 17, 2016; **Published:** January 13, 2017

Abstract: This study deals with estimation of electricity demand of Iran on the basis of economic criteria using a genetic-based approach called Gene Expression Programming (GEP) as an expression-driven approach. The GEP-based mathematical model is provided based on population, gross domestic product, exports, and imports. The proposed model is derived based on available data obtained from 1992 to 2006. To assess the forecasting accuracy of the model, the electricity demand from 2007 until 2012 are calculated by the GEP-based model and the obtained results are compared with the real data during this period. To show the accuracy of the model, the results obtained by GEP model are compared with the results obtained from Multi-Layer Perceptron (MLP) neural network and Multiple Linear Regression (MLR) as the two conventional methods. In addition, a five-fold cross-validation and future year prediction are used to show the robustness of the model in predicting the electricity demand. Future estimation of Iran's electric energy consumption is then projected up to 2030 according to three different scenarios. Finally, a sensitivity analysis is conducted to identify the most important independent variables affecting electricity demand.

Keywords: Data Mining Approach, Electric Power Consumption, Forecasting, Gene Expression Programming

1. Introduction

The contribution of electricity to economic development obviously is well-recognized [1]. Generally, there is a strong relationship between the generation and demand of the electricity energy. Optimum generation of the electricity energy is one of the main problems in energy management. Consequently, precise estimation of electricity demand could be one of the efficient methods to overcome this problem. Therefore, several techniques have been proposed to forecast the electricity demand. The proposed approaches can be categorized into three classes:

- i) Time series approach;
- ii) Multi-Linear Regression (MLR) approach;
- iii) Artificial Intelligence (AI) approach.

Auto Regressive Moving Average (ARMA) and Auto Regressive Integrated Moving Average (ARIMA) are the two well-known methods in time series analysis which have been widely used in the area of energy demand [1]. Azadeh et al. [2]

performed ARMA to predict estimation of electricity demand in China and Iran. Mohamed et al. [3] studied the double seasonal ARIMA model to predict the double seasonal (daily and weekly) Malaysia load demand time series. Furthermore, Ohtsuka et al. [4] have integrated ARMA and Spatial Auto-regression to predict electricity demand of Japan. Chujalet al. [5] applied ARMA and ARIMA to obtain a model that predict the electricity demand in a household for any given period. They concluded that the ARIMA model was the best model to find the most suitable forecasting period in a monthly and quarterly basis. In addition, ARMA model was the best model to obtain the most appropriate forecasting period on daily and weekly basis.

MLR is one of the most common techniques in behavioural modelling approaches, which have been applied successfully in energy demand forecasting. Bianco et al. [6] used linear regression to find a model to predict electricity consumption

using gross domestic product (GDP), gross domestic product per capita (GDP per capita) and population in Italy. The result obtained by the model was comparable to the official projections. In [7] authors carried out a research to forecast the amount of consumption of the electricity energy in Jordan using MLR. The results showed that MLR is a reliable method to analyse the electricity consumption. Although MLR and other statistical-based techniques have been used in the electricity demand prediction but the main shortcoming is that they model the nature of the corresponding problem by a pre-defined mathematical model [8].

In the past decade, AI approaches have received more attention from both academics and practitioners in solving real life problems. The AI approach designs computer programs based on historical dataset. Artificial Neural Network (ANN) is the most widely used technique among the AI approach which has been applied in the field of energy management [9-11]. Pao [12] investigated an ANN model with single output node structure to predict electricity price. To demonstrate the validity of the proposed ANN model, a cross validation technique was used. The results showed that the proposed ANN model is more accurate compared to MLR model. Murat and Ceylan [13] provided a non-linear ANN-based model for energy demand using gross national product (GNP), population and total annual average Veh-km in Turkey from 1970 to 2001. The results illustrated that ANN is a suitable technique to forecast the energy demand. Geem [14] proposed a Multi-Layer Perceptron (MLP) ANN to estimate the energy demand in South Korea based on gross domestic product (GDP), population, import, and export amounts. They compared the results derived from ANN with the results obtained from MLR and concluded that ANN is more reliable and accurate in estimating the energy demand. Support Vector machine is another AI-based technique, which has been used as a powerful predictive technique in energy demand estimation. Fan and Chen [15] integrated self-organized map (SOM) and support vector machine (SVM) based on an adaptive two-stage hybrid network to propose a load forecasting method. The proposed method is capable of adapting to different models automatically for regular and irregular days at the same time. In addition, Fan et al. [16] developed a hybrid forecasting model for day ahead electricity load forecasting. Bayesian clustering by dynamics (BCD) was combined with support vector regression (SVR) to handle the non-stationary of time series.

In addition to the AI-based techniques mentioned above, Fuzzy systems [17-20] especially Adaptive Fuzzy Inference System (ANFIS) [21-24] has been applied to forecast electricity demand. Zahedi et al. [25] modelled electricity demand in Ontario (Canada) from 1976 to 2005 using ANFIS based on employment, GDP, population, dwelling count and weather temperature. The results demonstrated that electricity demand is most sensitive to employment.

Moreover, the metaheuristic methods [26-28] are applied on the pre-defined mathematical functions to model the energy demand forecasting. Assareh et al. [29] used particle

swarm optimization (PSO) to estimate the energy demand of Iran from 1981 to 2005 and compared the results with GA. Ceylan [30] approximated energy demand on the basis of economic variables in Turkey using genetic algorithm (GA). The error of the proposed model was lower when they were compared with the Ministry of Energy and Natural Resources (MENR) projection.

Although aforementioned techniques are very efficient in forecasting but the major problem is that they have black box system [31]. This implies that they are not capable of providing an explicit mathematical model for the dependent variable (i.e. energy demand) using the independent variables (e.g. GDP, population, exports and imports). To overcome the black box problem of the AI-based techniques, a robust genetic-based intelligent technique has been introduced known as Genetic Programming (GP).

The strength of GP is that not only its power of prediction is higher than the conventional AI-based techniques but also it solves the weakness of black box feature. In the field of energy demand prediction, Mostafavi et al. [1] proposed a mathematical GP-based model for electricity demand based on GDP, Stock index (SI), Total revenue (TR) and population (P) in Thailand. They used Correlation coefficient (R), root mean squared error (RMSE) and mean absolute percent error (MAPE) as the model performance evaluation factors. The calculated results were compared with MLR and MLP neural network. The results depicted that the proposed GP-based model is more accurate than MLP ANN and MLR. Furthermore, some practitioners have proposed other predictor for energy demand forecasting via different AI-based techniques. Table 1 summarizes the AI-based methods which have been used for Long-term energy demand forecasting.

This study aims to introduce a new variant of GP namely Gene Expression Programming (GEP) which has not received much attention in the context of energy demand in modelling the electricity demand using the collected dataset of Iran (1992 to 2012).

Since the relationship between the electricity demand and the input variables is non-linear, proposing a model with high accuracy is a challenging task for smart grids as are equipped with advanced techniques [32-40].

Therefore, the main purpose of this study is to provide a precise mathematical model for the electricity demand based on the determined criteria using GEP to overcome the problems of both the black box system and precision of the model in prediction. The developed GEP function is a strategic model in analyzing the electricity demand for considering the energy policy of the country.

Table 1. IA-based methods applied for long-term energy consumption forecasting.

