Hybrid ANN and Artificial Cooperative Search Algorithm to Forecast Short-Term Electricity Price in De-Regulated Electricity Market

ALIREZA POURDARYAEI [1], **HAZLIE MOKHLIS** [1], **HAZLEE AZIL ILLIAS**[1],
S. HR. AGHAY KABOLI[2], **SHAMEEM AHMAD**[1], **AND SWEE PENG ANG**[3]

[1]Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia
[2]University of Malaya Power Energy Dedicated Advanced Centre (UMPEDAC), Level 4, Wisma Research and Development University Malaya (UM), Kuala Lumpur 59990, Malaysia
[3]Electrical and Electronic Engineering Programme Area, Universiti Teknologi Brunei, Mukim Gadong A BE1410, Brunei

Corresponding author: Hazlie Mokhlis (hazli@um.edu.my)

ABSTRACT Smart grid has evolved into a viable platform for participants of electricity market to effectively regulate their bidding strategies based on demand-side management (DSM) models ascribed to its immense technological advancements in recent years. Reliability of system operation as well as capital cost investments can improve greatly with responsiveness of market participants. In this regard, efficient design, implementation, evaluation of numerous demand response measures and development of robust short-term price forecasting in the day-ahead transactions are of the utmost importance. Accuracy and efficiency of the day-ahead price forecasting process are complex challenges in deregulated electricity market. The unstable nature of electricity price compared to load series causes lower accuracy. Therefore, this research proposes a hybrid method for electricity price forecasting via artificial neural network (ANN) and artificial cooperative search algorithm (ACS). In parallel, a feature selection technique based on the combination of mutual information (MI) and neural network (NN) is developed in this study to select the input variables subsets, which have substantial impact on forecasting of electricity price. Actual data sets are collected from Ontario electricity market of the year 2017 for the verification of simulation results. Finally, the simulation results validated the premise of the proposed hybrid method through enhanced accuracy compared to the results acquired by implementing hybrid support vector machine (SVM) and hybrid ANN optimization methods.

INDEX TERMS Artificial neural network, artificial cooperative search, deregulated electricity market, electricity price forecasting.

NOMENCLATURE

$\bar{Z}$	Normalized data
Z	Un normalized data
:=	Update operation
∂	Kernel's parameter
C	Soft margin parameter
υ	Fraction of error
TH	Threshold
t	The network output
n	The number of samples
y	The actual value
n_j	First hidden layer neuron
n_{Lj}	j^{th} neuron (Output signal) in hidden layer of total (N) neuron
$nPop$	Population size of host nests
$nVar$	Number of respective optimization variable
e(t)	The whiteness of estimated residuals
H(X)	Entropy of random variable X
P(X)	Probability function of X
P	Control parameter of ACS
H(X,Y)	Joint probability distribution of variable X and Y
Np	Number of nests
α_1, α_2 and β	Coefficient of penalty function
Pa	Probability of an alien egg

The associate editor coordinating the review of this article and approving it for publication was Hui Ma.

nf	Number of features are selected in first stage of feature selection
$EP(t)_{observed}$	actual electricity price at time t
$EP(t)_{forecasted}$	Predicted electricity price at time t
f	Activation function
P	Control parameter of ACS
U – statistic	Thiel's inequality coefficient
k or k'	Regression lines
hi	Actual output
$g+1$	Next generation
t_i	Predicted output
t	Hourly interval
h_{0j}	Weight assigned to the bias unit of jth neuron in output layer
h_{ij}	Connection strength between i^{th} neuron in last hidden layer and j^{th} neuron in output layer
L	Hidden layer
low_j	Lower search space limits of jth variable
M	Binary integer-valued matrix
R_m	Confirmed indicator
R_o^2	The squared correlation coefficient
$R_{o'}^2$	The squared correlation coefficient between experimental and predicted values
CSA	Cuckoo search algorithm
ACS	Artificial cooperative search
FSA	fish swarm algorithm
AI	Artificial intelligence
MLP	Multi-layer perceptron
ANFIS	Adaptive neuro-fuzzy inference system
AWNN	Adaptive wavelet neural network
BP	Back propagation
DE	Differential evolution
ED(t)	Electricity demand at time t
EP(t)	Electricity price at time t
PSO	Particle swarm optimization
RACF	Residuals autocorrelation function
RMSE	Root mean square error
CE	Conditional entropy
MI	Mutual Information
NL_{EP}	Number of lag order for electricity price
ARIMA	Auto regressive integrated moving average
permuting	Random shuffling function
a and b	Randomly generated number
randi	Random selection function
randperm	Random permutation function
U	Uniform distribution function
up_j	Upper limits of jth variable in search space
w_{0j}	Activation threshold
w_i	Initial weight
w_{ij}	Connection strength between neurons
$y_{i;\alpha}$	i^{th} sub-superorganism productivity related to α superorganism
$y_{i;\beta}$	i^{th} sub-superorganism productivity related to β superorganism

y_g	Global minimum
$/MW.h	Dollar per Megawatts hour
w_l^m	The weight connecting from l-th features to m-th hidden unit
w_m^o	The weight connecting from m-th hiden unit to network output
α	Learning rate
β	Distribution factor
ISO	Independent System Operator
BP	back propagation
PrNN	probability neural network
OED	orthogonal experimental design
PASA	Projected assessment of system adequacy
ANN	Artificial neural network
LSSVM	Least square support vector machine
WPT	wavelet packet transform
RBFNN	radial basis function neural network
VB	Virtual Budget
GA	Genetic algorithm
MAPE	Mean absolute percentage error
e (t)	he whiteness of estimated residuals
MW	Megawatts
NN	Neural network
EPF	Electricity Price Forecasting
SVM	Support vector machine
T&D	Transmission and distribution
DGs	Distributed generations
NL_{ED}	Number of lag order for electricity demand
HOED	Hourly Ontario electricity demand
HOEP	Hourly Ontario electricity price

I. INTRODUCTION

The recent electricity grids are expected to maintain high standards of quality to meet the growing diversity in demand whilst providing a constant and reliable supply. These intricate challenges are the driving force behind smart grid technologies to be constantly evolved and developed. Realization of the smart grid technologies faces complex challenges such as optimization of distributed generations (DGs) capacity, transmission and distribution (T&D) systems and efficient energy storage technologies. These challenges require extensive research and substantial investment. Therefore, electricity price forecast plays an instrumental role in today's advanced electricity market as well as smart grid operation. This forecasting helps every individual generator to determine the optimal bidding layout. Furthermore, decision of joint agreement and investment in a new generation facility in the long run are highly influenced by the price forecasting [1]. It is imperative to forecast electricity price for the generation companies or Independent System Operator (ISO) as well as different level of customers and investors. Basically, different bidders in competitive electricity market require the future electricity prices to gear up their profit. Since the recent energy markets are highly deregulated and nonlinear, the price forecasting has become more complex compared to previous days. Due to the nonlinearity and instability of

this system, the accuracy of price prediction has become lower [2]. Moreover, it leads to an explosive electricity market by affecting the bidding policies.

Due to the uncertainty nature of electricity market price, the supply and demand side managements are experienced with numerous difficulties in day-ahead electricity market [3]. The power suppliers may receive more privileges in their short-term prediction of their rational offers by knowing the preceding information of electricity market price variations [4]. Moreover, it helps the power suppliers to setup their bidding strategies to enhance their profit in maximum scale. On the other hand, it is very important for demand side management to have the knowledge of market price changes and variations to develop the short-term operational planning. Therefore, in recent years, the researches in electricity market for price forecasting have become more significant.

Forecasting techniques can be classified into three categories in accordance to the forecasting framework i.e. statistical models, time series methods and Artificial Intelligence (AI) based approaches. Among different techniques, AI-based forecasting approaches have gained significant traction in recent years as these approaches offer a remarkable advantage of assuring a certain level of estimation accuracy compared to high fluctuation of independent and dependent variables in the statistical model [5]. For instance, ANN was extensively adopted among different AI-based approaches in [6] to forecast electricity demand and price. However, there are criterions like convergence speed, weight adaptation algorithm and network architecture selection over which the ANN based methods' accuracy and robustness are heavily dependent. Support vector regression (SVR) was used in [7], [8] for electricity price forecasting as it is capable of adapting and encapsulating complicated relationships with the input data. Conversely, AI techniques have the capability to handle nonlinearity issues related to short term electricity price forecasting as these methods can remove different discriminators in complex environment, and from past experience they can recall, learn and store information which have made them popular in the area of electricity price forecasting.

Development of hybrid techniques has emerged as feasible approach to overcome the nonlinearity involved in short-term forecasting, which increases accuracy and improve reliability. In [9], another technique has been presented in which ARIMA procedure and wavelet transform (WT) are combined adopting the power market prediction. To be more precise, in this technique, first historical data is separated by wavelet transform. Then, the ARIMA technique is applied and the inverse wavelet transform is employed respectively to obtain the final prediction outcomes. In [10], ANN has been used to forecast the electricity price where the price and quantities are found by using the estimated future parameters and history. Exploiting fuel cost and demand as entry data, a three-layer back propagation (BP) neural network (NN) has also been introduced in this study. A combination of the probability neural network (PrNN) and orthogonal experimental design (OED) has been adopted in [11] to

forecast the power price. The PrNN and OED methods have been used for classification and locating the best variable respectively, which eventually increase the forecast precision. The support vector machine (SVM) and the projected assessment of system adequacy (PASA) have been used in [12] for prediction of price according to the price and load history, and the entry data respectively. The work has also used a regional data of south wales to carry out their evaluations.

In [13], the fish swarm algorithm (FSA) has been selected and used as a time series forecast procedure to select the SVM variable. The work used the power price as data entry. Wavelet packet transform (WPT) and feature selection have been combined in [14] to introduce least square support vector machine (LSSVM) algorithm for prediction purpose. A probabilistic power price prediction method by combining support vector regression (SVR) and ARIMA method has been proposed in [15], [16]. A combination of three methods as WT, radial basis function neural network (RBFNN) and ARIMA has also been used in [17]. In [18], another mixed model has been presented based on interaction of load and price prediction. Another model for price-directed demand response has been introduced in [19], which Virtual Budget (VB) approach is developed for couples' price and load prediction, and let automated morphing of a consumer's electricity demand. Another hybrid model has been proposed in [20] to predict load and price. The work applied a hybrid time-series and adaptive wavelet neural network (AWNN) to predict the price. In [21], different states of multi-block based forecast engine are applied for both price load forecasting purposes. In this work, mechanism of forecasting consists of multi-block neural network (NN) and was optimized by an intelligence algorithm to increase training time and forecasting capabilities. In addition, genetic algorithms, particle swarm optimization and cuckoo search technique have been proposed in [22], [23] to perform better NN training as network training is an important feature in ANN based price forecasting.

Though remarkable advancements in precise electricity forecasting have been achieved through ANN, SVR, ANFIS and hybrid techniques, a precise and more accurate method to enhance the accuracy of electricity price is still needed. Moreover, all the aforementioned research regarding electricity price forecasting performs well though the forecasting accuracy is still impoverished.

