

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

# An efficient framework for short-term electricity price forecasting in deregulated power market

Alireza Pourdaryaei<sup>1,2</sup>, Mohammad Mohammadi<sup>1</sup>, MunirAzam Muhammad<sup>3</sup>, Junaid Bin Fakhru Islam<sup>4</sup>, Mazaher Karimi<sup>5,\*</sup> and Amidaddin Shahriari<sup>6</sup>

<sup>1</sup>Department of Power and Control, School of Electrical and Computer Engineering, Shiraz University, Shiraz, 7194684334, Iran

<sup>2</sup>Department of Electrical and Computer Engineering, University of Hormozgan, Bandar Abbas, 7916193145, Iran

<sup>3</sup>Department of Electrical Engineering, Main Campus, Iqra University, Karachi 75300, Pakistan

<sup>4</sup>Department of Electrical Engineering, Faculty of Engineering, Univeristi Malaya, Kuala Lumpur, 50603, Malaysia.

<sup>5</sup>School of Technology and Innovations, University of Vaasa, Wolfintie 34, 65200 Vaasa, Finland

<sup>6</sup>Department of Electrical Engineering, Islamic Azad University, South Tehran branch, Tehran.

Corresponding author: Mazaher Karimi (e-mail: mazaher.karimi@uwasa.fi).

This research has been supported by University of Vaasa under Profi4/WP2 project with the financial support provided by the Academy of Finland.

**ABSTRACT** It is widely acknowledged that electricity price forecasting become an essential factor in operational activities, planning, and scheduling for the participant in the price-setting market, nowadays. Nevertheless, electricity price became a complex signal due to its non-stationary, non-linearity, and time-variant behavior. Consequently, a variety of artificial intelligence techniques are proposed to provide an efficient method for short-term electricity price forecasting. BSA as the recent augmentation of optimization technique, yield the potential of searching a closed-form solution in mathematical modeling with a higher probability, obviating the necessity to comprehend the correlations between variables. Concurrently, this study also developed a feature selection technique, to select the input variables subsets that have a substantial implication on forecasting of electricity price, based on a combination of mutual information (MI) and SVM. For the verification of simulation results, actual data sets from the Ontario energy market in the year 2020 covering various weather seasons are acquired. Finally, the obtained results demonstrate the feasibility of the proposed strategy through improved preciseness in comparison with the distinctive methods.

**INDEX TERMS** Backtracking search Algorithm, Electricity Market, Electricity Price Forecasting, Feature Selection, Support Vector Machine.

## NOMENCLATURE

$y_{m;x_g^m}$	Fitness of $(x_g^k)$	$b$	regression function intercept
$(\xi_m, \xi_m^*)$	Slack variables	$Z$	Un normalized data
$(\alpha_m, \alpha_m^*)$	Nonnegative Lagrange multipliers	$\bar{Z}$	Normalized data
$HOEP(t)_{Actual\ value}$	Observed value of EP at time t hour	$mixrate$	Control parameter of BSA
$HOEP(t)_{Estimated\ value}$	Forecasted value of EP at time t hour	$oldP$	Historical population
$rand/randi$	Distributed random numbers /Random selection function	$\varphi(x_k)$	Kernel function
$J$	Soft margin parameter	$:=$	Update operation
$y_j$	Minimum value of the $j^{th}$ objective function	$P_g$	Productivity of $i^{th}$ individual
$up_j$	Upper search space limits of $j^{th}$ variable	$low_j$	Lower search space limits of $j^{th}$ variable
$nPop$	Population size of host nests	$nVar$	Number of respective optimization variable
$Mutant$	Initial form of trial population	$U$	Uniform distribution function
$P_{best}$	Previous best position	$N$	Standard normal distribution
$v$	Fraction of error	$CE$	Conditional entropy
$\langle W, x_m \rangle$	Vector inner product of the predictors	$AI$	Artificial intelligence
$U$ – statistic	Thiel's inequality coefficient	$HOEP$	Hourly Ontario electricity price
$k$ or $k'$	Regression lines	$a$ and $b$ in BSA	Randomly generated numbers
$hi$	Real output	$y_g$	Global minimum

$t_i$	Predicted output	$T$	Generated offspring at the end of crossover process
Rm	Confirmed indicator	$NL_{HOEP}$	Number of lag order for electricity price
Ro2	The squared correlation coefficient	ED(t)	Electricity demand at time t
Ro'2	The squared correlation coefficient between experimental and predicted values	EP(t)	Electricity price at time t
$\bar{y}_m$	Estimated output of the regression function	HOED	Hourly Ontario electricity demand
$(\beta_n, \beta'_n)$	Lagrangian multipliers	EPF	Electricity Price Forecasting
$b$	weight vector in SVR	SVM	Support vector machine
$X$ and $Y$	Random variables	$NL_{HOED}$	Number of lag order for electricity demand
$\Gamma(\cdot)$	Gamma distribution function	<i>permuting</i>	Random shuffling function
MAPE	Mean absolute percentage error	MW	Megawatts
$\hat{\sigma}$	Adjustable parameter in Gaussian RBF	RACF	Residuals autocorrelation function
$F$	Wiener process	RMSE	Root mean square error
$g$	Transfer function	SVR	Support vector regression
$g+1$	Next generation	MI	Mutual Information
$g_{best}$	Overall best value	AI	Artificial intelligence
<i>map</i>	Binary integer-valued matrix	H(X)	Individual entropy
$\$/MW.h$	dollar per Megawatts hour	H(X,Y)	Joint entropy
e(t)	The whiteness of estimated residuals	CE	Conditional entropy