Models	Input Variables	Country
[13], [14], [41]	GNP, Population, and veh-km; GDP, population, import, and export amounts; Yearly ambient temperature, installed power	Turkey S. Korea Turkey

	Models	Input Variables	Country
		capacity, yearly per resident electricity consumption, and gross domestic product	
[17]	ANFIS	employment, GDP, population, dwelling count and weather temperature	Canada
[42], [43]	SVM	Time series; time series	Taiwan China
[19]	GA	Population, Gross National Product, import and export	Turkey
[18]	PSO	GDP, population, import, and export amounts	Iran
[44], [45]	Ant Colony	Population, GDP, import and export; population, GDP, import and export	Turkey; Turkey
[1], [46]	GP	population, GDP, stock index, total revenue; Time series	Thailand, S. Korea

In order to evaluate the performance of the GEP-based model, the Roy and Roy test is performed to show the predictive ability of the GEP-based model. MLP ANN and MLR as conventional techniques are used to compare the accuracy of the proposed GEP-based model. In addition, the 5-fold cross validation is used for further verification of the model. To find out which variables have the most influence on the electricity demand, a sensitivity analysis is conducted.

The rest of the paper is organized as follows: Section 2 presents a brief overview of the GEP and explains in detail how the GEP-based methodology is implemented. The obtained results of the electricity demand of Iran and its discussion are provided in section 3. The performance of the model is evaluated in section 4. Section 5 presents the results of the sensitivity analysis. The last section presents the conclusion.

2. Methodology

This part briefly reviews GEP and MLP ANN. It also explains comprehensively the proposed methodology applied to provide a genetically mathematical model for electricity demand of Iran.

2.1. Gene Expression Programming (GEP)

Holland [47] proposed GA as an evolutionary technique. In GA, using specific operators such as selection, mutation, and crossover, chromosomes (solutions) are generated. Genetic Programming (GP) is an extension of the genetic algorithms [48] (by manipulating the above operators). The major difference between genetic programming and genetic algorithms is the presentation of the solution. GA generates string of numbers as the solution, whereas the solution of GP is a computer program (tree structures). Based on a fitness landscape, a population of computer programs is optimized by GP. The fitness of each created program in the population is evaluated via a fitness function. Every program derived from GP is a structured tree as illustrated in Figure 1. In order to change the tree-based model into mathematical model (T), it should be read from left to right.

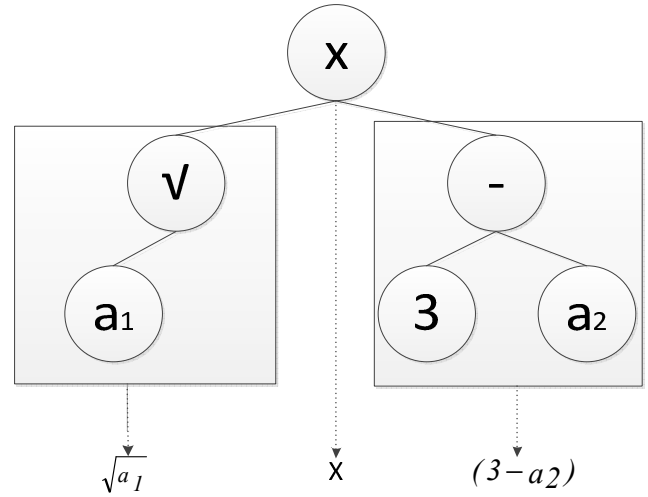


Figure 1. Tree representation of a GP model.

Generally, among the three different types of GP [49] (tree-based, linear-based, and graph-based), linear based GP has received most attention. Figure 2 shows the types of GP.



Figure 2. Different types of GP.

In Linear GP (LGP), the created programs are depicted as linear strings. These linear strings are decoded and expressed as nonlinear entities [50]. In comparison with tree-based GP, LGP is faster [51]. Therefore, more runs can be carried out by LGP in a period of time. As a robust LGP technique, GEP has been widely used in solving real problems [52-55]. GEP [56] was first invented by Ferreira in which functional set, terminal set, fitness function, control parameters, and termination condition are the most important elements. In the GEP algorithm, a fixed length of character strings is used to represent solutions to the problems, which are afterwards expressed as Expression Trees (ETs) in different sizes and shapes. Based on multi-genic nature of GEP, more complex programs composed of several sub-programs will be permitted to be created during the evolutionary process. Every GEP gene consist of a list of symbols which are elements from function and terminal sets like $\{+, -, /, \sin\}$ and the terminal set like $\{a_1, a_2, a_3, +4\}$. A typical GEP gene is as follows:

$$+.\times.\sin.a_1.-.+.\times.a_2.a_1.a_3.+4.a_2.a_1 \quad (1)$$

The above expression is termed as Karva notation or K-expression. A K-expression can be depicted by a diagram that is an ET. As an example, Figure 3 represents the expression tree of the sample gene.

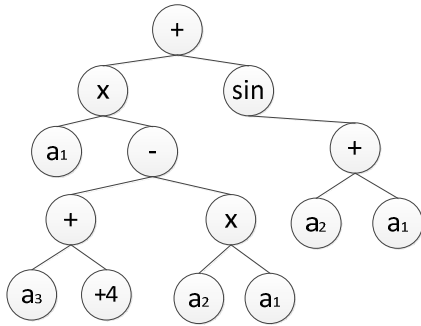


Figure 3. GEP ETs.

The conversion process is created from the start position in the K-expression, which corresponds to root of the ET, and reads through the string one by one. The mentioned GEP gene can be shown in a mathematical equation as:

$$a_1((a_3 + 4) - (a_2 \times a_1)) + \sin(a_2 + a_1) \quad (2)$$

The main four steps in GEP to achieve a terminal condition are as follows:

- I. Random generation of the fixed-length chromosome of each individual for the initial population.
- II. Expressing chromosomes as ET and evaluating fitness of each individual.
- III. Selecting the best individuals according to their fitness to reproduce with modification.
- IV. Repeating the above process for a definite number of generations or until a solution has been found.

In GEP, the individuals are chosen and copied into the next generation based on the fitness by roulette wheel sampling with elitism. This guarantees the survival and cloning of the best individual to the next generation. In GEP, by different combinations of genetic operators (including crossover, mutation and rotation), variation in the population is carried out. The rotation operator is performed to rotate two subparts of an element sequence in a genome based on a randomly selected point. This can also considerably restructure the ETs. For instance, the following gene rotates the first five elements of gene (1) to the end:

$$+.+. \times a_2. a_1. a_3. +4. a_2. a_1. +. \times. \sin. a_1. - \quad (3)$$

The solution function is built based on only the first 7 components, with the corresponding expression presented in Figure 4.

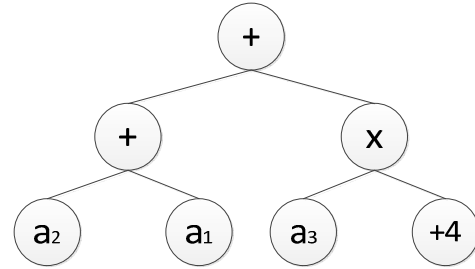


Figure 4. An instance of ET.