For instance, linearly structured time series are incapable of capturing the nonlinear patterns and the frequent regularity in underlying the data changes over time. Moreover, the input parameters for electricity price forecasting are carried out based on trial-and-error procedures and engineer experience. Few researches have been observed where AI techniques were employed for electricity price forecasting due to its excellent nonlinear modeling capability. Another vital aspect that enhances the accuracy and efficiency of the forecasting is the proper feature selection. However, the selection of proper features using the existing feature
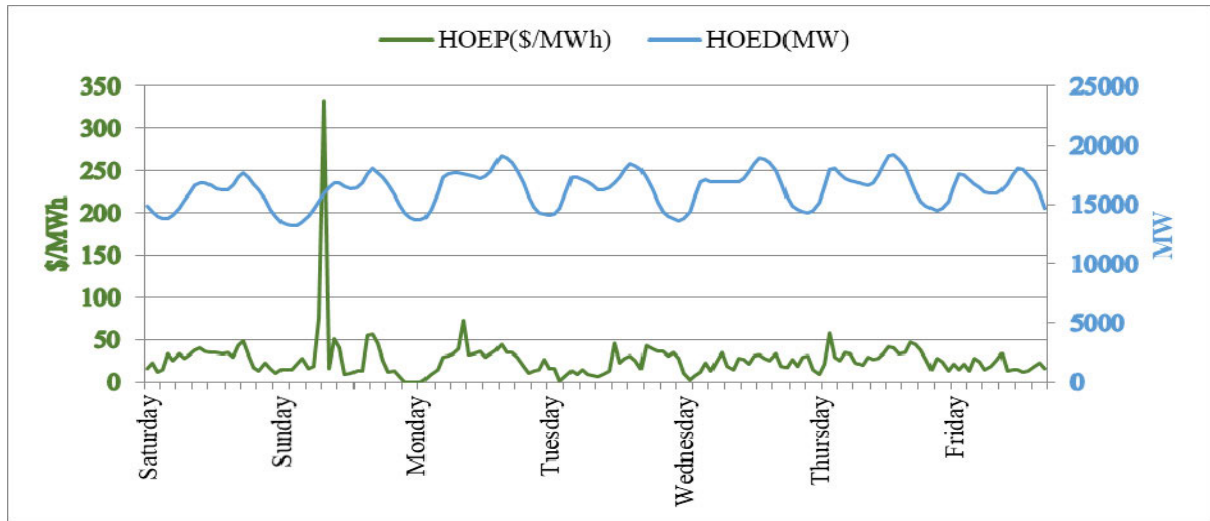
FIGURE 1. Ontario electricity market from 11.02.2017 to 17.02.2017.

selection techniques for electricity price forecasting by considering price non-linearity has been found to be difficult and complex task, which made it essential to explore enhanced feature selection techniques.

Therefore, in this research, to resolve the aforementioned issues, a new hybrid approach for short-term electricity price forecasting is proposed. The proposed approach is carried out through the combination of artificial cooperative search algorithm (ACS) and artificial neural network (ANN) to enhance the precision of the forecasting even more. ACS is a meta-heuristic algorithm, which has two-population. Unlike other metaheuristic methods, it has only a single parameter, which needs to be controlled and it does not possess high sensitivity towards the initial value of the control parameter. In addition, ACS operators exhibit high capabilities for exploitation of better results and problem exploration in search space due to its unique crossover and mutation process. Furthermore, mutual information and ANN techniques have been combined together to form a robust feature selection technique for enhancing the accuracy of the proposed hybrid ACS-ANN based price forecasting method. ANN has the potential of simulating data with nonlinear and complex relationships, which makes this approach preferable in the present study, since the feature selection will be based on the non-linear price signal.

To evaluate the accuracy of the methods proposed in this work, the Ontario electricity market was considered for the case studies. This is due the Ontario electricity market has been recognized as one of the unstable markets in the world as a result of its single settlement nature [24], [25]. Fig. 1 illustrates the relationship between hourly Ontario electricity price (HOEP) and hourly Ontario electricity demand (HOED) within one week. As shown in the figure, electricity price in Ontario electricity market is a function of electricity demand in the deregulated electricity market. There is a high competition for electricity price when electricity demand is very high and the generation is limited. Therefore, with the inherent correlation between electricity price and demand, prediction in smart grid environment such as Ontario electricity market is more complex than the conventional power systems. Hence, a novel prediction approach should be applied for this market to provide high accurate forecasts.

A summary of contributions of this research work is as follows:

➢ The main contribution of the study is to propose a hybrid electricity price forecasting technique based on efficient ACS algorithm along with ANN method for enhancing the accuracy of the price forecasting compared to existing forecasting methods. ACS has been used to search the most suitable biases and weights values of ANN for acquiring least error, which results in improved forecasting accuracy. Although ACS has simple structure, due to its effectiveness on solving multidimensional functions, it has been widely used in various numerical problems. Furthermore, only one factor is required to control in ACS algorithm.

➢ The second contribution of the work is related to the feature selection problem. To address this issue, a robust hybrid feature selection technique based on the combination of NN and mutual information techniques has been developed. Here, NN has been used to choose the best subset of features, where MI has been applied to extract input variables with minimum redundancy and maximum relevancy. By adding a penalty term to the error function of the network, redundant network connections can be distinguished from those relevant ones by their small weights when the network training process has been completed.

➢ Pertinence and precision of the proposed hybrid forecasting method are evaluated through comparison of
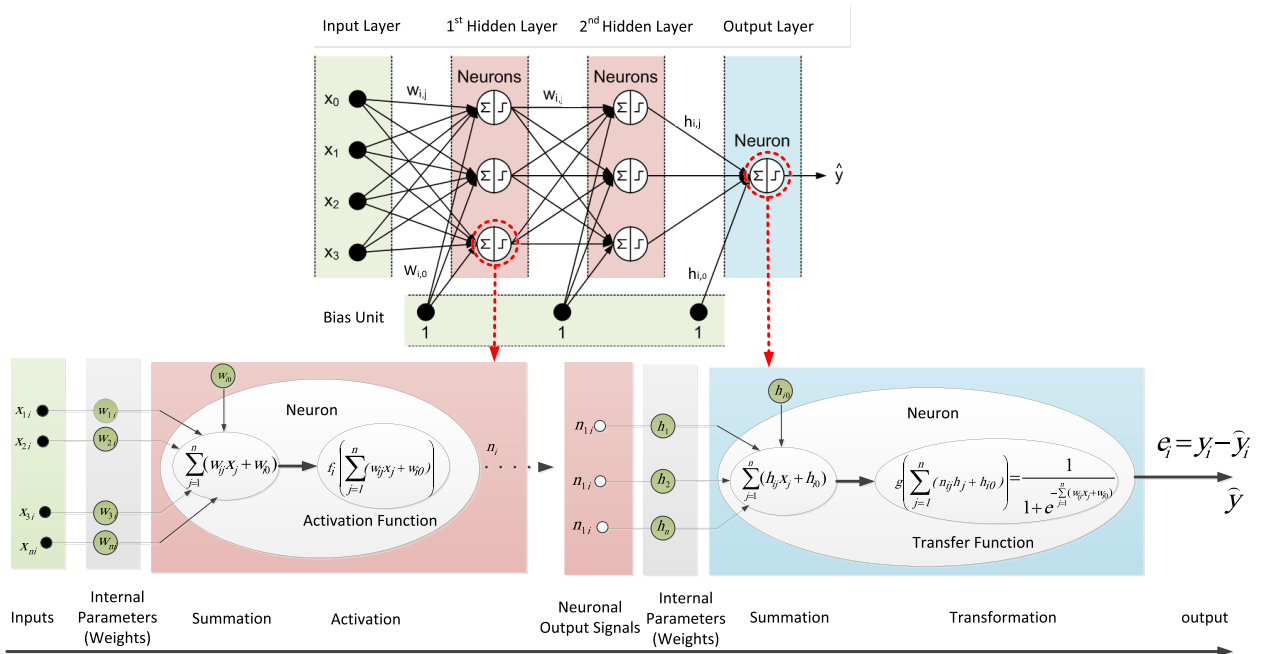
FIGURE 2. The MLP architecture.

achieved results with that of hybrid SVM and hybrid ANN methods, whose parameters are optimized by particle swarm optimization (PSO) and Cuckoo search algorithm (CSA).

Subsequent sections are organized as follows: Section 2 explains the short briefing on structure and evolution process of ANN and ACS respectively. Section 3 presents the development of short-term EP forecasting models. It also describes the proposed feature selection to provide the most influential features on short-term EP forecasting through filtering input variables using MI in first stage and developing neural network (NN) technique in second stage. Section 4 provides comprehensive discussion and results where statistical analysis is provided to ensure that the proposed approach is strongly suitable and applicable for future electricity price forecasting in de-regulated electricity market. Finally, section 5 concludes the work.

II. ARTIFICIAL NEURAL NETWORK (ANN)

Artificial neural networks (ANNs) are inspired by human brain consisting of millions of interconnected cells to interpret and process the information [26], [27]. Thus, a learning process will be simulated within the interconnected cell and collectively performs tasks that surpass the supercomputers with high-level computational capacity. As such, ANN is developed, which consists of a highly interconnected network known as neurons.

Fig. 2 shows the architecture of ANN, which consists of input layer, hidden layer and output layer that are interconnected to each other. In each layer, it contains several numbers of neurons. For example, in input layer, the number

of neurons is equal to the number of input data. For the output layer, the number of neurons represents the desired outputs or results. Whereas, in hidden layer, there is no limit to the number of neurons. However, these numbers of neurons will affect the quality of learning process, learning time and generalization process. The neurons are communicated with each other through the weighted connection consisting of weight and bias. In each neuron, the associated transfer function will interpret the input signal with respect to the output signal. Besides, the numerical weights and biases within the weighted connection are optimized in order to generalize and fit between the input and output data. Then, an accumulated experience and information during the ANN learning process is stored in the form of network file. The saved network file will have the capability to respond to the new input data, which does not involve during the learning process.

MLP networks are usually applied to perform supervised learning tasks, which involve iterative training methods to adjust the connection weights within the network. Generally, to achieve the preferred level of approximated accuracy, a number of passes is required. Standard error back-propagation algorithm is used for the adjustment of the correction weights, where the gradient decent method is applied to minimize the total error [28]. Back-propagation is a systematic method used for training MLP networks and its schematic diagram is briefly described in Table 1.

A. INITIALIZE WEIGHTS

The numerical or initial estimates of the linking strength between all neurons (w_{ij}) are allocated arbitrarily. Moreover, the activation threshold (w_{0j}) initial value is also randomly

TABLE 1. General structure of back-propagation algorithm.

1.   Initialize weights
Repeat
2.   Apply a sample
Calculate the output of each neuron:
Summation
Addition of bias
Activation
End
3.   Calculate the output errors
4.   Backpropagation of the error
5.   Update weights and biases
Until termination criteria are satisfied;

assigned to each neuron. This activation threshold is similar to an autonomous term of the linear combination of the outputs from the previous neurons. It is considered as a weight allotted to a fictitious neuron known as bias unit with an output value of 1. Therefore, the rule of the bias input (memory) is to shift the origin of activation function for better learning.