## I. INTRODUCTION

Recently power grids are required to sustain the high-quality standards to meet increasing variation in demand and ensure a reliable and consistent supply. These complex issues seem to be the driving factor beyond the smart grid technologies that are continually evolving and developing. Smart grid technology implementation faces numerous challenges including optimizing the distributed generation (DG) sizing, distribution, and transmission (D&T) networks, and technology of energy storage efficiency. These complexities necessitate a rigorous analysis and significant financial expenditure. Hence, nowadays in modern electricity market similarly as smart grid activities, the forecasting on the price of electricity has become the major role. The forecasting allows each generator to decide the optimal bidding layout. In addition, mutual agreement decisions and long-term investment in the new-developed facility of generation are strongly affected by the price prediction [1]. Electricity price forecasting is essential for utility developer or Independent System Operator (ISO) similarly to the consumer and the investors. Essentially, various bidders in the competitive energy sector requires future electricity rates to maximize their revenue. The price prediction has become increasingly complicated in relation to previous years, as the latest electricity markets are particularly deregulated and non-linear. The price forecast accuracy has deteriorated because of the nonlinearity and volatility of this system. It also contributes to an explosive in the energy market by impacting the policies of bidding. Because of the incertitude of the market price of energy, various challenges are faced by supply-side and demand-side management on the day-to-day power market [2]. The suppliers of energy can obtain more rights by understanding the prior information on the price variations of the electricity market by the short-term prophesy of the sensible offers. Furthermore, it assists the power suppliers in developing

their strategies of bidding in order to save the industry and energy economy from wasting a huge amount of dollars that maximize their benefit on a large scale [3]. On the other hand, demand side management needs to be aware of market price fluctuations and differences to improve the planning of the short-term operation. Consequently, research has become more relevant in recent years on the electricity market for price forecasts.

Predicting techniques in conjunction with the forecasting context can be categorized under three groups –, Artificial Intelligence (AI) based method, statistical models, and time series techniques. In comparison with the statistical model of high level of independent and dependent variables, amongst various methods, AI-based electricity pricing approaches such as ANFIS, ANN (artificial neural network), CNN (conventional neural network) in recent years have gained considerable momentum in recent years because they have a great opportunity in ensuring a certain degree of precision in estimation the price [4-6]. The typical techniques to forecast electricity demand and price, amongst various AI-based approaches in [7], ANN has been widely deployed. Another technique for the electricity price forecasting is SVR (Support Vector Regression), which is efficient in adaptation and encapsulation of intricate relationships with data entry, normally utilized is shown in [8, 9].

Advancement of hybrid technology was developed to improve accuracy and practicality in order to address the nonlinearity inherent in short-term forecasts. Furthermore, by deploying the power market prediction, a different technique was introduced, which combines a Wavelet Transformation and ARIMA process (WT) in [10]. The approach depicted the wavelet transformation used to separate the first historical data, meanwhile to form the finishing prediction results, the ARIMA method has adopted the reverse wavelet transformation to implement the process. Another research contribution [11] has forecast prices using ANN technology

to identify prices and quantities by using past and estimation on future parameters. A three-layer back propagation (BP) neural network (NN) represents the contribution by taking the demand and the cost of fuel as entry data on the market. At the same time a combination of the orthogonal experimental design (OED) and probability neural network (PrNN) is utilized explained in [12]. For grouping and identifying the best variable, the PrNN and OED methods were used, which resulted in increased forecast accuracy.

In [13], the support vector machine (SVM) and estimated appraisal of the device adequacy (PASA) have been utilized to forecast price based on price and load history, as well as entry data. To carry out their assessments in this work, regional data from South Wales have also been deployed. To pick the SVM attribute, the fish swarm algorithm (FSA) was chosen and applied as a time series forecast method in [14]. The power price has been used as a data entry method in this work. By merging wavelet packet transform (WPT) and feature selection, the least square support vector machine (LSSVM) for forecasting been introduced in [15]. In [16] and [17], proposed a probabilistic power price prediction approach that is combined with SVR and ARIMA techniques. In [18], three combined methods recognize as WT, radial basis function neural network (RBFNN), and ARIMA have been utilized. A mixed model based on load-prediction interactions is presented in [19]. For price directed demand response, another method the Virtual Budget (VB) was introduced in [17], which provides price and load prediction and allows automatic morphing of electricity demand for a consumer. To forecast price and load, a hybrid model was proposed in [20]. To predict the price, the researchers used a hybrid time-series and adaptive wavelet neural network (AWNN). Besides, authors in [21] develops an optimal model based on LSTM-NN to predict electricity load and price. Distinct states of a multi-block based forecast engine are used in [22] for price prediction and both price load prediction. The forecasting mechanism in this study is a multi-block neural network (NN) that has been optimized by an intelligence algorithm to maximize the training time as well as improve the capabilities of forecasting. Other methods to model the price and forecast the demand have been suggested by [23, 24].