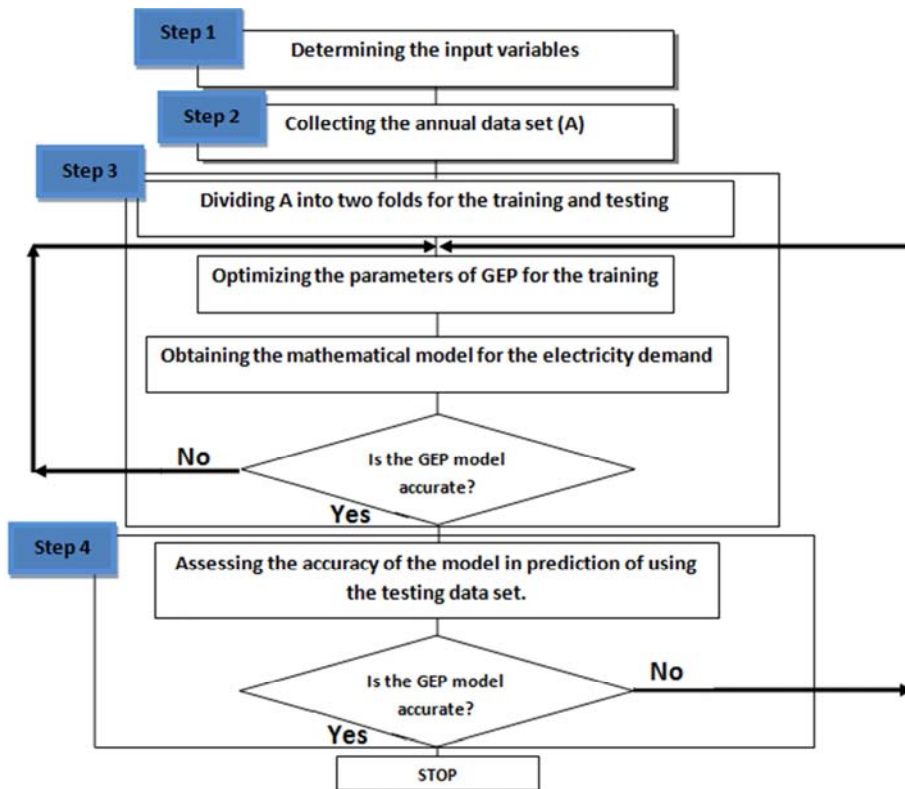


Figure 5. Proposed model.

2.2. The Proposed GEP-based Model

As shown in Figure 5, in order to find the most accurate mathematical GEP-based model for the electricity demand estimation, the following steps are carried out.

- Determining variables as inputs influence the electricity demand.
- Collecting the data set.
- Dividing the collected data set for training and testing.
- Running GEP for the training process. In this step, the parameters of GEP are optimized and consequently the best mathematical model is found.
- Evaluating the accuracy provided mathematical model developed by GEP using the testing data set.

To evaluate the performance of the model in both the training and testing (in terms of the precision), correlation coefficient (R), root mean squared error (RMSE) and mean absolute percent error (MAPE) are used as the statistic measures (Equations 4, 5, and 6).

$$R = \frac{(\sum_{i=1}^n (O_i - \bar{O}_i)(t_i - \bar{t}_i))}{\sqrt{\sum_{i=1}^n (O_i - \bar{O}_i)^2 \sum_{i=1}^n (t_i - \bar{t}_i)^2}} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - t_i)^2}{n}} \quad (5)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left[\frac{|O_i - t_i|}{O_i} \right] \quad (6)$$

where, O_i and t_i are the actual and predicted output values for the i^{th} output respectively, \bar{O}_i and \bar{t}_i are the average of the actual and predicted outputs, and n is the number of samples respectively.

3. Real Application of the Proposed Model

This research proposed a genetic-based model for the electricity demand of Iran from 1992 to 2012. Estimation of the electricity demand can be calculated for short, medium or long-term periods. The short-term electricity demand prediction is required for controlling, scheduling or dispatching of power systems. In this study, GEP approach has been used to provide a mathematical model based on long-term (annual) electricity demand data set in Iran.

The first step in modelling is to select the most appropriate inputs which are influential on output (s). To carry out this study, population (P) (million), GDP, Import (Im) (USD) and export (Ex) (USD) of Iran were selected as the inputs (independent variables) and the electricity demand (ED) was selected as the output (dependent variable) as shown in Figure 6. It is obvious that demand is related to population as the population increases, more electricity will be consumed. GDP is a criterion of all economic activities and rising GDP means improved living standards and therefore increased energy use. Finally, Imports and exports for Iran (as an developing country) are related to manufacturing processes and therefore

robustly influence electricity demand. Table 2 shows the statistic features of the inputs and output variables as the results of the first step. Note that the data set was collected from the following addresses:

Population (P) [57]; GDP [58]; Imports (I) and Exports (E) [59].

Table 2. Statistic feature of the collected dataset.

	P (10 ⁶)	GDP (\$10 ⁹)	I (\$10 ⁹)	E (\$10 ⁹)	ED (TWh)
Mean	67.721	178.500	30	10	148113
Max	76.420	528.430	54	338	254265
Min	58.310	50.400	11	26	63982
Std	5.695	135.000	17	10	62557

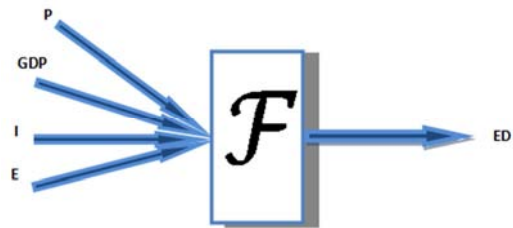


Figure 6. Topology of the electricity demand model.

After collecting the data related to electricity demand of Iran, the dataset was divided into two parts for the training (optimizing GEP parameters and finding the best mathematical model) in the third step and testing (as untrained dataset) for evaluating the prediction ability of the model as the forth step. It is recommended that 75% of the data set is dedicated to the training and 25% is dedicated to testing. To find the most accurate GEP-based mathematical model, several runs have been carried out for determining the best parameterization for the GEP until no important minimization of error was observed through the run.

There are four main parts for setting GEP parameters including: General Setting; Fitness Function; Genetic Operators; Numerical Constant. Table 3 shows the optimized parameters of GEP for finding the most accurate model based on R, RMSE and MAPE.

Table 3. The best parameters for the GEP algorithm.
(y can be any number)

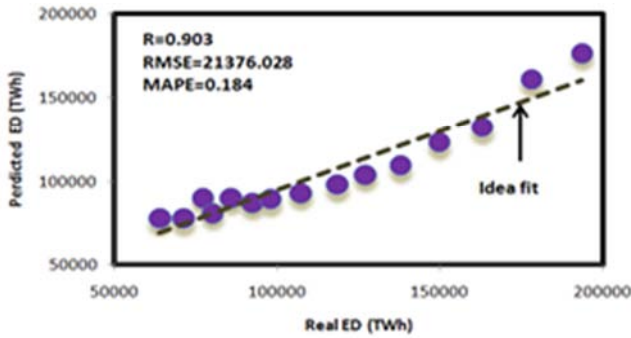
	Parameters	Value
General	Chromosomes	30
	Function set	$\times, /, +, -, \text{power}(x, y^*), e^y, \text{Tan}$
	Number of genes	4
	Head size	8
Fitness Function	Linking function	Addition
	RMSE	
Genetic Operators	Mutation rate	0.044
	One-point recombination rate	0.2
	Two-point recombination rate	0.3
	Gene recombination rate	0.1
	Gene transportation rate	0.1
Numerical Constant	Constants per gene 2	
	Data type	Floating Point
	Lower bound	-10
	Upper bound	+10

The optimal GEP-based mathematical model of the electricity demand (ED) is as formulated below:

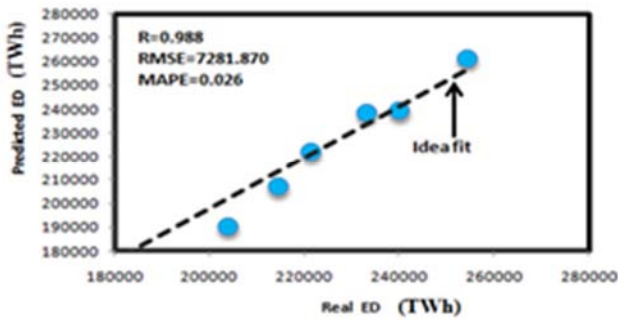
$$\left(x_{(GDP)} - \left(\frac{x_{(Im)} - 7.72x_{(P)}}{x_{(Ex)} \sin(x_{(Ex)}x_{(GDP)})} \right) \right) + \left(x_{(GDP)} - \left(\frac{\log\left(\frac{x_{(Im)}}{x_{(GDP)}}\right)^2}{\sin(x_{(P)})} \right) \right) + \left(x_{(GDP)} - \frac{x_{(P)}}{\sin(-5.56 - x_{(P)})} \right) + \left(\left(\sqrt[3]{\cos x_{(GDP)} + x_{(Ex)}} + (\log x_{(GDP)})^3 \right) x_{(P)} \right) \quad (7)$$

In which $x_{(p)}$, $x_{(GDP)}$, $x_{(Ex)}$, and $x_{(Im)}$ are the inputs (independent) variables and ED is the output (dependent) variable. Figure. 7 presents the results obtained by developed GEP model.

Additionally, the Expression Tree (ET) of developed model and its code in different languages are provided in appendix.



(a)



(b)

Figure 7. Real versus predicted electricity demand using the GEP model (a) training data, (b) testing data.

4. Performance Evaluation of the GEP Model

In this section, different methods are applied to validate the proposed GEP-based model. On the basis of a rational hypothesis, Smith [1] recommended the following factors for assessing the performance of a predictive model:

I. If a model gives $|R| > 0.8$, a strong correlation exists between the predicted and real values.

II. If a model gives $0.2 < |R| < 0.8$ a correlation exists between the predicted and real values.

III. If a model gives $|R| < 0.2$, a weak correlation exists between the predicted and real values.

In all conditions, the error values (e.g. MSE, MAE) should be at the minimum. It can be seen from Figure 7 that the GEP model predicts accurately both training ($R_{\text{training}}=0.903$, $\text{RMSE}_{\text{training}}=21376.028$, $\text{MAPE}_{\text{training}}=0.184$) and testing ($R_{\text{testing}}=0.988$, $\text{RMSE}_{\text{testing}}=7281.87$, $\text{MAPE}_{\text{testing}}=0.026$) data set. Besides, new factors suggested by Golbraikh and Tropsha [1] were checked for external validation of the models on the validation data sets. It is recommended that at least one slope of the regression lines (k or k') through the origin should be close to 1. It should be noted that k and k' are the slopes of the regression lines between the regressions of actual output (h_i) against predicted output (t_i) or t_i against h_i through the origin, i.e. $h_i = k t_i$ and $t_i = k' h_i$, respectively. In addition, the performance indexes of m and n should be less than 0.1 (m and n are the two factors for evaluating the model performance). Newly, Roy and Roy [1] presented a confirmed indicator (R_m) of the external predictability of models. For $R_m > 0.5$, the condition is satisfied. Either the squared correlation coefficient (through the origin) between predicted and experimental values (R_o^2), or the squared correlation coefficient between experimental and predicted values ($R_o'^2$) should be close to R^2 and to one. The considered validation criteria and the relevant results obtained by the models are given in Table 4. As it is observed, the developed model satisfies all the requisite conditions. The validation phase ensures that the proposed model is strongly suitable and applicable.

In order to understand how well is the predictive ability of the GEP model, MLP as a useful network and MLR as a prevalent predictor technique have been used.

MLP as feedforward neural network is usually trained with Back Propagation (BP) learning algorithm. There are at least three layers in an MLP network including an input layer, one hidden layer of neurons (at least) and an output layer. Each layer consists of several nodes. Determining the number of the hidden layers, nodes, learning rate, epochs and type of motivation function is very important in optimizing the MLP structure. Therefore, several networks with different setting were run to find the best structure with minimum error. Similar to the GEP model, the criteria of the MLP model were the yearly values of P , GDP , Im , and Ex and the single output was the ED of the same year. The same data set were used for training and testing the MLP. To implement the MLP neural network, Neuro Solution 5 software was used. As said before, the learning algorithm is BP learning algorithm. Numbers of nodes in the input layer are equal to numbers of independent variables (P , GDP , Im and Ex). In output layer, there is only one node to show the ED . Learning rule of the MLP model is momentum (the learning rate is 0.7) and trained for 1000 epochs. Figure 8 depicts the accuracy of the prediction GEP model in comparison with MLP and MLR for the

testing dataset.

As it can be seen, the results derived by the two AI-based approaches (GEP and MLP) are a significant improvement over those calculated by the MLR model. In general, the performance of ANN model is satisfactory, but in comparison with the GEP model, it can be seen that for all years the error prediction of the GEP model is less than the prediction error of the ANN model. Moreover, the major shortcoming of this

neural-based method is that it does not provide practical prediction equations. Therefore, it is very complicated for practitioners to use and analyse the developed ANN models.

Generally, AI techniques may need large data set for effective application. In the area of energy demand forecasting, obtaining large data set for all the inputs (P, GDP, Im and Ex) is difficult.

Table 4. Statistical factors of the decision model for external validation.

	Formula	Condition	Testing
1	R	$0.8 < R$	0.988
2	$k = \frac{\sum_{i=1}^n (h_i \times t_i)}{h_i^2}$	$0.85 < k < 1.15$	0.995
3	$k' = \frac{\sum_{i=1}^n (h_i \times t_i)}{t_i^2}$	$0.85 < k' < 1.15$	1.003
4	$m = \frac{R^2 - R_o^2}{R^2}$	$m < 0.1$	-0.022
5	$n = \frac{R^2 - R_o'^2}{R^2}$	$n < 0.1$	-0.022
6	$R_m = R^2 \times \left(1 - \sqrt{ R^2 - R_o^2 }\right)$	$0.5 < R_m$	0.832
Where	$R_o^2 = 1 - \frac{\sum_{i=1}^n (t_i - h_i^o)^2}{\sum_{i=1}^n (t_i - \bar{t}_i)^2}, h_i^o = k \times t_i$	$R_o^2 = 0.9979$	
	$R_o'^2 = 1 - \frac{\sum_{i=1}^n (h_i - t_i^o)^2}{\sum_{i=1}^n (h_i - \bar{h}_i)^2}, t_i^o = k' \times h_i$	$R_o'^2 = 0.9976$	

Table 5. The performance of GEP, ANN and regression models on data sets for five-fold cross-validation.