B. APPLY A SAMPLE
The input vector $\{X_1, X_2, X_3\ldots, X_k\}$ having desired output vector $\{y_1, y_2, y_3, \ldots, y_m\}$ is applied.

C. FEED-FORWARD COMPUTATION
Starting from the first hidden layer and propagating toward the output layer, each input unit (x_i) assigns initial weight (w_i) and broadcast this weight to all neuron in the first hidden layer.
i. Each neuron in the first hidden layer (n_j) sums its input weights by

$$n\_in_j = \sum_{i=1}^{k} w_{ij}X_i + w_{0j} \qquad (1)$$

ii. The activation function (f) process the output signal of each neuron (n_j) in hidden layers by

$$n_j = f\left(n\_in_j\right) \qquad (2)$$

Generally, the activation functions are one of the linear, logistic sigmoid and bipolar sigmoid (hyperbolic tangent) activation functions. The linear activation function is

$$n_j = f\left(n\_in_j\right) = n\_in_j \qquad (3)$$

In MLP networks, if the neurons have linear activation functions, the capabilities of the network is no better than a single layer network with linear activation function. Thus, nonlinear activation functions (sigmoid functions) are used, which usually limit the output signal of each neuron to the values between two asymptotes.

The logistic sigmoid function is the most widely used activation function in MLP, given by

$$n_j = f\left(n\_in_j\right) = \frac{1}{1 + e^{-(n\_in_j)}} \qquad (4)$$

The hyperbolic tangent function as formulated by Eq. (5) is another sigmoidal function used as activation function for neurons in hidden layer of MLP networks. The hyperbolic tangent function is closely related to the bipolar sigmoid function as

$$n_j = f\left(n\_in_j\right) = \frac{1 - e^{-(n\_in_j)}}{1 + e^{-(n\_in_j)}} = \frac{1 - e^{-(n\_in_j)}}{1 + e^{-(n\_in_j)}} \qquad (5)$$

iii. The output signal j^{th} neuron of total (N) neuron in hidden layer (L) denoted by (n_j^L) is transferred to next hidden layer as follows:

$$n_j^L = f\left(\sum_{i=1}^{N} w_{ij}n_i^{L-1} + w_{0j}\right) \qquad (6)$$

iv. The output signal of each neuron in the last hidden layer is propagated toward the output layer as follows:

$$\hat{y}_j = g\left(\sum_{i=1}^{m} h_{ij}n_i + h_{0j}\right) \qquad (7)$$

where (g), (h_{ij}) and (h_{0j}) are the activation function of output layer known as transfer function, the connection strength (weight) between i^{th} neuron in last hidden layer and j^{th} neuron in output layer, and the weight assigned to the bias unit of jth neuron in output layer respectively.

D. CALCULATE THE OUTPUT ERRORS
The error information term of each neuron in output layer is computed as follows:

$$\delta_j = \left(y_j - \hat{y}_j\right) g'\left(\sum_{i=1}^{m} h_{ij}n_i + h_{0j}\right) \qquad (8)$$

E. BACKPROPAGATION OF THE ERROR
The error backward is propagated to the input layer through each hidden layer using the error information term.

The backward weight correction term from output layer to last hidden layer and its bias correction term are computed by

$$\Delta h_{ij} = \alpha \delta_j n_i \qquad (9)$$
$$\Delta h_{0j} = \alpha \delta_j \qquad (10)$$

where α is the learning rate.

The error information term of each neuron in last hidden layer is calculated from multiplying the summation of its backward weights correction by derivative of its activation function as follows:

$$\Delta_j = \left(\sum_{i=1}^{m} \delta_j h_{ij}\right) \times f'\left(\sum_{i=1}^{N} w_{ij}n_i + w_{0j}\right) \qquad (11)$$

The backward weight correction term from the hidden layer (L) is sent to its hidden layer below (L-1) and its bias correction term are computed as follows:

$$\Delta w_{ij} = \alpha \Delta_j n_j^{L-1} \tag{12}$$

$$\Delta w_{0j} = \alpha \Delta_j \tag{13}$$

The error information term of each neuron in hidden layer is calculated as follows:

$$\Delta_{\delta j} = \left(\sum_{i=1}^{N} \Delta_j w_{ij} \right) \times f' \left(\sum_{i=1}^{N} w_{ij} n_j^{L-1} + w_{0j} \right) \tag{14}$$

The backward weight correction term from first hidden layer is sent to the input layer and its bias correction term are computed as follows:

$$\Delta_{\delta} w_{ij} = \alpha \Delta_{\delta j} X_i \tag{15}$$

$$\Delta_{\delta} w_{0j} = \alpha \Delta_{\delta j} \tag{16}$$

F. UPDATE WEIGHTS AND BIASES

Each neuron in the output layer updates its bias and weights by

$$h_{ij}(t+1) = h_{ij}(t) + \Delta h_{ij} \tag{17}$$

$$h_{0j}(t+1) = h_{ij}(t) + \Delta h_{0j} \tag{18}$$

Each neuron in hidden layer updates its bias and weights by

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij} \tag{19}$$

$$w_{0j}(t+1) = w_{ij}(t) + \Delta w_{0j} \tag{20}$$

Each neuron in first hidden layer updates its bias and weights by

$$w_{ij}(t+1) = w_{ij}(t) + \Delta_{\delta} w_{ij} \tag{21}$$

$$w_{0j}(t+1) = w_{ij}(t) + \Delta_{\delta} w_{0j} \tag{22}$$

Finding the optimal values of different weights and biases is achieved by training of the networks. Generally, to find the appropriate values of weights and biases of the ANN, different techniques are used. The optimum training of ANN is achieved by ACS.

III. ARTIFICIAL CORPORATIVE SEARCH ALGORITHM

The ACS is an algorithm that involves two-population search method and it is based on coevolution process. A few drawbacks of metaheuristic approaches have been overcome by this ACS; e.g. the parameters of initial values are over sensitive, have too many control parameters and when more time is required to compute between exploitation of better results and exploration of problem's search space [29].

ACS can be controlled by only one parameter and this is not sensitive to the initial value. In order to keep balance between exploration and exploitation, crossover and mutation are utilized in this algorithm. The functions of these operators are different in the crossover and mutation strategies than other methods, such as GA and DE. The advantage of ACS is that it has a memorization process to explore the feeding areas. It comprises of seven stages: initialization, selection of Predator, selection of Prey, mutation, crossover, boundary control, and export the best individual. Table 2 shows the generic structure of ACS.

IV. DAY-AHEAD SHORT-TERM EP FORECASTING

From the relation between electricity price and demand situations, it exhibits a time series spread over hourly intervals, which can be seen in a competitive electricity market. On the contrary, the price of electricity is a function of demand of electricity. The price of electricity depends on its present value t as well as electricity price's and demand's past values. It can be expressed as (30), as shown at the bottom of this page, where NL_{ED} is the electricity demand lag order, $EP(t)$ and $ED(t)$ are the price and demand of electricity at instantaneous time t assuming them as a time series and NL_{EP} is the lag order number for the price of electricity.

This work is focused on the implementation of ANN-ACS method for hourly Ontario electricity price (HOEP) forecasting. From [24], the input HOED and HOEP past data sets for the year of 2017 have been acquired. Historical data has variable range and Eq. (31) is used to normalize the independent and dependent variables. Normalizing specific data entails calibrating the data collected on distinct scales to an estimate common scale, usually applied before data processing.

$$\bar{Z}(t) = \frac{Z(t) - min(Z)}{max(Z) - min(Z)} + 1 \tag{31}$$

where normalized data is represented by $\bar{Z}$, the data to be normalized is represented by Z and the hourly interval is represented by t.

For the purpose of electricity price forecasting in this study, only one week of exogenous variables ($NL_{EP} = NL_{ED} = 168$) with hourly lagged values are assumed, where, total 336 exogenous variables lagged values are available. Overburdening the machine learning algorithms with excess amount of features results in a sluggish learning process, rendering a deplorable performance and overfitting the training data. Therefore, only features significantly affecting the output (for electricity price forecasting process) should be assigned to machine learning algorithm.

In the context of statistics and machine learning, feature (variables or predictors) selection, also known as attribute selection, variable selection or variable subset selection is a method for selecting a subset of relevant features in model

$$EP(t) = f \left(\begin{array}{c} EP(t-1), E\,P(t-2), E\,P(t-3), \ldots, E\,P(t-NL_{EP}) \\ ED(t), E\,D(t-1), E\,D(t-2), E\,D(t-3), \ldots, E\,D(t-NL_{ED}) \end{array} \right) \tag{30}$$

TABLE 2. General framework of ACS.

Step 1: Initialization	
The ACS algorithm consists of two superorganisms; α and β which have artificial sub-superorganisms equal to the population dimension. By using Eq. (23) the individual initial values of ith sub-superorganism belong to each superorganism (α and β) are defined.	$\alpha_{i,j;g=0} \sim U(low_j, up_j)$, $y_{i;\alpha} = f(\alpha_i)$ $\beta_{i,j;g=0} \sim U(low_j, up_j)$, $y_{i;\beta} = f(\beta_i)$ *for* $i = \{1,2,3,...,nPop\}$, $j = \{1,2,3,...,nVar\}$ (23) where, $nVar$ is respective optimization variable number; up_j and low_j are the jth variable's upper and lower search space limits; $nPop$ is the individuals population size in the superorganisms α and β; U is the constant distribution function; g is the generation number; $y_{i;\beta}$ and $y_{i;\alpha}$ are the ith sub-superorganism productivity linked to α and β superorganisms;

Step 2: Selection of Predator	
The Predator sub-superorganism of ACS algorithm is determined randomly through the 'if-then-else' decision rule from two superorganisms in each generation using Eq. (24).	$if\ a < b\Big|_{a,b \sim U(0,1)} then\ predator := \alpha,\ y_{predator} := y_\alpha,\ key := 1$ $else\ predator := \beta,\ y_{predator} := y_\beta,\ key := 2$ (24) where: key is a memory to trail in each iteration the origin of Predator; $:=$ is the update action; b and a are generated numbers randomly;

Step 3: Selection of Prey	
From two superorganisms, the arbitrary selection of Prey sub-superorganism will be executed using Eq. (25) in this step.	$if\ a < b\Big|_{a,b \sim U(0,1)} then\ prey := \alpha,\ else\ prey := \beta$ (25) $prey := permuting\ (prey)$

Step 4: Mutation	
The ACS method's mutation process will perform biological interaction of location x among Prey sub-superorganisms and Predator based on Eq. (26).	$x = predator + R.(prey - predator)$ $R = 4.a.(b\text{-}c)\Big|_{a,b,c \sim U(0,1)}$ (26)

Step 5: Crossover	
In the crossover process, the active individuals of the Predator indicated by a binary integer-valued matrix (M), are determined using Eq. (27).	$M_{i,j} = 1,$ $if\ a < (p.b)\Big|_{a,b \sim U(0,1)} then\ M_{i,randperm(randi(nVar))} = 0$ $else\ M_{i,randi(nVar)} = 0$ $if\ M_{i,j} > 0\ then\ x_{i,j} := predator_{i,j}$ *for* $i = \{1,2,3,...,nPop\}$, $j = \{1,2,3,...,nVar\}$ (27) where, $randperm(nVar)$ is arbitrary permutation function, P is the optimization algorithm control parameter, $randi(nVar)$ is arbitrary selection function.