Though ANN, SVR and hybrid approaches have made significant advances in the precise electricity price forecasting technique, a more perfect and precise method to improve the exactness of price forecasting is yet required. Furthermore, all of the aforesaid research on electricity price forecasting works admirably since the accuracy of those forecasting methods are still improvised. For example, a linearly organized time series are incompetent in expressing nonlinear patterns and frequent regularity in data variations over the time. Moreover, based on the engineer's experience and trial and error procedure, the electricity price forecasting input parameters are determined. Because of having nonlinear model capability of AI techniques, few research

applied AI techniques to forecast electricity price forecasting. Appropriate feature selection is another important prospect in price forecasting since it enhances the efficiency and accuracy of the model. However, by applying the existing feature selection methods, the appropriate feature selection of the electricity price forecasting with considering the non-linearity of the price has been regarded as a difficult and complex task that essentially forced to investigate an improved enhanced feature selection technique.

Therefore, to solve the aforementioned issues in electricity price forecasting, a new hybrid approach is proposed in this research. The proposed method is combined with backtracking search algorithm (BSA) and support vector regression (SVR) to improve the precision of the forecasting. BSA is a metaheuristic technique consists of two population. Dissimilar to other metaheuristic techniques, BSA has just one parameter that has to be regulated and it does not have a high sensibility to the control parameter's starting value. Moreover, with the property of unique crossover and mutation process, BSA has high capability of problem exploration in the search space and exploitation of better result. Furthermore, the combination of mutual information and SVM methods form a new feature selection method which is robust and enhances the price forecasting technique. In addition, the feature selection based on the non-linear price signal, SVM has the ability to simulate data with nonlinear and complicated connections, making it a better fit for the present study.

The Ontario electricity market is being used as a case study to test the feasibility of the suggested methods. Because of its single settlement existence, the Ontario electricity market is regarded as one of the most volatile in [25, 26]. The correlation between the Ontario hourly energy price (HOEP) and Ontario hourly power demand (HOED) for a week is exhibited in Fig. 1. As can be seen in the figure, the prices of electricity in the deregulated Ontario electricity market depend on the electricity demand. When the demand for electricity is so high and the supply is limited, there is intense competition for electricity prices. Therefore, the price forecast of the smart grid scenario such as the Ontario power market is more complicated than usual electricity systems with an intrinsic interrelation within price and demand, and a novel prediction approach in that market should be introduced to deliver high accurate prediction.

The contributions of this paper are summarized as follows:

- 1) The most significant contribution of this study is the development of a hybrid electricity price forecasting technique relying on an effective BSA algorithm in conjunction with the SVR method, which enhances the efficiency of price forecasting as compared to existing forecasting techniques. BSA has been used to find the most appropriate weights with the least error, resulting in increased prediction accuracy. Despite of its simplicity of the structure, BSA has

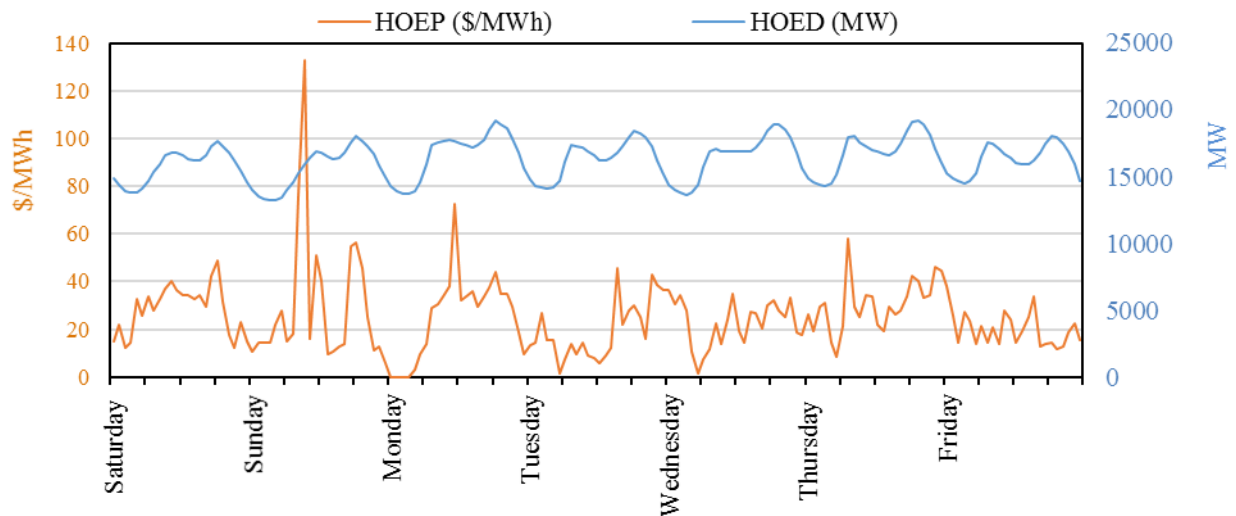


FIGURE 1. Ontario electricity market from 11.02.2020 to 17.02.2020.

been frequently employed in a variety of numerical problems due to its effectiveness in solving multidimensional functions.

- 2) In this paper, a robust hybrid feature selection technique, which integrates SVM and mutual information (MI) techniques, has been devised to handle the feature selection problem. For this case, SVM has been used to retrieve the most influential features on electricity price forecasting after MI has removed input variables in the first stage with the least amount of redundancy and the greatest amount of relevancy. Moreover, by penalizing every feature in the dual formulation of SVM, the important features during classifier construction have been determined. This method is known as Kernel-penalized SVM (KP-SVM), and it optimizes the shape of an anisotropic RBF Kernel by removing features for the classifier performance that are less significant.
- 3) The proposed procedure's relevance and precision are assessed by comparing achieved results to various machine learning approaches with proven results.