	GEP Model		GEP Model		GEP Model	
	Training dataset	Testing dataset	Training dataset	Testing dataset	Training dataset	Testing dataset
	R	R	RMSE	RMSE	MAPE	MAPE
1	0.991	0.645	10667.240	18165.027	0.050	0.232
2	0.987	0.942	13866.221	4720.022	0.110	0.030
3	0.995	0.963	11296.213	14768.531	0.094	0.080
4	0.963	0.999	13586.003	10414.718	0.102	0.078
5	0.979	0.320	11994.418	14091.030	0.072	0.132
Average (GEP)	0.983	0.7738	12276.623	12431.870	0.085	0.110
Average (MLP)	0.862	0.697	19450.231	17338.0963	0.117	0.164
Average (MLR)	0.783	0.649	24914.753	31243.423	0.243	0.218

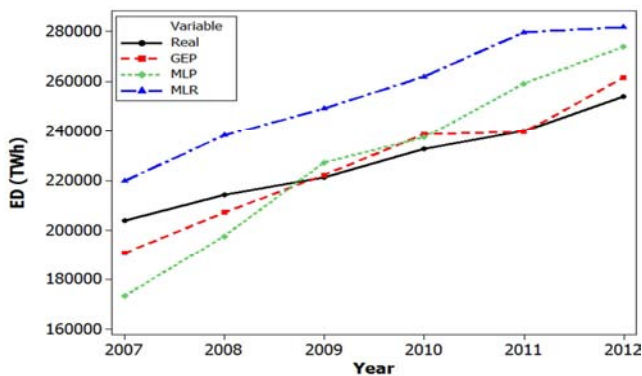


Figure 8. A comparison of the electricity demand predictions made by different models.

GEP (R=0.988, RMSE=7281.870, MAPE=0.026)

MLP (R=0.906, RMSE=18480.912, MAPE=0.072)

MLR (R=0.872, RMSE=28260, MAPE=0.119)

Thus, scholars and researchers usually collect small size data set for modelling the energy demand. Data partitioning techniques such as cross-validation methods have been

proposed to deal with small datasets [50]. The advantages of this method is that the limited number of data point can be used in both training and testing set as many times as possible. In this study, a five-fold cross validation has been used to demonstrate the accuracy of the GEP-based model in terms of sampling variation. Training and testing process were carried out five times. To implement the five-fold cross-validation, the dataset was divided into five parts and repeatedly ran the GEP model for five times. Each testing dataset contained 6 years and the rest of the dataset should be considered as the training dataset (15 year). Since there were 21 samples and 6 samples were selected for testing, separated subsets were not mutually exclusive and the testing datasets for five-fold cross-validation are as follows. Samples 1 to 6 for testing for the first time; samples 7 to 12 for the second time; samples 13 to 18 for the third time, samples 19, 20, 21, 1, 2, 3 for the fourth time; and samples 4 to 9 for the fifth time. R, RMSE, and MAPE were applied as assessment criteria. After training and testing, the average of R, RMSE, and MAPE were utilized to obtain the model precision. Table 4 illustrates the results of the GEP-based in five-fold cross-validation for both the

training and testing and compares the results of the GEP model with MLP and MLR. As can be seen, the GEP model performs superior to MLP and MLR models in terms of R, RMSE, and MAPE.

In order to use GEP model for future projections, each input variable should be forecasted in the future time domain. Following scenarios are defined for forecasting each socio-economic indicator in the future [60]:

Scenario I: It is assumed that the annual average growth rates of population, GDP, import and export were 1.6%, 4.5%, 6%, and 3.5% during 2014-2030.

Scenario II: It is assumed that the annual average growth rates of population, GDP, import and export were 1.4%, 4.5%, 6.5%, and 4.5% during 2014-2030.

Scenario III: It is assumed that the annual average growth rates of population, GDP, import and export were 1.5%, 5%, 7.5%, and 2.5% during 2014-2030.

Table 6. shows the electricity demand in the future time domain (2014-2030) by GEP model with tree different scenarios for socio-economic indicators.

Table 6. The electricity demand in the future time domain.

ED (TWh)			
Years	Scenario I	Scenario II	Scenario III
2013	270147.3708	267541.5608	268644.452
2014	279455.0047	278098.8827	277401.6644
2015	283614.7633	283868.0312	284098.3323
2016	292098.7104	292415.9845	292449.8797
2017	262936.9717	300354.7164	300545.2028
2018	311938.9657	305259.3237	310400.8181
2019	318773.8046	317689.277	317596.9504
2020	327637.7961	326651.0428	335208.878
2021	337169.604	333884.0616	336585.4935
2022	344914.7279	345326.9227	343584.8211
2023	352854.6683	354070.1726	355069.8989
2024	367008.2407	365490.4583	363606.2985
2025	375901.424	376060.6674	376035.6035
2026	388193.655	386657.307	386328.5616
2027	525322.8882	398604.0099	399324.8506
2028	407812.761	416600.743	403025.762
2029	419860.0118	425408.2985	420934.0227
2030	420336.4014	432351.9508	428988.1372

5. Sensitivity Analysis

Sensitivity analysis of the inputs that affect the electricity demand was carried out in order to appraise the importance of the independent variables to the prediction of the electricity demand of Iran. A frequency value equal to 100 percent for an input shows that this variable has appeared in all of the best thirty programs evolved by GEP. Figure 9 illustrates the frequency values (percentages) of the inputs. According to this figure, the electricity demand of Iran is particularly sensitive to

all of the determined variables. However, the electricity demand seems to be more affected by P, GDP and Ex.

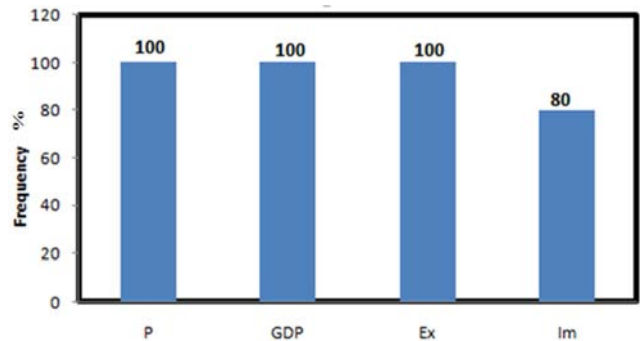


Figure 9. Contributions of the variables in the GEP analysis.

6. Conclusion

Multivariable influence the electricity demand. Therefore, modeling the electricity demand is a complicated process. Moreover, the conventional methods such as regression are not capable of modeling the electricity demand based on inputs with high accuracy. So, alternative methods such as ANN, SVM, and ANFIS have been used (as AI techniques) to predict the electricity demand precisely. Although the mentioned AI techniques can find the non-linear relationship more accurately than regression but their major shortcoming is the black box system. This means the aforementioned AI techniques do not propose an explicit mathematical model for the electricity demand based on the determined inputs (independent variables).

This study introduces a new robust genetic-based approach namely GEP for the empirical modeling of the electricity demand in Iran on the basis of historical data from 1992 to 2012. In this study, not only a mathematical GEP-based model was provided for the electricity demand, but also it was demonstrated that the proposed model is more accurate than MLP and MLR as the two efficient techniques in this context. In order to evaluate the accuracy of the GEP model in terms of sampling variation, a five-fold cross validation was carried out. For further verification of the validity of the GEP model in contrast with the conventional methods, the data related to year 2013 as the unseen and untrained data was used and the electricity demand of that year was predicted by the GEP model. The derived results showed that the precision of the GEP model was higher than MLP and MLR. Meanwhile, a sensitivity analysis was conducted and the results depicted that the electricity demand is influenced more by population, GDP, and exports.

Acknowledgment

This work has been supported by HIR (high impact research) secretariat at university of Malaya through the “Campus network smart grid system for energy security” project (under grant number: H-16001-00-D000032).