Step 6: Boundary control	
In case of habitat limits are violated by obtained interaction location x, then boundary control will be used to update the related interaction location using Eq. (28).	$if\ (x_{i,j} < low_j)\ or\ (x_{i,j} > up_j)\ then\ x_{i,j} \sim U(low_j, up_j)$ (28)

Step 7: Export the global minimum:	
In the final step, as global minimum the best individual fitness value will be exported and based on Eq. (29) its position will be determined which will be considered as global minimizer.	$if\ f(x_i) < y_{i;predator}\ then\ predator_i := x_i,\ y_{i;predator} := f(x_i)$ (29) $if\ key = 1\ then\ \alpha := predator,\ y_\alpha := y_{predator}$ $else\ \beta := predator,\ y_\beta := y_{predator}$ $y_g = min(y_{predator})$ $if\ y_g < y_{g-1}\ then\ global\ minimum := y_g,\ global\ minimizer := predator_g,$ $g = g+1$ *for* $i = \{1,2,3,...,nPop\}$, $g = \{1,2,3,...,gMax\}$

construction. The objective of using feature selection techniques is a three-fold [30]:

1. Improving the prediction performance of the predictors by reducing overfitting (formally, reduction of variance).

2. Providing faster and more cost-effective process to construct the model (facilitate learning process).

3. Providing a simplified model that makes it easier to interpret (improving the generalization ability).

For electricity market price forecasting, mutual information (*MI*) technique has been broadly employed in [31]. However, this technique is facing difficulties due to the lagged values of the candidate input comprising of price, load demand and other variables provided by the electricity market. Thus, the individual probability distribution and the joint probability distribution of the candidate input are difficult to be obtained. Besides, it is noted that the electricity price is a time variant signal. Therefore, long history of the candidate input is not relevant to be used as the market conditions evolve every time. As such, it can mislead or give inaccurate price forecast process due to the lack of information values [32].

The main purpose of mutual information is to obtain the mutual correlation between two arbitrary variables X and Y, while, by using one variable information amount is achieved in this technique with regard to another variable which is random in nature. In other words, the mutual information is zero if variable X does not have any information related to variable Y and vice versa. As such, these two random variables are independent. High mutual information is obtained if variable X is a deterministic function of variable Y as well as variable Y as a deterministic function to variable X [33]. The link between *MI* and *CE* is depicted in Fig. 3. It can be seen that X and Y are closely related and dependent on each other when *MI* is large.

Apart of the entropy, conditional entropy (*CE*) is also observed. It is a measure of the average uncertainty of the first random variable after the second random variable. As shown in Fig. 3, $P_{XY}(X,Y)$'s joint probability distribution is used to achieve mutual information among X and Y and $MI(X,Y)$ random variables' as the random variable entropy is complexly related with the theme of mutual information.

Let $X = \{EP(t-1), EP(t-2), EP(t-3), \ldots, EP(t-NL_{EP}), ED(t), ED(t-1), ED(t-2), ED(t-3), \ldots, ED(t-NL_{ED})\}$ as a vector of all input features and $Y = EP(t)$ as the target or output feature. The electricity price and demand are denoted by *EP (t)* and *ED (t)* respectively. Fig. 3 shows a figure with the equation of computation of the mutual information (*MI*) between each individual input and output feature. For example, *MI {EP (t-1); ED (t)}* calculates the mutual information between the first lagged value of electricity price and electricity demand. Then, based on the calculated value
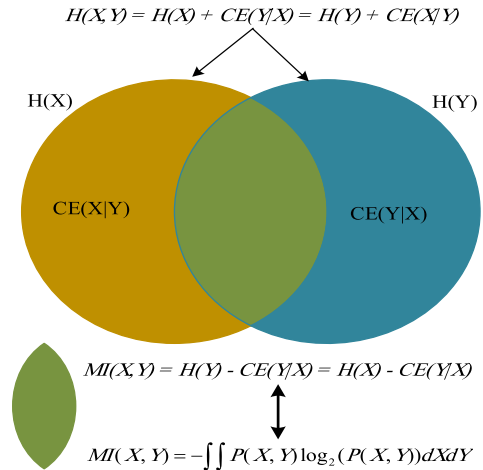

FIGURE 3. Mathematical relationship between conditional entropy and mutual information.

of MI, all the input features are sorted in descending order. Strong dependency between each variable of input and output is represented by higher value of MI. Lower MI value than the threshold *TH* is removed as it indicates less significant influence on the output while the remaining of the input features will form a subset, $SX \subset X$.

In the second stage, elimination of the redundant features is done and the neural network algorithm focusing on finding and removing those features is applied. Fig. 4 shows the utilization of the applied algorithm of three-layer feed forward neural networks. The input to the hidden layer and from the hidden to the output layer is weak for the redundant input. Therefore, it will be eliminated due to having less significant effect on the network accuracy. Commonly, the error function defined during the training process is defined as follows:

$$S = \frac{1}{n} \sum_{i=1}^{n} (t_i - y_i)^2 \tag{32}$$

where n represents the number of observation, t represents output of the network and y represents the real value. In order to detect irrelevant and redundant features, a penalty function is added to the error function as per Eq. (33), as shown at the bottom of this page, where α_1, α_2 and β are coefficients that control the influence of the penalty function. h is the number of hidden units, nf is the number of features selected in the first stage, is the weight connecting from *l-th* attribute to *m-th* hidden unit, and is the weight connecting from *m-th* hidden unit to network output.

The NN model will first evaluate the accuracy of the network, N using the set of input features $SX = \{X1, \ldots, X_{nf}\}$, $SX \subset X$, $nf < (NL_{EP} + NL_{ED})$. Then, the number of features

$$P(W) = \alpha_1 \left(\sum_{m=1}^{h} \sum_{l=1}^{nf} \frac{\beta(w_l^m)^2}{1 + \beta(w_l^m)^2} + \sum_{m=1}^{h} \frac{\beta(w_m^o)^2}{1 + \beta(w_m^o)^2} \right) + \alpha_2 \left(\sum_{m=1}^{h} \sum_{l=1}^{nf} (w_l^m)^2 + \sum_{m=1}^{h} (w_m^o)^2 \right) \tag{33}$$
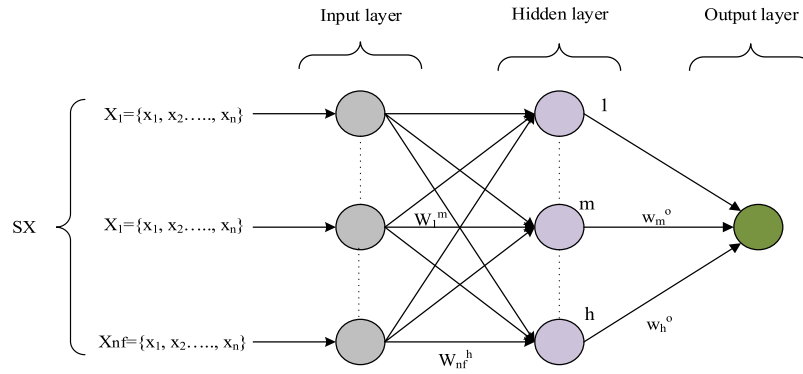
FIGURE 4. Second stage of feature selection based on the neural network model (NN).

is sequentially reduced to form a new set of input features and subsequently evaluate the accuracy of that network, N_k where $k = \{1, 2, \ldots, nf\}$. The accuracy of the network will be computed to determine the total number of features that can be eliminated. The steps of applied feature selection are outlined as follows:

1. Given the input vector $SX = \{X1, \ldots, X_{nf}\}$, $SX \subset X$ is divided into two data set, which are training set, SX_{tr} and testing set SX_{ts}. The network N is trained and the accuracy of the trained network is calculated for both SX_{tr} and SX_{ts}. In this algorithm, the value of α_1, α_2 and β are set to 10^{-1}, 10^{-4} and 0.03 respectively as mentioned in [34].

2. Suppose $SX = \{X1, \ldots, X_{nf}\}$ is the input features of the network, N_k does not include several nf features. For instance, $SX = \{4, 5, \ldots, nf\}$ are the input features of network N_3. The network N_3 is trained and the accuracy of the training and testing set, ASX_{tr} and ASX_{ts} respectively are calculated.

3. The network N_k is ranked based on the accuracy of the testing set as $ASX_{ts}^1 \geq ASX_{ts}^2 \geq \ldots \geq ASX_{ts}^{nf}$. Subsequently, the average value of this accuracy ASX_{ts}^{avg} is computed.

4. The process is continued by updating the penalty parameter. If the accuracy of the network N_k denoted by ASX_{ts}^{nf} is smaller than ASX_{ts}^{avg}, the weight values w_l^m and w_m^o are multiplied by 1.1. Otherwise, the values will be divided by 1.1. This allows significant input to have higher connection magnitude after the network is retrained. It is important to note that the input feature, which has weak connection magnitude, will be eliminated in this algorithm.

Feature selection technique is of utmost concern for selecting the important input variables. By selecting a feature, the contributions of the final predictor variables (HOED and HOEP in preceding hours) in the best NN model were evaluated. After developing and controlling several models with different combinations of input variables, these variables were identified. A hybrid feature selection is applied in order to reduce the running time. In the first stage of hybrid feature selection, relevancy threshold of TH = 0.46 has been chosen

for filtering the redundant features, and after filtering 60 relevant features are selected. In the second stage, (NN) is used in this study to select the input variables subsets, which have substantial impact on forecasting of electricity price. In this stage, hybrid MI and NN has been used to choose 31 dissimilar and most relevance features among the previously selected 60 candidates, which for the process of forecasting have been used as input. The description of the implemented feature selection method is presented in Fig. 5(a). The input variables subsets chosen by (MI + NN) as follows:

ED(t), EP(t-1), ED(t-1), EP(t-2), ED(t-2), EP(t-3), EP(t-23), EP(t-24), ED(t-24), EP(t-25), ED(t-25, EP(t-48), EP(t-49), EP(t-72), EP(t-73), ED(t-73), EP(t-96), EP(t-97, ED(t-97), EP(t-121), ED(t-121), EP(t-144), ED(t-144), EP(t-145), EP(t-168), ED(t-168), EP(t-169), EP(t-192), EP(t-193), EP(t-337), EP(t-505)

V. SIMULATION RESULTS AND DISCUSSION

In this paper, the electricity price forecasting (EPF) accuracy of Ontario mainland is forecasted by employing ANN-ACS, which is known as the one of the most volatile electricity market. Significant features are determined by developing a feature selection (MI + NN) as input for forecasting analysis in this section. Moreover, to evaluate the usefulness of ANN-ACS for short-term EPF accuracy, the proposed method is compared with well-known AI- techniques that include ANN-PSO, ANN-CSA, ANN, SVR-ACS, SVR-PSO, SVR-CSA and SVR. Fig. 5(b) shows the methodology used for forecasting the short-term electricity price. Sequential steps to obtain AI-based models for short-term EP forecasting are carried out for all models as follows:

Step 1: Considering one month in each season (i.e., winter (February), spring (May), summer (August) and autumn (November)), the efficacy of the applied methods has been evaluated for EPF in different seasons due to seasonal effects. The Ontario electricity market particular data (HOEP and HOED) for 2017 is selected by taking independent variables and HOEP as the dependent variable. Both of them consist of two subsets, at first for training of design phase, the first three
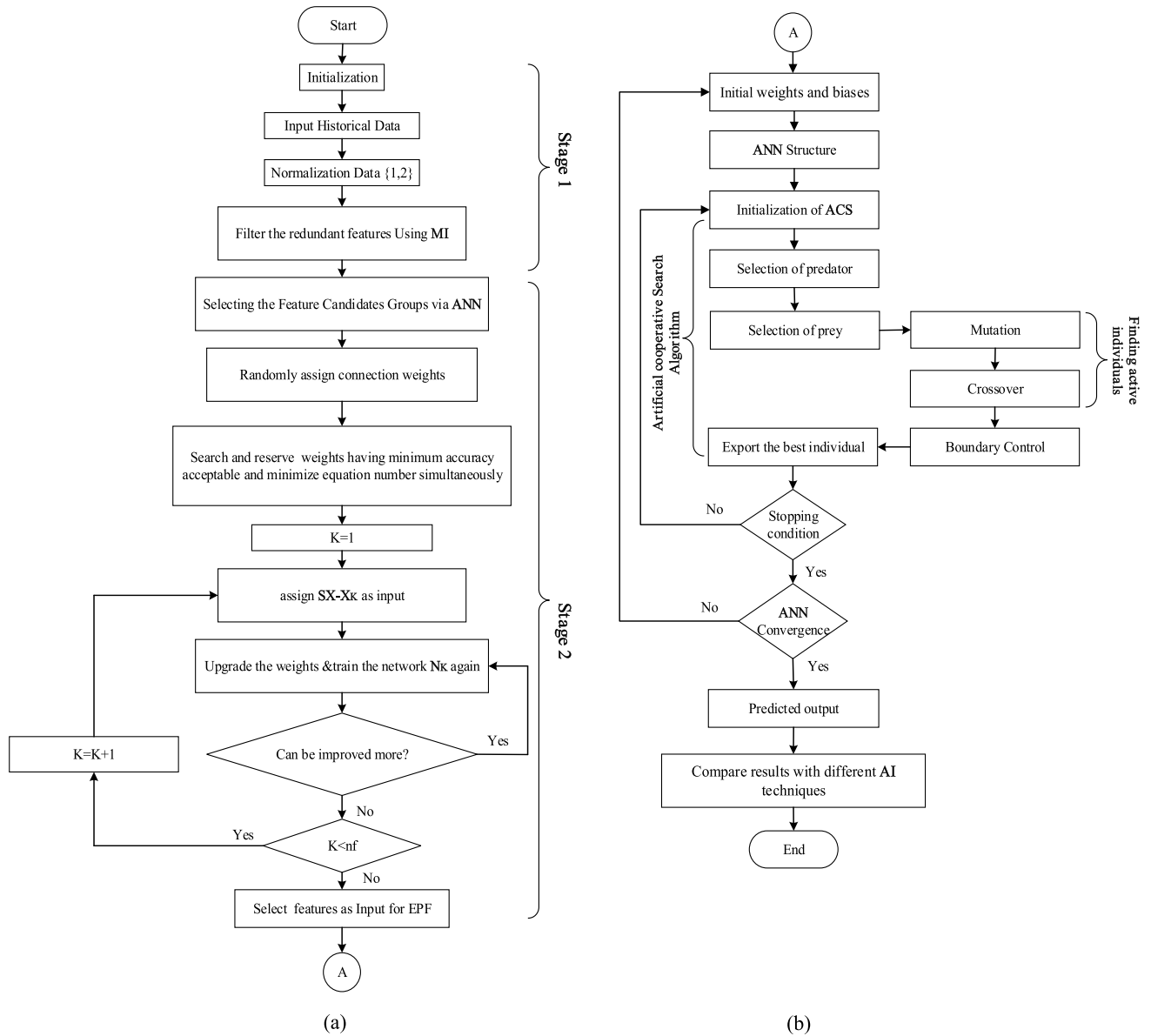
FIGURE 5. Framework of (a) 2-stages feature selection technique and (b) Forecasting of electricity price.

weeks hourly data is used and where for testing phase the last week data of each month is exploited.

Step 2: Designing training phase entail derivation of algorithms responsible for connecting the input variables to the output variables and Eq. (31) is used to normalize the input and output variables to make the learning process swift.

Step 3: To predict the electricity price (*EP*) precisely, metaheuristic methods is implemented for seeking the optimal coefficients of ANN and SVM by minimizing the cost function as follows:

$$F = \sum_{t=1}^{N} \left| (EP(t)_{observed} - EP(t)_{forecasted}) \right| \quad (34)$$

where $EP\ (t)_{observed}$ and $EP(t)_{forecasted}$ are the actual and predicted electricity price respectively and N represents the number of observation.

The coefficients of ANN and SVM models (*w*) are determined by

$$w = arg\ min\ F \quad (35)$$

Step 4: The purpose of designing a testing phase is to evaluate the model performance on the results of AI approaches applied on datasets having no function in building models. Various assessment criteria are applied to quantify the performance of the prediction models, such as root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), and Thiel's inequality coefficient (*U – statistic*) as:

$$MAPE\%$$
$$= \frac{1}{N} \sum_{t=1}^{N} \frac{\left| (EP(t)_{observed} - EP(t)_{forecasted}) \right|}{EP(t)_{observed}} \times 100 \quad (36)$$

TABLE 3. Parameter tuning of investigated methods.

Applied methods	Parameters setting			
ANN (**MLP**)	**Hidden layer**:2 **Transfer function**: logarithmic sigmoid **Learning algorithm**: Levenberg-Marquardt PB			
SVM (**SVR**)	**RBF kernel**	∂:1/6 **C**:1 υ:0.5		
ACS	**Number of individuals** = 100 **Control parameter** (P)= 0.15			
PSO	**Swarm population (Np)**=100 $w_{max} = 0.9, w_{min} = 0.4$ $c_1 = c_2 = 2$			
CSA	**Number of nests (Np)** = 100 **Probability of an alien egg (Pa)** = [0, 1] **Distribution factor (ß)** = 1.5			

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(EP(t)_{observed} - EP(t)_{forecasted})^2} \quad (37)$$

$$U = \frac{RMSE}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(EP(t)_{observed})^2} + \sqrt{\frac{1}{N}\sum_{t=1}^{N}(EP(t)_{forecasted})^2}} \quad (38)$$

U-statistic always generates binary results [0, 1], where zero represents higher forecasting precision and one represents estimation is as inaccurate as a naïve guess.

Appropriateness description of a given data series obtained through models is ensured through the whiteness test, also known as the Durbin-Watson test [26], acquired after a confirmatory analysis. The main objective of this confirmatory analysis is that it can confirm the whiteness of estimated residuals ($e(t)$) and also confirms the un-correlation between

them. Residuals autocorrelation function (*RACF*) is used to provide this calculation defined by:

$$RACF = \frac{\left|\sum_{t=2}^{N}(e(t)\,e(t-1))\right|}{\sum_{i=1}^{N}(e(t))^2} \quad (39)$$

Results of *RACF* are ranged between 0 and 1. *RACF* value falls outside a confidence level if it substantially differs from zero, implying un-correlation (whiteness) of residuals and hinting that a crucial independent variable has been excluded from the investigated model.

Successful methodologies in the literature of electricity price forecasting are taken under consideration to define the control parameters of the simulated methods due to highly problem dependent parameter setting of AI based methods and the lack of consensus regarding their optimal values in this study. All parameter settings of the applied methods and ANN-ACS model are summarized in Table 3.

The machine learning methods performances in winter season of 2017 for Ontario EPF is presented in Table 4. From the results tabulated in Table 4, it can be said that the whiteness of the estimated residuals for all developed models has been validated by the calculated *RACF* values, which are in an affirmed confidence range. Moreover, all developed models are able to describe the given set of data sufficiently. The analysis that has been done on Ontario electricity market regarding the evaluation of forecasting accuracy of methods concludes that according to multi-criteria decisions using the mean rank of the methods, each indicator (absolute error, *RMSE*, *U*-statistic and *MAPE*) is ranked as ANN-ACS > ANN-PSO = SVR-ACS > ANN-CSA > SVR-PSO > SVR-CSA > ANN > SVR. The comparison of the developed models with the existing similar models concludes that ANN-ACS approach performs exceptionally

TABLE 4. Comparison among the studied methods' electricity price forecasting accuracy in one month of Winter 2017 of Ontario.

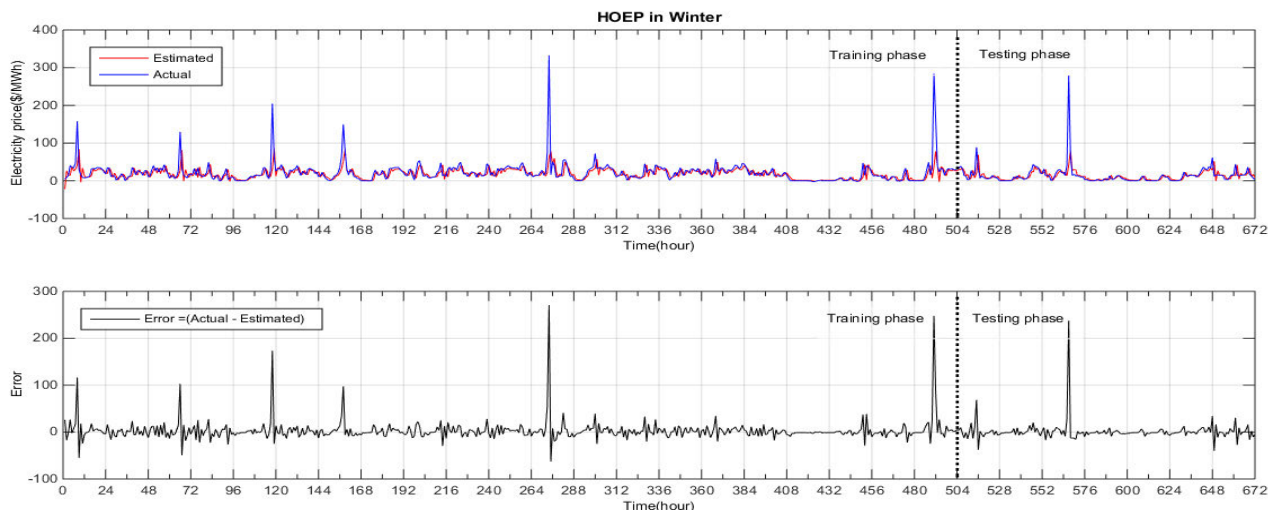| Performance Indexes		Methods							
		SVM	SVM-CSA	SVM-PSO	SVM-ACS	ANN	ANN-CSA	ANN-PSO	ANN-ACS
RACF	Training	0.0002	0.0033	0.0019	0.0024	0.0008	0.0007	0.0021	0.0032
	Testing	0.0011	0.0112	0.0009	0.0031	0.0012	0.0014	0.0015	0.0019
	Whole set	0.0017	0.0014	0.0016	0.0077	0.0014	0.0023	0.0004	0.0003
Absolute error	Training	21.7654	21.0076	19.9976	19.1654	21.5438	20.2265	19.7476	18.9976
	Testing	8.2789	8.0356	7.8795	7.3245	8.0234	7.7657	7.2358	7.0021
	Whole set	30.0443	29.0432	27.8771	26.4899	29.5672	27.9922	26.9834	25.9997
RMSE	Training	0.0912	0.0897	0.0884	0.0862	0.0902	0.0873	0.0867	0.0847
	Testing	0.0883	0.0879	0.0868	0.0853	0.0876	0.0862	0.0851	0.0840
	Whole set	0.0898	0.0884	0.0872	0.0860	0.0887	0.0870	0.0862	0.0845
U-statistic	Training	0.0492	0.0465	0.0460	0.0425	0.0483	0.0456	0.0429	0.0412
	Testing	0.0467	0.0458	0.0451	0.0431	0.0455	0.0447	0.0416	0.0405
	Whole set	0.0474	0.0462	0.0458	0.0429	0.0471	0.0450	0.0421	0.0410
MAPE (%)	Training	5.3481	5.1256	5.1124	4.7864	5.2346	5.0345	4.8564	4.6781
	Testing	5.1876	4.9981	4.9861	4.7641	5.0121	4.8719	4.6486	4.4917
	Whole set	5.2119	5.0237	4.9996	4.7775	5.1455	4.9228	4.7334	4.5760