Rest of the paper are arranged as follows: Section II and III explain the short description on evolution process and structure of SVR and BSA. In section IV, short term electricity price forecasting and formulating models are presented as well as the proposed feature selection method to ensure the most suitable features by filtering the input variables employing MI in first stage and applying SVM in second stage to forecast the EP on short term basis. Section VI provides extensive discussions with results and statistical analysis is presented which ensured that the proposed method is applicable and strongly suitable for electricity price forecasting in de-regulated electricity market. Lastly, conclusion is described in section VI.

## II. SUPPORT VECTOR REGRESSION (SVR)

The data is mapped into higher dimensional feature spaces using kernel-based techniques with the expectation that perhaps the data are linearly segregated or have a better structure in the higher dimensional spaces.

The best-known member of Kernel-based techniques is SVMs, which are an extended version to the nonlinear model of the generalized portrait algorithm has the capability to either categorized data input or identifying complicated interactions in data input. SVR refers SVMs that are used for function approximation and forecasting meanwhile SVM dealing with the problems of classification is just termed for vector classification (SVC) supporter. SVC is transformed to SVR with only a few slight alterations. As a result, SVM can be developed into SVR, using statistical learning theory-based, as an effective function approximation technique [27]. Using the data set provided, consider the following:

$$D = \left\{ (x_1, y_1), \dots, (x_k, y_k), \dots, (x_m, y_m), \right\} \quad (1)$$

$m = 1, 2, 3, \dots, N$

Where,  $x_m \in R^n$  signifies the  $m^{th}$  component in an  $n$  - dimensional vector input,  $y_m \in R$  represent the recorded time step  $m$  response values with  $N$ -observation total number. The following expression is linear SVR estimation function:

$$f(x_m) = \langle W, \phi(x_m) \rangle + b \quad (2)$$

where  $b$  and  $W$  are the regression function intercept and weight vector respectively,  $\bar{y}_m$  denotes the output model approximation at time phase,  $m$ . meanwhile the vector inner product of the indicators is denoted by  $\langle W, x_m \rangle$ . SVR approach uses Kernel trick by adding the Kernel function  $W$  in order to facilitate nonlinear modelling  $\phi(x_m)$  that is a non-linear projection from the space of input to a space of high-dimensional feature [28]. Hence, the SVR estimation function has been written in the following form:

$$\bar{y}_m = f(x_m) = \langle W, \varphi(x_m) \rangle + b \quad (3)$$

The aim of the SVR approach is to find a function that diverges from  $y_m$  by no more than  $\varepsilon$  for each data point ( $m = 1, 2, \dots, N$ ), meanwhile  $W$  is as flat as it can be. As a result, for improved generalization performance and to guarantee that  $W$  is as flat as it can be, it is necessary to minimize the norm of vector  $W$ , which calculates the flatness of the function as defined by Eq. (4):

$$J(W)_{SVR} = \frac{1}{2} \|W\|^2 \quad (4)$$

under the condition that all residuals have a less than epsilon value:

$$\forall k : |y_m - \bar{y}_m| \leq \varepsilon \quad (5)$$

The SVR is oblivious to minor errors, and it penalizes if exceed the  $\varepsilon$  value. In SVR, this role is accomplished by using  $\varepsilon$ -insensitive loss function as defined in Eq. (6).

$$|y_m - \bar{y}_m|_\varepsilon = \begin{cases} 0, & \text{if } |y_m - \bar{y}_m| \leq \varepsilon \\ |y_m - \bar{y}_m| - \varepsilon, & \text{otherwise} \end{cases} \quad (6)$$

If the forecasted value is inside the  $\varepsilon$ -insensitive tube, the measured loss is equal to zero, according to the  $\varepsilon$ -insensitive loss function. When the  $\varepsilon$ -insensitive loss function is considered, the objective function recognized as the regularized risk function is obtained. Hence, in SVR, the task is to minimize the next regular risk function by approximating the value threshold ( $b$ ) and vector weight ( $W$ ):

$$\text{minimize } R(J) = \frac{J}{N} \sum_{m=1}^N (y_m, \bar{y}_m)_\varepsilon + \frac{1}{2} \|W\|^2 \quad (7)$$

where the trade-off between model flatness and the penalty levied on forecasted values that fall beyond the  $\varepsilon$ -insensitive tube is established by  $J$  as a positive constant regularization parameter. Since the parameter dictates the degree of tolerable errors (accuracy) in SVR, the regularization parameter ( $J$ ) defines the balance between efficiency and generalization capability[29]. At this point, it is likely that no SVR approximation function can find a way to fulfill the constraints for any data point ( $m = 1, 2, \dots, N$ ), so the constraint that has feasible values are not assured. In order to contend with other infeasible restrictions, the two slack variables ( $\xi_m, \xi_m^*$ ) as seen in graphical illustration of Figure 2 that reflect the distance from real values to the corresponding boundary values of  $\varepsilon$ -insensitive for each data point.