Appendix

Expression Tree (ET) of developed GEP model:

Ex=d0,

GDP=d1,

Im=d2,

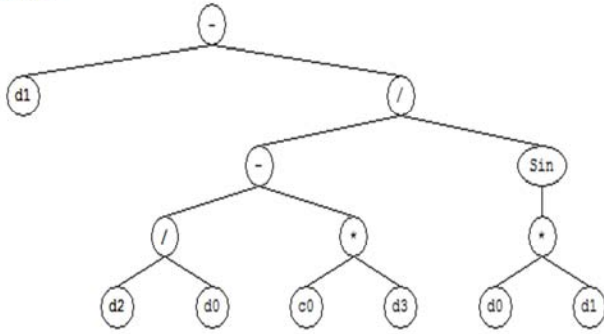
P=d3,

Sub-ET1 C0=7.725403

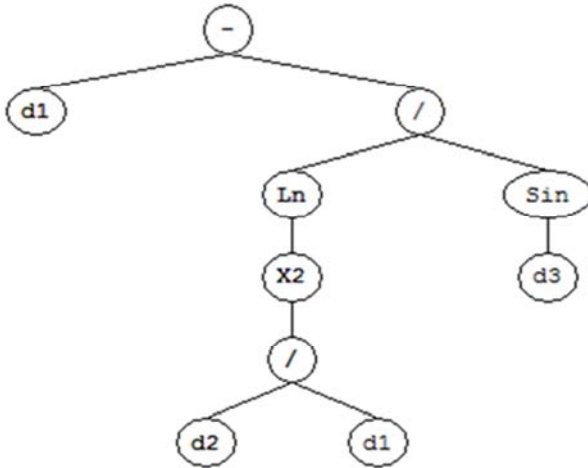
Sub-ET3 C0=-5.56546

ED=Sub-ET1+Sub-ET2+Sub-ET3+Sub-ET4

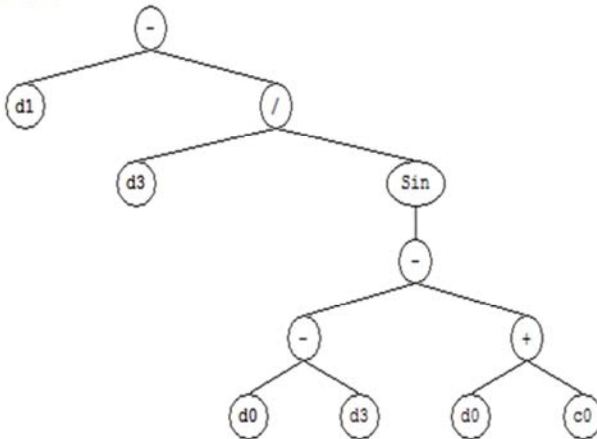
Sub-ET 1



Sub-ET 2



Sub-ET 3



Sub-ET 4

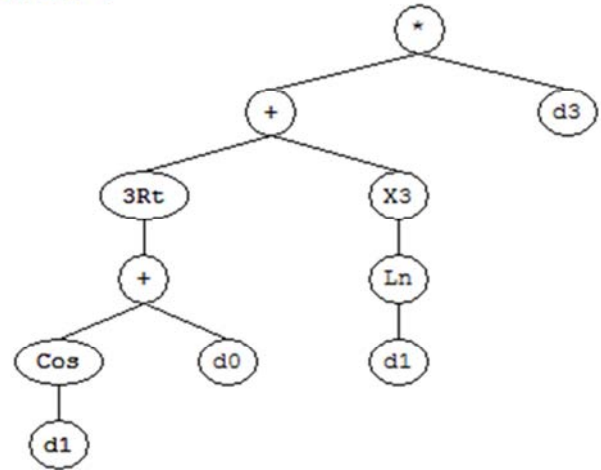


Figure A1. Expression Tree (ET) of developed GEP model (ED=Sub-ET 1+ Sub-ET 2+ Sub-ET 3+ Sub-ET 4).

MATLAB code of developed GEP model:

```
%-----
% Code generated by Gene Xpro Tools 4.0
% Fitness Function: RRSE
% Ex=d (1);
% GDP=d (2);
% Im=d (3);
% P=d (4);
% varTemp=ED;
%-----
function result=gepModel (d)
G1C0=7.725403;
G1C1=-7.681702;
G2C0=-6.125427;
G2C1=6.147918;
G3C0=-5.56546;
G3C1=-1.669892;
G4C0=6.733307;
G4C1=4.782715;
varTemp=0.0;
varTemp=(d (2)-(((d (3)/d (1))-(G1C0*d (4)))/sin (((d (1)*d
(2)))));
varTemp=varTemp + (d (2)-(log (((d (3)/d (2))^2)/sin (d
(4)))));
varTemp=varTemp + (d (2)-(d (4)/sin (((d (1)-d (4))-(d
(1)+G3C0)))));
varTemp=varTemp + (((cos (d (2))+d (1))^(1.0/3.0))+ (log
(d (2)^3))*d (4));
result=varTemp;
```

C++ code of developed GEP model:

```
//-----
// Code generated by Gene Xpro Tools 4.0
// Fitness Function: RRSE
// Ex=d [0]
```

```
// GDP=d [1]
// Im=d [2]
// P=d [3]
//dblTemp=ED
//-----
using System;
class gepModel
{ public double Calculate (double [] d)
{
const double G1C0=7.725403;
const double G1C1=-7.681702;
const double G2C0=-6.125427;
const double G2C1=6.147918;
const double G3C0=-5.56546;
const double G3C1=-1.669892;
const double G4C0=6.733307;
const double G4C1=4.782715;
double dblTemp=0.0;
dblTemp=(d [1]-(((d [2]/d [0])-(G1C0*d [3]))
/Math.Sin ((d [0]*d [1]))));
dblTemp +=(d [1]-(Math.Log (Math.Pow ((d [2]/d
[1]),2))/Math.Sin (d [3])));
dblTemp +=(d [1]-(d [3]/Math.Sin (((d [0]-d [3])-(d
[0]+G3C0)))));
dblTemp +=((Math.Pow ((Math.Cos (d [1])+d [0]),
(1.0/3.0))+ Math.Pow (Math.Log (d [1]),3))*d [3]);
return dblTemp;
}}}
```

JAVA code of developed GEP model:

```
//-----
// Code generated by Gene Xpro Tools 4.0
// Fitness Function: RRSE
// Ex=d [0]
// GDP=d [1]
// Im=d [2]
// P=d [3]
//dblTemp=ED
//-----
class gepModel
{ public double Calculate (double [] d)
{
final double G1C0=7.725403;
final double G1C1=-7.681702;
final double G2C0=-6.125427;
final double G2C1=6.147918;
final double G3C0=-5.56546;
final double G3C1=-1.669892;
final double G4C0=6.733307;
final double G4C1=4.782715;
double dblTemp=0.0;
dblTemp=(d [1]-(((d [2]/d [0])-(G1C0*d [3]))/Math.sin ((d
[0]*d [1]))));
dblTemp +=(d [1]-(Math.log (Math.pow ((d [2]/d
[1]),2))/Math.sin (d [3])));
dblTemp +=(d [1]-(d [3]/Math.sin (((d [0]-d [3])-(d
```

```
[0]+G3C0)))));
dblTemp +=((Math.pow ((Math.cos (d [1])+d [0]),
(1.0/3.0))+Math.pow (Math.log (d [1]),3))*d [3]);
return dblTemp;
}}}
```