FIGURE 6. ANN-ACS performance during training of testing phase and design phase and its corresponding error for EPF in one month of winter 2017 of Ontario.

TABLE 5. Comparison among the studied methods' electricity price forecasting accuracy in one month of Spring 2017 of Ontario.

Performance Indexes		Methods							
		SVM	SVM-CSA	SVM-PSO	SVM-ACS	ANN	ANN-CSA	ANN-PSO	ANN-ACS
RACF	Training	0.0027	0.0007	0.0009	0.0038	0.0038	0.0020	0.0021	0.0027
	Testing	0.0017	0.0025	0.0011	0.0005	0.0041	0.0007	0.0009	0.0038
	Whole set	0.0002	0.0012	0.0024	0.0022	0.0015	0.0019	0.0024	0.0003
Absolute error	Training	6.4791	6.4579	6.4660	6.4412	6.4734	6.4121	6.3125	6.1247
	Testing	2.4827	2.4106	2.4111	2.3916	2.5973	2.3974	2.2745	2.0542
	Whole set	8.9618	8.8685	8.8771	8.8328	9.0707	8.8095	8.5870	8.1789
RMSE	Training	0.0338	0.0324	0.0329	0.0320	0.0334	0.0315	0.0307	0.0294
	Testing	0.0276	0.0253	0.0258	0.0249	0.0281	0.0254	0.0248	0.0213
	Whole set	0.0319	0.0310	0.0317	0.0303	0.0326	0.0301	0.0272	0.0255
U-statistic	Training	0.0214	0.0200	0.0203	0.0197	0.0209	0.0189	0.0176	0.0167
	Testing	0.0115	0.0102	0.0106	0.0099	0.0118	0.0103	0.0091	0.0084
	Whole set	0.0195	0.0191	0.0194	0.0187	0.0199	0.0180	0.0168	0.0159
MAPE (%)	Training	1.3512	1.3130	1.3245	1.3017	1.3481	1.2849	1.2734	1.2656
	Testing	1.1405	1.1312	1.1368	1.1275	1.1456	1.1317	1.1268	1.1120
	Whole set	1.2855	1.2519	1.2592	1.2489	1.2887	1.2403	1.2189	1.2044

better. Table 4 also shows that the ANN-ACS based method is superior term of *MAPE* = 4.58 %, *U*-statistic = 0.04, *RMSE* = 0.08 and absolute error = 25.99. Fig. 6 presented the performance test of ANN-ACS method in winter season of 2017 when the training is executed for both testing and design phase.

Table 5 shows a tabulated performance of the applied machine learning techniques in Ontario region for EPF in spring season of 2017. According to the *RACF* values, the conclusion can be drawn that the estimated residuals of all obtained models are white at a confidence interval level. The accuracy of the methods is ranked as ANN-ACS > ANN-PSO > ANN-CSA > SVR-ACS > SVR-CSA > SVR-PSO > SVR > ANN, which is based on multi-criteria

decisions adopting the mean rank of the methods for each indicator (*U*-statistic (0.02), *MAPE*(1.2%), *RMSE* (0.03), and absolute error (8.18)). Hence, the most efficient model is ANN-ACS. Fig.7 illustrates the performance of ANN-ACS in spring season for the same region.

The performance results of different machine learning methods are tabulated in Table 6 for EPF for the same area in summer 2017. The *RACF* values, like in summer season, declare the whiteness of estimated residuals at a confidence interval level for all obtained models. The forecasting accuracy of these methods is ranked as ANN-ACS > SVR-ACS > ANN-CSA > ANN-PSO > SVR-CSA > SVR-PSO > SVR > ANN in terms of multi-criteria decisions, which the mean rank of the methods for each indicator (*U*-statistic,
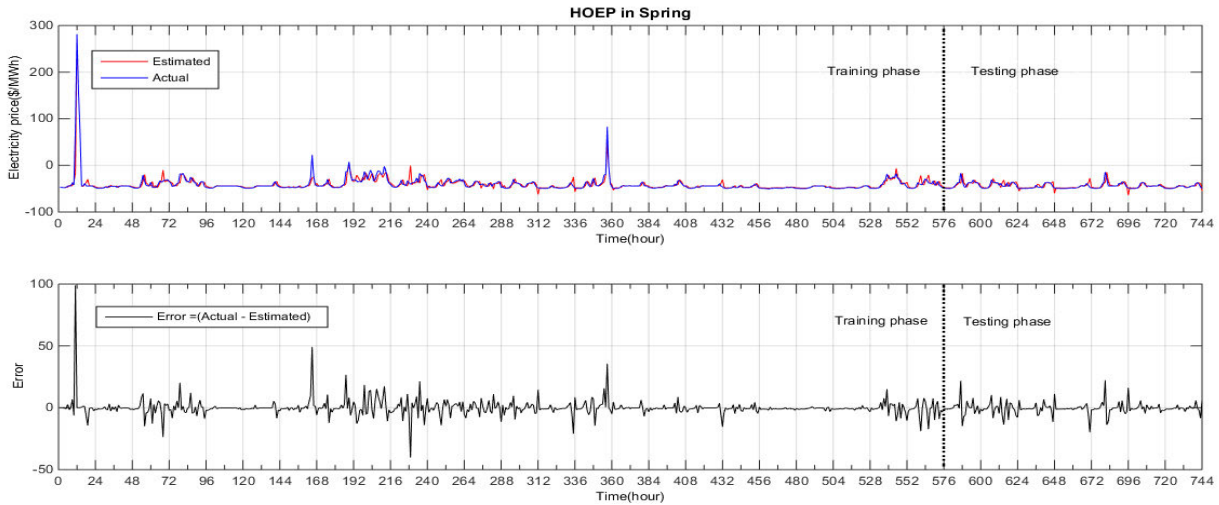
FIGURE 7. ANN-ACS performance during training of testing phase and design phase and its corresponding error for EPF in one month of spring 2017 of Ontario.

TABLE 6. Comparison among the studied methods' electricity price forecasting accuracy in one month of Summer 2017 of Ontario.

Performance Indexes		Methods							
		SVM	SVM-CSA	SVM-PSO	SVM-ACS	ANN	ANN-CSA	ANN-PSO	ANN-ACS
RACF	Training	0.0022	0.0015	0.0021	0.0002	0.0013	0.0006	0.0012	0.0023
	Testing	0.0030	0.0018	0.0010	0.0007	0.0018	0.0023	3.2E-6	0.0017
	Whole set	0.0019	0.0002	0.0006	0.0011	0.0025	0.0018	0.0014	0.0004
Absolute error	Training	17.9023	17.6621	17.8920	16.4623	18.1209	17.5581	17.5632	15.6175
	Testing	4.2757	4.2268	4.2561	4.0235	4.3213	4.1085	4.1955	3.9247
	Whole set	22.1780	21.8889	22.1481	20.4858	22.4422	21.6666	21.7587	19.5422
RMSE	Training	0.0736	0.0719	0.0727	0.0682	0.0748	0.0691	0.0708	0.0674
	Testing	0.0444	0.0425	0.0436	0.0395	0.0456	0.0402	0.0416	0.0380
	Whole set	0.0723	0.0706	0.0719	0.0673	0.0731	0.0685	0.0695	0.0659
U-statistic	Training	0.0398	0.0384	0.0392	0.0367	0.0410	0.0375	0.0379	0.0352
	Testing	0.0246	0.0235	0.0242	0.0211	0.0257	0.0213	0.0221	0.0194
	Whole set	0.0367	0.0353	0.0361	0.0324	0.0378	0.0336	0.0342	0.0308
MAPE (%)	Training	2.8496	2.7998	2.8312	2.7138	2.8534	2.7622	2.7834	2.6423
	Testing	2.7543	2.6866	2.7385	2.5941	2.7980	2.6330	2.6532	2.5514
	Whole set	2.7965	2.7651	2.7811	2.6895	2.8267	2.7254	2.7442	2.6241

MAPE, *RMSE*, and absolute error) is used to extract this ranking. To be more precise, *U*-statistic, *MAPE*, *RMSE* and absolute error are 0.03, 2.62%, 0.07 and 19.54 respectively for ANN-ACS. Additionally, ANN-ACS can be concluded as the most promising and advanced model than any other applied methods used for Ontario's short-term electricity price forecasting. The performance analysis of ANN-ACS for EPF in summer 2017 is depicted in Fig. 8 during both training and testing design phases for Ontario region.

The ANN-ACS performance for EPF of Ontario region in autumn season of 2017 is compared with other AI- methods for the purpose of further examination of solution methodology as shown in Table 7. From the values obtained for RACF, it can be shown that the estimated residuals of all models are uncorrelated and the obtained models sufficiently describe the given set of data. Based on Table 6, the forecasting accuracy of the applied methods is ranked as ANN-ACS > ANN-CSA > ANN-PSO > SVR-ACS > SVR-CSA > ANN > SVR-PSO > SVR. By using the mean rank of multi-criteria decisions methods, the ranking for each indicator (absolute error, *RMSE*, *U*-statistic and *MAPE*) in the whole set has been performed. According to the findings, from the comparison of studied methods for electricity market in Ontario region in autumn season, ANN- ACS approach performs much better than other methods in terms of electricity forecasting since it has higher precision.