Even though this solution resembles the SVC definition of soft margin, the key distinction lies in the slack variables used. For each data point, SVR requires assigning two slack variables ( $\xi_m, \xi_m^*$ ), while in SVC one slack variable ( $\xi_m$ ) is allocated for every data point. Incorporating two

slack variables yields the primal formula that must be optimized as follows:

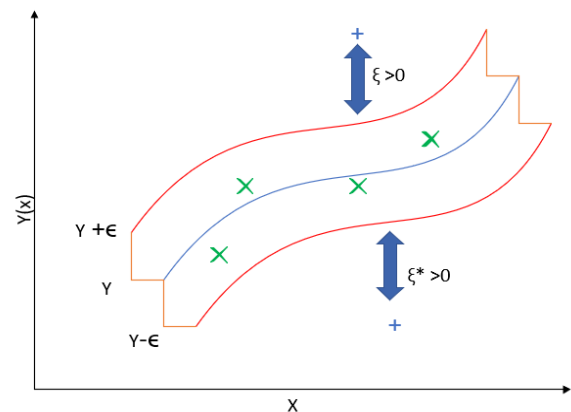


FIGURE 2. The  $\varepsilon$ -insensitive loss function in SVR

$$\text{minimize } R(W, \xi_m, \xi_m^*) = \frac{J}{N} \sum_{m=1}^N (\xi_m + \xi_m^*) + \frac{\|W\|^2}{2} \quad (8)$$

S.t.

$$\forall k : y_m - \bar{y}_m \leq \varepsilon + \xi_m \quad (9)$$

$$\forall k : \bar{y}_m - y_m \leq \varepsilon + \xi_m^* \quad (10)$$

$$\forall k : \xi_m, \xi_m^* \geq 0 \quad (11)$$

The parameter ( $\nu$ ) is substituted parameter  $\varepsilon$  is deployed to control the amount of support vectors and training errors inside the spectrum of  $[0, 1]$  to enhance SVR output

precision. The primal formula is converted into the following formulation after swapping ( $\varepsilon$ ) with ( $\nu$ ):

$$\text{minimize } R(W, \xi_m, \xi_m^*, \varepsilon) = R((\nu \cdot \varepsilon) + (\frac{1}{N} \sum_{k=1}^N (\xi_m + \xi_m^*))) + \frac{\|W\|^2}{2} \quad (12)$$

S.t.

$$\forall k : y_m - \bar{y}_m \leq \varepsilon + \xi_m \quad (13)$$

$$\forall k : \bar{y}_m - y_m \leq \varepsilon + \xi_m^* \quad (14)$$

$$\forall k : \xi_m, \xi_m^* \geq 0 \quad (15)$$

The primal Lagrangian form as developed by Eq. (16) is utilized to address this constrained optimization problem (minimizing the primal formula).

$$L(W, b, \xi_m, \xi_m^*, \alpha_m, \alpha_m^*, \beta_m, \beta_m^*, \eta) = \frac{\|W\|^2}{2} + J \cdot \nu \cdot \varepsilon + \frac{J}{N} \sum_{m=1}^N (\xi_m + \xi_m^*) - \eta \cdot \varepsilon - \sum_{m=1}^N (\beta_m \xi_m + \beta_m^* \xi_m^*) - \sum_{m=1}^N \alpha_m (\xi_m + y_m - \bar{y}_m + \varepsilon) - \sum_{m=1}^N \alpha_m^* (\xi_m^* + \bar{y}_m - y + \varepsilon) \quad (16)$$

Where Lagrangian multipliers is referred to  $\alpha_m, \alpha_m^*, \beta_m, \beta_m^*$  and ( $\eta$ ).

The primal Lagrangian type saddle point is fined which minimizes the primal variables ( $\xi_m, \xi_m^*$ ), ( $b$ ) and ( $W$ ) and, maximize over the Lagrangian multipliers ( $\alpha_m, \alpha_m^*$ ), ( $\beta_m, \beta_m^*$ ) and ( $\eta$ ). Therefore, the Lagrangian form derivations are set to zero for the primary variables, which provides the complementarity criteria for the following Karush-Kuhn-Tucker (KKT) [30]. Thus, the derivatives of Lagrangian form with respect to the primal variables are set to zero, which yields the following Karush-Kuhn-Tucker (KKT) complementarity conditions:

$$\frac{\partial L_{SVM}}{\partial W} = 0 \rightarrow W = \sum_{K=1}^N (\alpha_k^* - \alpha_k) \varphi(x_k) \quad (17)$$

$$\frac{\partial L_{SVM}}{\partial W} = 0 \rightarrow J \cdot \nu = \sum_{K=1}^N (\alpha_k^* + \alpha_k) - \eta \quad (18)$$

$$\frac{\partial L_{SVM}}{\partial W} = 0 \rightarrow \sum_{m=1}^N (\alpha_m^* - \alpha_m) = 0 \quad (19)$$

$$\frac{\partial L_{SVM}}{\partial W} = 0 \rightarrow \frac{J}{N} = \alpha_m^* - \beta_m^* = 0, \quad (20)$$

$m = 1, 2, \dots, N$

$$\text{maximize } L(\alpha) = \sum_{m=1}^N (\alpha_m - \alpha_m^*) y_m - \frac{1}{2} \sum_{m=1}^N \sum_{j=1}^N (\alpha_m - \alpha_m^*) (\alpha_j - \alpha_j^*) K_{mj} \quad (21)$$

As the problem formulation meets a constraint and convex, dual Lagrangian solution is determined for the cost of optimal solution to the primary problem, regardless of the

fact that primal and dual Lagrangian optimal values must be the same, and their disparity is regarded as gap of duality.