PASCAL code of developed GEP model:

```
{-----}
{ Code generated by Gene XproTools 4.0 }
{ Fitness Function: RRSE }
{Ex=d [0] }
{GDP=d [1] }
{ Im=d [2] }
{P=d [3] }
{rlTemp=ED }
{-----}
program gepModel (input,output);
var
d: array [0..3] of real;
rlTemp, result: real;
const G1C0=7.725403;
const G1C1=-7.681702;
const G2C0=-6.125427;
const G2C1=6.147918;
const G3C0=-5.56546;
const G3C1=-1.669892;
const G4C0=6.733307;
const G4C1=4.782715;
function gepXPower3 (x: real): real;
begin
if (x<0) then gepXPower3:=-exp (3*ln (abs (x))) else
if (x=0) then gepXPower3:=0.0 else
gepXPower3:=exp (3*ln (abs (x)));
end;
function gepXPower3Rt (x: real): real;
begin
if (x=0) then gepXPower3Rt:=0.0 else
gepXPower3Rt:=exp ((1.0/3.0)*ln (x));
end;
begin
rlTemp:=0.0;
rlTemp:=(d [1]-(((d [2]/d [0])-(G1C0*d [3]))/sin ((d [0]*d
[1]))));
rlTemp:=rlTemp + (d [1]-(ln (sqr ((d [2]/d [1])))/sin (d
[3])));
rlTemp:=rlTemp + (d [1]-(d [3]/sin (((d [0]-d [3])-(d
[0]+G3C0)))));
rlTemp:=rlTemp + ((gepXPower3Rt ((cos (d [1])+d
[0]))+gepXPower3 (ln (d [1])))*d [3]);
result:=rlTemp;
end.
```

VISUAL BASIC code of developed GEP model:

```
'-----
' Code generated by GeneX pro Tools 4.0
```

```
' Fitness Function: RRSE
' Ex=d (0)
' GDP=d (1)
' Im=d (2)
' P=d (3)
' dblTemp=ED
'-----
Function gepModel (ByRef d () As Double) As Double
Const G1C0 As Double=7.725403
Const G1C1 As Double=-7.681702
Const G2C0 As Double=-6.125427
Const G2C1 As Double=6.147918
Const G3C0 As Double=-5.56546
Const G3C1 As Double=-1.669892
Const G4C0 As Double=6.733307
Const G4C1 As Double=4.782715
Dim dblTemp As Double
dblTemp=0.0
dblTemp=(d (1)-(((d (2)/d (0))-(G1C0*d (3)))/Sin ((d
(0)*d (1))))))
dblTemp=dblTemp (d (1)-(Log (((d (2)/d (1))^2)/Sin (d
(3))))))
dblTemp=dblTemp + (d (1)-(d (3)/Sin (((d (0)-d (3))-(d
(0)+G3C0))))))
dblTemp=dblTemp + (((Cos (d (1))+d (0))^ (1.0/3.0))+
(Log (d (1)^3))*d (3))
gepModel=dblTemp
End Function
```

References

- [1] E. S. Mostafavi, S. I. Mostafavi, A. Jaafari, F. Hosseinpour. A novel machine learning approach for estimation of electricity demand: An empirical evidence from Thailand. *Energy Conversion and Management*. 74 (2013) 548-55.
- [2] A. Azadeh, M. Saberi, S. Ghaderi, A. Gitiforouz, V. Ebrahimipour. Improved estimation of electricity demand function by integration of fuzzy system and data mining approach. *Energy Conversion and Management*. 49 (2008) 2165-77.
- [3] N. Mohamed, M. H. Ahmad, Z. Ismail, S. Suhartono. Double seasonal ARIMA model for forecasting load demand. *Matematika*. 26 (2010) 217-31.
- [4] Y. Ohtsuka, T. Oga, K. Kakamu. Forecasting electricity demand in Japan: A Bayesian spatial autoregressive ARMA approach. *Computational Statistics & Data Analysis*. 54 (2010) 2721-35.
- [5] P. Chujai, N. Kerdprasop, K. Kerdprasop. Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models. *Proc IMECS Conf, Hong Kong* 2013.
- [6] V. Bianco, O. Manca, S. Nardini. Electricity consumption forecasting in Italy using linear regression models. *Energy*. 34 (2009) 1413-21.
- [7] A. Al-Ghandoor, M. Samhouri. Electricity consumption in the industrial sector of Jordan: application of multivariate linear regression and adaptive neuro-fuzzy techniques. *JJMIE*. 3 (2009).
- [8] A. H. Alavi, A. H. Gandomi. A robust data mining approach for formulation of geotechnical engineering systems. *Engineering Computations*. 28 (2011) 242-74.
- [9] K. Ermis, A. Midilli, I. Dincer, M. Rosen. Artificial neural network analysis of world green energy use. *Energy Policy*. 35 (2007) 1731-43.
- [10] J. Catalão, S. Mariano, V. Mendes, L. Ferreira. Short-term electricity prices forecasting in a competitive market: a neural network approach. *Electric Power Systems Research*. 77 (2007) 1297-304.
- [11] A. Azadeh, S. Ghaderi, S. Tarverdian, M. Saberi. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Applied Mathematics and Computation*. 186 (2007) 1731-41.
- [12] H.-T. Pao. Forecasting electricity market pricing using artificial neural networks. *Energy Conversion and Management*. 48 (2007) 907-12.
- [13] Y. S. Murat, H. Ceylan. Use of artificial neural networks for transport energy demand modeling. *Energy Policy*. 34 (2006) 3165-72.
- [14] Z. W. Geem, W. E. Roper. Energy demand estimation of South Korea using artificial neural network. *Energy Policy*. 37 (2009) 4049-54.
- [15] S. Fan, L. Chen. Short-term load forecasting based on an adaptive hybrid method. *Power Systems, IEEE Transactions on*. 21 (2006) 392-401.
- [16] S. Fan, L. Chen, W.-J. Lee. Machine learning based switching model for electricity load forecasting. *Energy Conversion and Management*. 49 (2008)
- [17] Houshyar Asadi, S. Hr, et al. "Fuzzy Logic Control Technique in Li-Ion Battery Charger."
- [18] Houshyar Asadi, S. Hr, et al. "Fuzzy Control Based Five-Step Li-Ion Battery Charger by Using AC Impedance Technique."
- [19] Asadi, H., et al. "A Review on Li-Ion Battery Charger Techniques and Battery Charger Performance by Fuzzy Logic." *International Conference on Information and Intelligent Computing*. 2011.
- [20] Asadi, Houshyar, et al. "Fuzzy-control-based five-step Li-ion battery charger by using AC impedance technique." *Fourth International Conference on Machine Vision (ICMV 11)*. International Society for Optics and Photonics, 2012.
- [21] Sadati, N., et al. "Online optimal neuro-fuzzy flux controller for dtc based induction motor drives." *Applied Power Electronics Conference and Exposition, 2009. APEC 2009. Twenty-Fourth Annual IEEE. IEEE*, 2009.
- [22] Kaboli, S. Hr A., et al. "A hybrid adaptive Neural-Fuzzy tuned PI controller based Unidirectional Boost PFC converter feeds BLDC drive." *Power Electronics, Drive Systems and Technologies Conference (PEDSTC)*, 2013 4th. IEEE, 2013.
- [23] Mansouri, Mahdi, et al. "A hybrid Neuro-Fuzzy—PI speed controller for BLDC enriched with an integral steady state error eliminator." *Control System, Computing and Engineering (ICCSCE)*, 2012 IEEE International Conference on. IEEE, 2012.
- [24] Tavana, M., Fallahpour, A., Di Caprio, D., & Santos-Arteaga, F. J. (2016). A Hybrid Intelligent Fuzzy Predictive Model with Simulation for Supplier Evaluation and Selection. *Expert Systems with Applications*.
- [25] G. Zahedi, S. Azizi, A. Bahadori, A. Elkamel, S. R. Wan Alwi. Electricity demand estimation using an adaptive neuro-fuzzy network: a case study from the Ontario province—Canada. *Energy*. 49 (2013) 323-8.