The values of *U*-statistic (0.04), *MAPE* (3.79%), *RMSE* (0.08), and absolute error (29.73) are presented in Table 7.
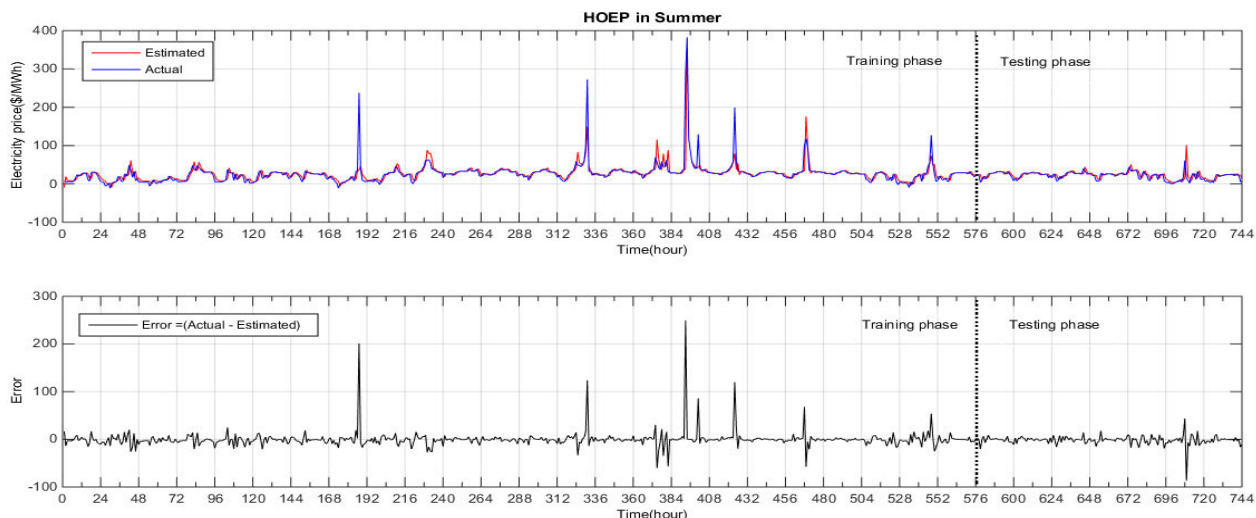
FIGURE 8. ANN-ACS performance during training of testing phase and design phase and its corresponding error for EPF in one month of summer 2017 of Ontario.

TABLE 7. Comparison among the studied methods' electricity price forecasting accuracy in one month of Autumn 2017 of Ontario.

Performance Indexes		Methods							
		SVM	SVM-CSA	SVM-PSO	SVM-ACS	ANN	ANN-CSA	ANN-PSO	ANN-ACS
RACF	Training	0.0025	0.0031	0.0027	0.0017	0.0003	0.0026	0.0021	0.0010
	Testing	0.0018	4.5E-6	0.0008	0.0006	0.0026	0.0011	0.0014	0.0004
	Whole set	0.0004	0.0013	0.0016	0.0005	0.0009	8.5E-6	0.0012	0.0015
Absolute error	Training	30.1020	29.6511	30.0013	28.4102	29.7746	25.1034	26.5052	24.5157
	Testing	8.4558	7.9803	8.2000	7.4699	8.0012	5.8814	6.7417	5.2168
	Whole set	38.5578	37.6314	38.2013	35.8801	37.7761	30.9848	33.2469	29.7325
RMSE	Training	0.0892	0.0830	0.0841	0.0824	0.0833	0.0805	0.0812	0.0801
	Testing	0.0820	0.0779	0.0813	0.0746	0.0805	0.0658	0.0681	0.0610
	Whole set	0.0867	0.0821	0.0835	0.0812	0.0827	0.0791	0.0802	0.0783
U-statistic	Training	0.0577	0.0546	0.0562	0.0530	0.0559	0.0458	0.0521	0.0427
	Testing	0.0383	0.0362	0.0375	0.0351	0.0366	0.0301	0.0345	0.0275
	Whole set	0.0549	0.0505	0.0524	0.0486	0.0518	0.0406	0.0463	0.0382
MAPE (%)	Training	4.6910	4.5013	4.5788	4.4436	4.5463	4.1279	4.2357	4.0324
	Testing	4.1015	3.9447	3.9934	3.6232	3.9721	3.3948	3.5468	3.0880
	Whole set	4.5968	4.2451	4.4319	4.1864	4.4170	3.9761	4.0145	3.7865

Fig. 9 depicts the ANN-ACS performance for forecasting the electricity price in autumn season of 2017 during the training and testing phase for Ontario region. From the demand side management perspective, the negative price can be controlled in whole seasons by providing incentive for using the electricity in a particular time to curtail the consumption accordingly when the demand is low in order to stabilize the frequency and voltage of grid.

The validity of mathematical models developed by ANN-ACS is verified through the application of different statistical methods as external validation. To evaluate the performance of the developed model, following attributes are recommended [36]–[39]:

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

II. If a model generates $0.2 < |R| < 0.8$, a correlation exists between the observed and predicted values.

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

The studied model's statistical factors for forecasting are computed for different month of Ontario mainland as well. The develop model fulfill all the statistical requirements as tabulated in Table 8. Findings indicate that the developed model is a promising and optimistic approach to be implemented for forecasting the future electricity price in de-regulated electricity market.
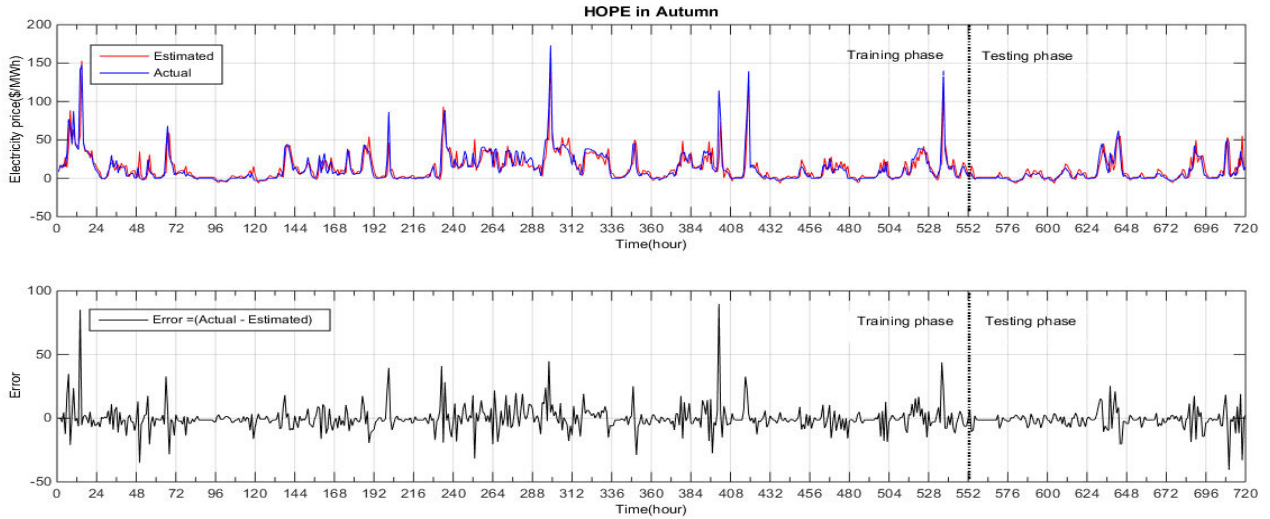
FIGURE 9. ANN-ACS performance during training of testing phase and design phase and its corresponding error for EPF in one month of autumn 2017 of Ontario.

TABLE 8. ANN-ACS model's Statistical factors for EPF in one month of Winter, Spring, Summer and Autumn 2017 of Ontario.

Item	Formula	Condition	ANN-ACS approach			
			Winter	Spring	Summer	Autumn
1	R	$0.8<R_0$	0.9987	0.9933	0.9850	0.9998
2	$K = \dfrac{\sum_{i=1}^{n}\left(h_i \times t_i\right)}{\sum_{i=1}^{n} h_i^2}$	$0.85<k<1.15$	0.9976	0.9985	0.9962	0.9970
3	$K' = \dfrac{\sum_{i=1}^{n}\left(h_i \times t_i\right)}{\sum_{i=1}^{n} t_i^2}$	$0.85<k'<1.15$	1.0030	1.0011	1.0023	1.0008
4	$m = \dfrac{R^2 - R_o^2}{R^2}$	$\lvert m \rvert <0.1$	-0.0033	-0.0039	-0.0042	-0.0024
5	$n = \dfrac{R^2 - R_o'^2}{R^2}$	$\lvert n \rvert <0.1$	-0.0033	-0.0039	-0.0041	-0.0024
6	$R_m = R^2 \times \left(1 - \sqrt{\left\lvert R^2 - R_o^2 \right\rvert}\right)$	$0.5<R_m$	0.9974	0.9923	0.9865	0.9980
Where	$R_o^2 = 1 - \dfrac{\sum_{i=1}^{n}\left(t_i - h_i^o\right)^2}{\sum_{i=1}^{n}\left(t_i - \bar{t_i}\right)^2}, h_i^o = k \times t_i$	$0.8<R_0^2<1$	0.9998	1.0000	1.0000	0.9996
	$R_o'^2 = 1 - \dfrac{\sum_{i=1}^{n}\left(h_i - t_i^o\right)^2}{\sum_{i=1}^{n}\left(h_i - \bar{h_i}\right)^2}, t_i^o = K' \times h_i$	$0.8<R_0'^2<1$	0.9998	1.0000	1.0000	0.9996

VI. CONCLUSION

In this research, for forecasting day ahead of electricity price, a hybrid ANN-ACS method has been developed, which has been verified with the Ontario electricity market based on hourly electricity demand and hourly electricity price.

- Feature selection is a fundamental element of machine learning algorithms. High predictive accuracy can be achieved through proper feature selection, which entails selection of relevant attributes of data amidst a vast pool of irrelevant and redundant features. In this work, by combining ANN and mutual information techniques, a hybrid feature selection technique has been proposed for the selection of optimum subset of features within a pool of 60 features, to be employed as input for direct

prediction method. The robustness of the proposed technique is evident through efficient selection of the most suitable features by removal irrelevant and redundant attributes.

- The results have shown the robustness of the developed ANN-ACS model in Ontario electricity market. In the case of electricity price forecasting, it provides a higher forecasting precision and simplicity compared to other AI methods in terms of $MAPE = 4.58\%$, 1.2%, 2.62% and 3.79 % in winter, spring, summer and autumn respectively.

- In developing the sustainable smart grid in forthcoming days, the importance of the proposed approach is inevitable to EP forecasting. Therefore, the presented synthesis can be a profound contrivance to develop energy strategies for the electricity participants in bidding as well as for the EP forecasting researchers plainly.

REFERENCES

[1] I. P. Panapakidis and A. S. Dagoumas, "Day-ahead electricity price forecasting via the application of artificial neural network based models," *Appl. Energy*, vol. 172, pp. 132–151, Jun. 2016.

[2] W. Gao, V. Sarlak, M. R. Parsaei, and M. Ferdosi, "Combination of fuzzy based on a meta-heuristic algorithm to predict electricity price in an electricity markets," *Chem. Eng. Res. Des.*, vol. 131, pp. 333–345, Mar. 2018.

[3] Z. Yang, L. Ce, and L. Lian, "Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods," *Appl. Energy*, vol. 190, pp. 291–305, Mar. 2017.