The development of Lagrangian function from the primal function using two non-negative Lagrange multipliers ( $\alpha_m, \alpha_m^*$ ) for each observation ( $x_m$ ). in order to obtain the dual formula. After that, by plugging Eqs. (17) – (20) into Eq (16) dual Lagrangian is calculated. As a result, the dual formula, as formulated as follows, should be maximized:

$$\sum_{m=1}^N (\alpha_m - \alpha_m^*) = 0 \quad (22)$$

$$\sum_{m=1}^N (\alpha_m + \alpha_m^*) \leq C \cdot \nu \quad (23)$$

$$\forall k : 0 \leq \alpha_m, \alpha_m^* \leq \frac{C}{N} \quad (24)$$

Where the two vectors  $x_k$  and  $x_j$  of the inner product  $\varphi(x_m)$  and  $\varphi(x_j)$  in the feature space respectively is represent by  $K_{kj} = k(x_m, x_j) = \varphi(x_m)^T \varphi(x_j)$ , defined as the Kernel function that has to fulfill the condition of Mercer. Kernel matrix has only non-negative Eigen values which defines by positive semi-definite known as the Mercer's theorem. The problem of optimization is transformed to the convex optimization problem by adding positive definite kernel such that the single solution is guaranteed. There are generally several kernels in SVM models which can be used. The linear function, polynomial function, Gaussian RBF (radial basis function), exponential RBF, and sigmoid function of the Kernel functions are most often used.

• **Linear Kernel function**

A linear kernel described by an inner product  $\langle x_m, x_j \rangle$  as shown below is the simplistic kernel function:

$$K_{mj} = k(x_m, x_j) = ax_m^T x_j + c_0 \quad (25)$$

where, optional constant value represents by  $c_0$ .

• **Polynomial Kernel function**

A kernel function that is non-stationary well-known as polynomial kernel. The polynomial kernel function, as defined in Eq. (26), is well adapted for problems involving normalized input data.

$$K_{mj} = k(x_m, x_j) = (ax_m^T x_j + c_0)^d \quad (26)$$

where, the adjustable parameters for the function Kernel are represented as  $c_0$ , the constant term expression,  $a$  is the slope, and  $d$  is the degree of the polynomial.

- **Gaussian RBF (radial basis function)**

Gaussian RBF is given in following equation:

$$K_{mj} = k(x_m, x_j) = \exp\left(-\frac{\|x_m - x_j\|^2}{2\gamma^2}\right) \quad (27)$$

Where constant parameter is defined by  $\gamma$ , instead the implementation of Gaussian RBF is as follows:

$$K_{mj} = k(x_m, x_j) = \exp(-\partial \|x_m - x_j\|^2) \quad (28)$$

where, the parameter that can be adjusted is expressed as  $\partial$ , does a substantial contribution in the Gaussian RBF efficiency. The Gaussian RBF is linear and the projection of higher-dimensional begins to lose its power of the non-linear form for the overlook of this adjustable parameter. Furthermore, if the parameter is underestimated, the Gaussian RBF regularization efficiency begins to be lost, and the determination boundary for the training data set will be extremely vulnerable to noise.

- **Exponential RBF**

As formulated by Eq. (29), the exponential kernel function is very close to the Gaussian RBF, with the square of the norm omitted.

$$K_{mj} = k(x_m, x_j) = \exp(-\partial \|x_m - x_j\|) \quad (29)$$

- **Hyperbolic Tangent Kernel function**

The function of Hyperbolic Tangent Kernel is an ANNs-derived sigmoid function. The Hyperbolic Tangent Kernel function can be written as the MLP kernel function since the bipolar sigmoid function is utilized as an activation function for artificial neurons. In general, SVM models are usually closely linked to ANNs, and an SVM model that uses a function of the MLP Kernel is analogous to a perceptron ANN with two-layer. The use of an MLP kernel function in SVM is a type of training for ANNs that allows unknown parameters to be acquired by resolving a quadratic programming problem with linear constraints instead of by solving a non-convex, uncompromising optimization problem, such as in ANNs. Furthermore, MLP Kernel function as formulated by Eq. (30) indicates satisfactory efficiency in operation even though it is only conditionally optimistic.

$$K_{mj} = k(x_m, x_j) = \tanh(ax_m^T x_j + c_0) \quad (30)$$

Where, Kernel function adjustable parameters is in form of  $c_0$  and  $a$  expression. Generally,  $(\text{data dimension})^{-1}$  is standard value for adjusting  $a$ .