- [26] Modiri-Delshad, Mostafa, et al. "An iterated-based optimization method for economic dispatch in power system." *Clean Energy and Technology (CEAT)*, 2013 IEEE Conference on. IEEE, 2013.
- [27] Modiri-Delshad, Mostafa, et al. "Economic dispatch in a microgrid through an iterated-based algorithm." *Clean Energy and Technology (CEAT)*, 2013 IEEE Conference on. IEEE, 2013.
- [28] Modiri-Delshad, Mostafa, et al. "Backtracking search algorithm for solving economic dispatch problems with valve-point effects and multiple fuel options." *Energy* 116 (2016): 637-649.
- [29] E. Assareh, M. Behrang, M. Assari, A. Ghanbarzadeh. Application of PSO (particle swarm optimization) and GA (genetic algorithm) techniques on demand estimation of oil in Iran. *Energy*. 35 (2010) 5223-9.
- [30] O. E. Canyurt, H. K. Ozturk. Application of genetic algorithm (GA) technique on demand estimation of fossil fuels in Turkey. *Energy Policy*. 36
- [31] A.R. Rafieerad, A.R. Bushroa, B. Nasiri-Tabrizi, S.H.A. Kaboli, S. Khanahmadi, Ahmad Amiri, J. Vadivelu, F. Yusof, W.J. Basirun, K. Wasa, Toward improved mechanical, tribological, corrosion and in-vitro bioactivity properties of mixed oxide nanotubes on Ti-6Al-7Nb implant using multi-objective PSO, *Journal of the Mechanical Behavior of Biomedical Materials*, Volume 69, May 2017, Pages 1-18, ISSN 1751-6161.
- [32] Kabolia, S. Hr Aghay, et al. "Controlling methods for power factor correction of storage devices connected into smart grid." *Clean Energy and Technology (CEAT)* 2014, 3rd IET International Conference on. IET, 2014.
- [33] Mansouria, M., et al. "Online adaptive power factor correction controller for DC-DC converters." *Clean Energy and Technology (CEAT)* 2014, 3rd IET International Conference on. IET, 2014.
- [34] Sebtahmadi, S. Sina, et al. "A 12-sector space vector switching scheme for performance improvement of matrix-converter-based DTC of IM drive." *IEEE Transactions on Power Electronics* 30.7 (2015): 3804-3817.
- [35] Mansouri, Mahdi, et al. "A review of single phase power factor correction AC-DC converters." *Clean Energy and Technology (CEAT)*, 2013 IEEE Conference on. IEEE, 2013.
- [36] Oladazimi, Maysam, et al. "A Review for Feature Extraction of EMG Signal Processing." *International Conference on Computer and Automation Engineering*, 4th (ICCAE 2012). ASME Press, 2012.
- [37] Fallahpour, A., Olugu, E. U., Musa, S. N., Khezrimotlagh, D., & Singh, S. (2014b). Supplier selection under fuzzy environment: a hybrid model using KAM in DEA. *Recent Developments in Data Envelopment Analysis and its Applications*, 342-348.
- [38] Fallahpour, A., Olugu, E. U., Moghassem, A., & Musa, S. N. (2014a). Ranking Alternatives in Rotor-Spun Knitting Process Using Extended VIKOR on Interval Data. *Journal of Engineered Fabrics & Fibers (JEFF)*, 9 (4).
- [39] Fallahpour, A., & Moghassem, A. (2012). Evaluating applicability of VIKOR method of multi-criteria decision making for parameters selection problem in rotor spinning. *Fibers and Polymers*, 13 (6), 802-808.
- [40] Fallahpour, A., & Moghassem, A. (2013a). Spinning preparation parameters selection for rotor spun knitted fabric using VIKOR method of multicriteria decision-making. *Journal of the Textile Institute*, 104 (1), 7-17.
- [41] L. Ekonomou. Greek long-term energy consumption prediction using artificial neural networks. *Energy*. 35 (2010) 512-7.
- [42] P.-F. Pai, W.-C. Hong. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*. 74 (2005) 417-25.
- [43] T. Xiong, Y. Bao, Z. Hu. Interval forecasting of electricity demand: A novel bivariate EMD-based support vector regression modeling framework. *International Journal of Electrical Power & Energy Systems*. 63 (2014) 353-
- [44] M. D. Toksarı. Estimating the net electricity energy generation and demand using the ant colony optimization approach: case of Turkey. *Energy Policy*. 37 (2009) 1181-7.
- [45] M. Duran Toksarı. Ant colony optimization approach to estimate energy demand of Turkey. *Energy Policy*. 35 (2007) 3984-90.
- [46] D. G. Lee, B. W. Lee, S. H. Chang. Genetic programming model for long-term forecasting of electric power demand. *Electric power systems research*. 40 (1997) 17-22.
- [47] J. H. Holland. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press 1975.
- [48] J. R. Koza. *Genetic programming: on the programming of computers by means of natural selection*. MIT press 1992.
- [49] Fallahpour, A., Olugu, E., Musa, S., Khezrimotlagh, D., & Wong, K. (2015). An integrated model for green supplier selection under fuzzy environment: application of data envelopment analysis and genetic programming approach. *Neural Computing and Applications*, 1-19.
- [50] M. Oltean, C. Grosan. A comparison of several linear genetic programming techniques. *Complex Systems*. 14 (2003) 285-314.
- [51] Fallahpour, A., Olugu, E. U., & Musa, S. N. A hybrid model for supplier selection: integration of AHP and multi expression programming (MEP). *Neural Computing and Applications*, 1-6.
- [52] Rafieerad, A. R., et al. "GEP-based method to formulate adhesion strength and hardness of Nb PVD coated on Ti-6Al-7Nb aimed at developing mixed oxide nanotubular arrays." *Journal of the mechanical behavior of biomedical materials* 61 (2016): 182-196.
- [53] Moghassem, A., & Fallahpour, A. (2011). Processing parameters optimization of draw frame for rotor spun yarn strength using gene expression programming (GEP). *Fibers and Polymers*, 12 (7), 970-975.
- [54] Fallahpour, A., & Moghassem, A. (2013b). Yarn Strength Modelling Using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Gene Expression Programming (GEP). *Journal of Engineered Fabrics & Fibers (JEFF)*, 8 (4).
- [55] Fallahpour, A., Amindoust, A., Antuchevičienė, J., & Yazdani, M. (2016). Nonlinear genetic-based model for supplier selection: a comparative study. *Technological and Economic Development of Economy*, 1-18.

- [56] C. Ferreira. Gene expression programming: a new adaptive algorithm for solving problems. arXiv preprint cs/0102027. (2001).
- [57] <http://www.sabteahval>.
- [58] www.tradingeconomics.com.
- [59] <http://www.tccim.ir>.
- [60] Kaboli, S. Hr Aghay, J. Selvaraj, and N. A. Rahim. "Long-term electric energy consumption forecasting via artificial cooperative search algorithm." *Energy* 115 (2016): 857-871.