[4] W. Gao, H. Moayedi, and A. Shahsavar, "The feasibility of genetic programming and ANFIS in prediction energetic performance of a building integrated photovoltaic thermal (BIPVT) system," *Solar Energy*, vol. 183, pp. 293–305, May 2019.

[5] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1460–1495, Sep. 2014.

[6] F. Keynia, "A new feature selection algorithm and composite neural network for electricity price forecasting," *Eng. Appl. Artif. Intell.*, vol. 25, no. 8, pp. 1687–1697, Dec. 2012.

[7] H. Shayeghi and A. Ghasemi, "Day-ahead electricity prices forecasting by a modified CGSA technique and hybrid WT in LSSVM based scheme," *Energy Convers. Manage.*, vol. 74, pp. 482–491, Oct. 2013.

[8] H. Shayeghi, A. Ghasemi, M. Moradzadeh, and M. Nooshyar, "Simultaneous day-ahead forecasting of electricity price and load in smart grids," *Energy Convers. Manage.*, vol. 95, pp. 371–384, May 2015.

[9] N. Li-Xiao, W. A. N. G. Zheng-Fang, Z. A. N. G. Chuan-Zhi, S. Wen-Li, and Z. Sheng-Shan, "Hybrid model based on wavelet transform and ARIMA for short-term electricity price forecasting," *Appl. Res. Comput./Jisuanji Yingyong Yanjiu*, vol. 31, no. 3, p. 688, 2014.

[10] C. Hamzaçebi, "Improving artificial neural networks' performance in seasonal time series forecasting," *Inf. Sci.*, vol. 178, no. 23, pp. 4550–4559, 2008.

[11] W.-M. Lin, H.-J. Gow, and M.-T. Tsai, "Electricity price forecasting using enhanced probability neural network," *Energy Convers. Manage.*, vol. 51, pp. 2707–2714, Dec. 2010.

[12] D.-X. Niu, Q. Wang, and J.-C. Li, "Short term load forecasting model based on support vector machine," in *Advances in Machine Learning and Cybernetics*. New York, NY, USA: Springer, 2006, pp. 880–888.

[13] D.-S. Gong, J.-X. Che, J.-Z. Wang, and J.-Z. Liang, "Short-term electricity price forecasting based on novel SVM using artificial fish swarm algorithm under deregulated power," in *Proc. 2nd Int. Symp. Intell. Inf. Technol. Appl.*, Dec. 2008, pp. 85–89.

[14] H. Ebrahimian, S. Barmayoon, M. Mohammadi, and N. Ghadimi, "The price prediction for the energy market based on a new method," *Econ. Res.-Ekonomska Istraživanja*, vol. 31, no. 1, pp. 313–337, 2018.

[15] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong, "A hybrid approach for probabilistic forecasting of electricity price," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 463–470, Jan. 2013.

[16] A. Baziar and A. Kavousi-Fard, "Short term load forecasting using a hybrid model based on support vector regression," *Int. J. Sci. Technol. Res.*, vol. 4, pp. 189–195, May 2015.

[17] W.-Y. Chang, "Short-term load forecasting using radial basis function neural network," *J. Comput. Commun.*, vol. 3, no. 11, pp. 40–45, 2015.

[18] N. Amjady and A. Daraeepour, "Mixed price and load forecasting of electricity markets by a new iterative prediction method," *Electric Power Syst. Res.*, vol. 79, pp. 1329–1336, Sep. 2009.

[19] M. Alamaniotis, N. Gatsis, and L. H. Tsoukalas, "Virtual Budget: Integration of electricity load and price anticipation for load morphing in price-directed energy utilization," *Electr. Power Syst. Res.*, vol. 158, pp. 284–296, May 2018.

[20] L. Wu and M. Shahidehpour, "A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid," *IET Gener., Transmiss. Distrib.*, vol. 8, no. 12, pp. 1937–1950, 2014.

[21] W. Gao, A. Darvishan, M. Toghani, M. Mohammadi, O. Abedinia, and N. Ghadimi, "Different states of multi-block based forecast engine for price and load prediction," *Int. J. Elect. Power Energy Syst.*, vol. 104, pp. 423–435, Jan. 2019.

[22] M. K. Kim, "Short-term price forecasting of Nordic power market by combination Levenberg–Marquardt and Cuckoo search algorithm," *IET Gener., Transmiss. Distrib.*, vol. 9, no. 13, pp. 1553–1563, Oct. 2015.

[23] K. Sarada and V. Bapiraju, "Comparison of day-ahead price forecasting in energy market using neural network and genetic algorithm," in *Proc. Int. Conf. Smart Electr. Grid (ISEG)*, Sep. 2014, pp. 1–5.

[24] (Nov. 17, 2018). *IESO*. [Online]. Available: http://www.ieso.ca/

[25] H. Zareipour, K. Bhattacharya, and C. A. Cañizares, "Electricity market price volatility: The case of Ontario," *Energy Policy*, vol. 35, pp. 4739–4748, Sep. 2007.

[26] G. Landeras, J. J. López, O. Kisi, and J. Shiri, "Comparison of gene expression programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain)," *Energy Convers. Manage.*, vol. 62, pp. 1–13, Oct. 2012.

[27] M. Kankal, A. Akpınar, M. I. Kömürcü, and T. S. Özsahin, "Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables," *Appl. Energy*, vol. 88, pp. 1927–1939, May 2011.

[28] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 1352–1372, Oct. 2015.

[29] P. Civicioglu, "Artificial cooperative search algorithm for numerical optimization problems," *Inf. Sci.*, vol. 229, pp. 58–76, Apr. 2013.

[30] E. T. Renani, M. F. M. Elias, and N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," *Energy Convers. Manage.*, vol. 118, pp. 193–203, Jun. 2016.

[31] N. Amjady and M. Hemmati, "Energy price forecasting-problems and proposals for such predictions," *IEEE Power Energy Mag.*, vol. 4, no. 2, pp. 20–29, Mar./Apr. 2006.

[32] N. Amjady and A. Daraeepour, "Design of input vector for day-ahead price forecasting of electricity markets," *Expert Syst. Appl.*, vol. 36, pp. 12281–12294, Dec. 2009.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[34] R. Setiono and H. Liu, "Neural-network feature selector," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 654–662, May 1997.

[35] M. R. AlRashidi and K. El-Naggar, "Long term electric load forecasting based on particle swarm optimization," *Appl. Energy*, vol. 87, pp. 320–326, Jan. 2010.

[36] E. S. Mostafavi, S. M. Mousavi, and F. Hosseinpour, "Gene expression programming as a basis for new generation of electricity demand prediction models," *Comput. Ind. Eng.*, vol. 74, pp. 120–128, Aug. 2014.

[37] A. Golbraikh and A. Tropsha, "Beware of q^2!" *J. Mol. Graph. Model.*, vol. 20, no. 4, pp. 269–276, Jan. 2002.

[38] P. P. Roy and K. Roy, "On some aspects of variable selection for partial least squares regression models," *QSAR Combinat. Sci.*, vol. 27, no. 3, pp. 302–313, 2008.

[39] A. H. Alavi, P. Aminian, A. H. Gandomi, and M. A. Esmaeili, "Genetic-based modeling of uplift capacity of suction caissons," *Expert Syst. Appl.*, vol. 38, pp. 12608–12618, Sep. 2011.

ALIREZA POURDARYAEI received the bachelor's degree in electrical and electronic engineering from Hormozgan University, Iran, in 2010, and the M.Eng.Sc. degree in industrial electronic and control from the University of Malaysia (UM), in 2014, where he is currently pursuing the Ph.D. degree in power system. His current research interests include distribution automation, artificial intelligence (AI), energy management, and forecasting.

HAZLIE MOKHLIS (M'01–SM'18) received the B.Eng. and M.Eng.Sc. degrees in electrical engineering from the University of Malaya (UM), Malaysia, in 1999 and 2002, respectively, and the Ph.D. degree from The University of Manchester, Manchester, U.K., in 2009. He is currently a Professor with the Department of Electrical Engineering, UM, and the Head of the UM Power and Energy System (UMPES) Research. His research interests include fault location, distribution automation, power system protection, and renewable energy. He is also a Chartered Engineer in U.K., and a Professional Engineer in Malaysia.

HAZLEE AZIL ILLIAS received the bachelor's degree in electrical engineering from the University of Malaya (UM), Malaysia, in May 2006, and the Ph.D. degree in electrical engineering from the University of Southampton, U.K., in May 2011. He immediately joined Freescale Semiconductors (M) Sdn., Bhd., as a Product Engineer before pursuing his Ph.D. studies, in January 2008. Since January 2017, he has been an Associate Professor with the Department of Electrical Engineering, UM. He is also an Active Reviewer for numerous high quality journals, which include the IEEE TRANSACTIONS ON DIELECTRICS AND ELECTRICAL INSULATION, the IEEE TRANSACTIONS ON POWER DELIVERY, and *Measurement*. Since November 2011, he has been the Head of the University of Malaya High Voltage Laboratory and Research Group. His main research interests include modeling and measurement of partial discharge phenomena in solid dielectric insulation, lightning overvoltage, transmission line modeling, optimization methods, and artificial intelligence techniques. He is currently a member of IET (Institution of Engineering and Technology).

S. HR. AGHAY KABOLI received the B.Eng. degree in electrical engineering from the Islamic Azad University of Isfahan (IAU), Isfahan, Iran, in 2009, and the M.Eng. and Ph.D. degrees in power system engineering from the University of Malaya (UM), Kuala Lumpur, Malaysia, in 2012 and 2018, respectively. Since 2011, he has been a High Impact Research Assistant for Campus Network Smart Grid for Energy Security Project with UM, UM Power Energy Dedicated Advanced Center. His research interests include artificial intelligence (AI), smart grid, and power electronic. He is a Reviewer for the IEEE TRANSACTIONS, *Elsevier*, and *Springer journals*. He also has been serving as an editor for the American Journal of Electrical Power and Energy Systems, the International Journal of Computer Systems (IJCS), and the Guest Editor for the special issue *Some Novel Algorithms for Global Optimization and Relevant Subjects* under the *Journal of Applied and Computational Mathematics* (ACM).

SHAMEEM AHMAD was born in Chittagong, Bangladesh, in 1985. He received the B.E. degree in electrical and electronics engineering from Visveswaraiah Technological University (VTU), Belgaum, India, in 2009, and the master's degree in electrical engineering from the University of Malaya, Malaysia, in 2014, where he is currently pursuing the Ph.D. degree. His research interests include FACTS controllers, power system stability, load flow, and artificial intelligence.

SWEE PENG ANG was born in Bruneik, Darussalam. He received the B.Eng. degree (Hons) in electronic and electrical engineering from the University of Glasgow, U.K., in 1995, the M.Sc. degree in electrical power engineering from the University of Manchester Institute Science and Technology (UMIST), U.K., in 2002, and the Ph.D. degree from the University of Manchester, U.K., in 2010. He is currently a Principal Lecturer in electrical and electronic enginnering programme area with Universiti Teknologi Brunei, Negara Brunei Darussalam. His research interests include transformer modeling, ferroresonance studies in power systems, solar power systems, and control in power systems.

• • •