The regression parameter  $W$  can no longer specifically be determined according to the non-linear mapping of Kernel functions in feature space. Eq. (21) considers as the maximization problem formulation that is in a form quadratic programming (QP) problem. The universal solution of two

non-negative Lagrange multipliers  $(\alpha_k, \alpha_k^*)$  is guaranteed by means of a quadratic problem of programming for each data set ( $k=1, 2, \dots, N$ ). The spares model is derived from the data point where the pair of non-negative Lagrange multipliers  $(\alpha_k \alpha_k^* = 0)$  is vanished while the support vector is a data point that does not disappear the pair of non-negative Lagrange multipliers  $(\alpha_k \alpha_k^* \neq 0)$ . The QP problem only desires to be determined for the support vectors in the sparse model  $(\alpha_k \alpha_k^* = 0)$ . As a result, the regression hyperplane optimum ideal weight vector is determined in the following equation:

$$W = \sum_{m=1}^N (\alpha_m - \alpha_m^*) \varphi(x_m) \quad (31)$$

And the regression hyperplane bias  $b$  is determined so that each of the support vectors fulfill the following requirement.

$$\varepsilon - y_m + \bar{y}_m = 0 \quad (32)$$

Finally, it is possible to express the regression function in the following way:

$$\bar{y}_m(x) = \sum_{m=1}^N (\alpha_m - \alpha_m^*) k(x, x_m) + b \quad (33)$$

### III. BACKTRACKING SEARCH ALGORITHM

BSA is a newly proposed evolutionary algorithm with a straightforward structure. However, because of its large potential in solving multimodal functions, that allows easy adjustments to various numerical optimization problems. With the usage of one control parameter and two advanced crossover and mutation operators, the problem search space is explored by harmonizing exploitation of improved findings. For the purpose of establishing a trial population, it employs two advanced crossover and mutation operators. These operators are unique and very distinct from the crossover and mutation processes that have been outlined by other evolutionary approaches (e.g. GA and DE). The strategies of the BSA to build trial populations and control the space boundaries and to adjust the search-direction matrix amplitude offer effective capabilities for exploration and exploitation.

In order to avoid the shortcomings of metaheuristic approaches, BSA is designed (for example, over-sensitivity and control parameters to initial values such as premature convergence and time-consuming computation) as it only has a single control parameter which is not too sensorial to the primary value of the parameters. When creating a trial population, it can also use the memory it possesses to draw on the experiences of previous generations. For the usage in producing the search-direction matrix, it specifically stores a population which is selected randomly from the earlier generations in its memory. In comparison to other well-known evolutionary approaches, statistical analysis in [31] shows that BSA is a viable optimization strategy for addressing high multimodal optimization benchmarks. Various stages for BSA algorithm are presented in Fig. 3.

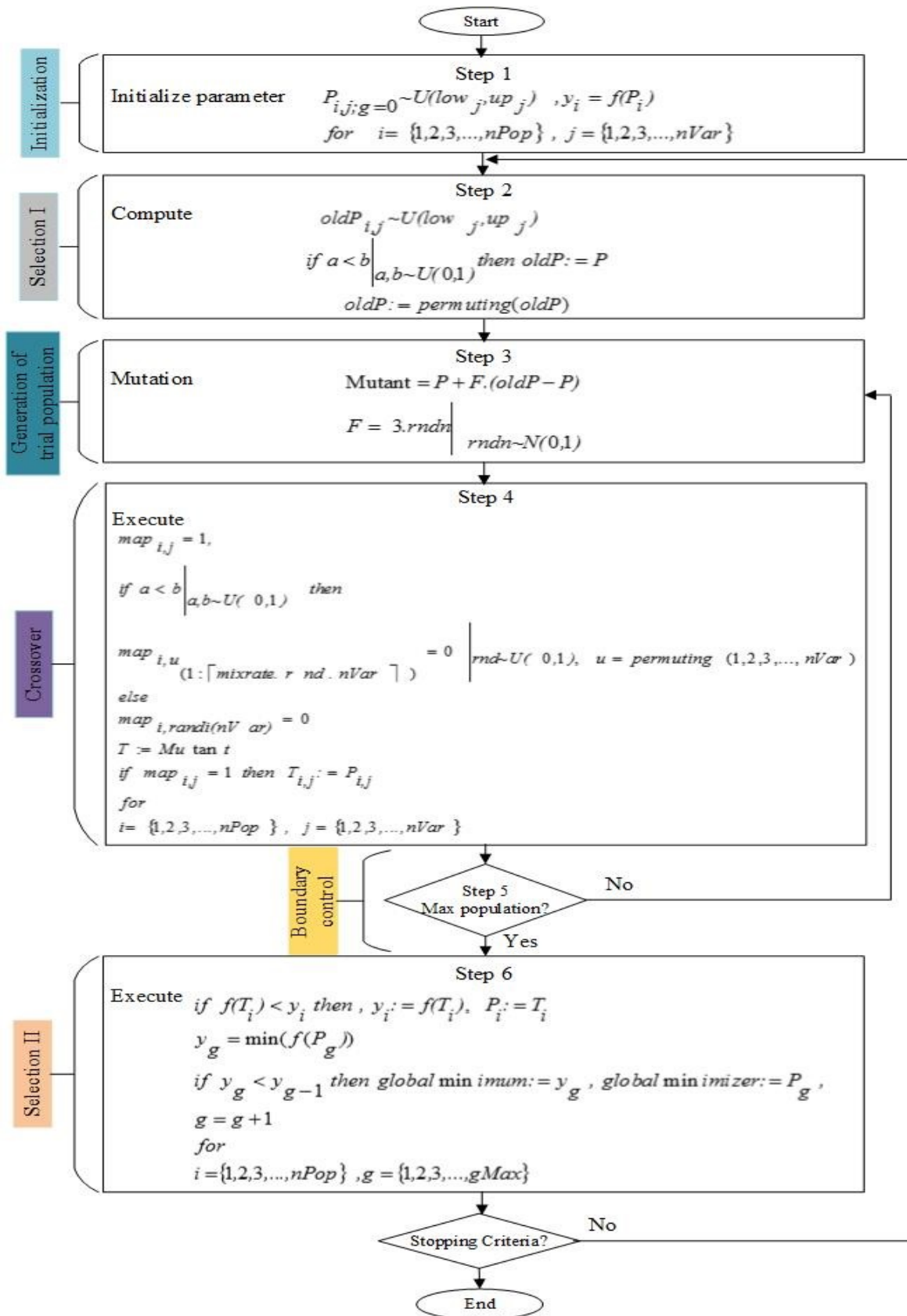


FIGURE 3. BSA optimization procedure.



#### IV. DESIGN OF INPUT VECTOR FOR EFF

The relationship between energy price and demand in previous hours exemplifies in time series characteristics and spreads over hourly intervals in the competitive energy market. In addition, electricity price is a function of the demanded electrical energy. The prediction of the prices of electrical energy in time  $t$  relies, not only on the demand of electricity in time  $t$ , but also the values of electricity demand in past and even past values which are set forth as follows:

$$HOEP(t) = F \left( \begin{array}{l} HOEP(t-1), HOEP(t-2), HOEP(t-3), \dots, HOEP(t-NL_{HOEP}), \\ HOED(t), HOED(t-1), HOED(t-2), HOED(t-3), \dots, HOED(t-NL_{HOED}) \end{array} \right) \quad (34)$$

Where  $HOEP(t)$  and  $HOED(t)$  express the hourly Ontario electricity price and hourly Ontario electricity demand at time  $t$  whilst they are estimated to be a time series with  $t$  interval,  $NL_{HOEP}$  displays the number of lag orders for the electricity price and equivalently  $NL_{HOED}$  indicates the demand of electricity in lagging order.

The goal of this study is the employment of SVR-BSA technique to forecast hourly Ontario electricity price (HOEP) of the Ontario market. The required input data HOED and HOEP for the year of 2020 are obtained from [25]. Equation (35) is applied for the normalization of dependent and independent variable since the past data has a variable range. Normalizing a particular data requires converting the collected data from different scales to a single estimated scale, which is generally done before data processing.

$$\bar{Z}(t) = \frac{Z(t) - \min(Z)}{\max(Z) - \min(Z)} + 1 \quad (35)$$

Where  $\bar{Z}$  is expressed the normalized data,  $Z$  represents the data that to be normalized and  $t$  represented the hourly interval.

To propose the electricity price forecasting model, in this research exogenous variables ( $NL_{HOEP} = NL_{HOED} = 168$ ) of one week with one hour lagged values are taken where the total number of available exogenous variables are 336. The machine learning algorithms with too many features slow down the learning process that resulting in poor performance and over fitting the training data. Consequently, only the features that have significant effect on the output, (in electricity price forecasting procedure) should be provided to the algorithm.

For the process of machine learning and statistics algorithm, feature (predictors or variable) selection can be termed as attribute selection where variable subset or variable selection is a procedure of selection of a subset of related features in model development. The aim of applying feature selection approaches is three-fold [32].

1. To improve the predictors prediction performance by decreasing over fitting that is formally reduction of variance.
2. To provide a fast and more cost-effective method to build the model.

3. Presenting a simplified model which is easier to interpret.

The mutual information (MI) approach has been widely used in [33] for forecasting electricity market prices. The lagged values of the candidate input, which includes load demand, price, and other factors given by the electrical market, arise a difficulty for this approach. Hence, it is difficult to achieve the probability distribution for both individual and joint candidate.

It should also be remembered that the electricity price is a signal that changes over time. As a result, the candidate input's long history is irrelevant to utilize since market circumstances change all the time. Thus, it can provide inaccurate or mislead price forecasting due to insufficiency of the information [34].

The basic objective of mutual information is to find out interrelation within two random variable  $X$  and  $Y$  where, one variable is random in nature and the other variable information is gathered by using the previous variable information [35]. The mutual information will be zero if there is no information of one variable  $X$  with respect to  $Y$  and vice versa. As a result, both random variables are independent in nature. Mutual information with a high value can be attained if variable  $Y$  is a deterministic function of variable  $X$  as same as variable  $X$  is a deterministic function of variable  $Y$ . the relationship within CE and MI is showed in Figure 4. In case of large MI, it is found that the variable  $X$  and  $Y$  are tightly dependent and linked to each other.

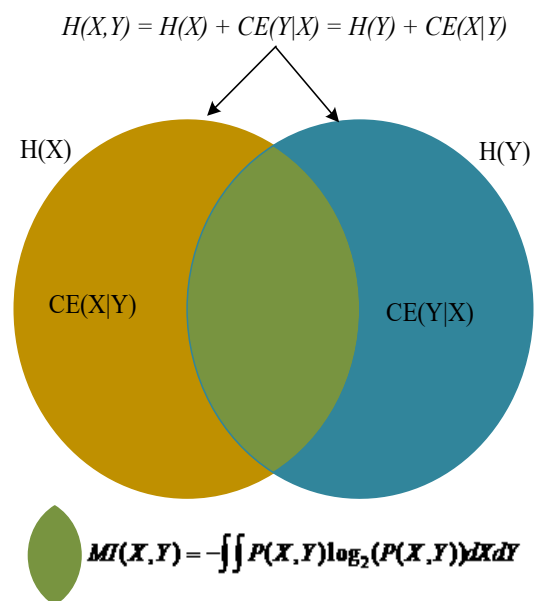


FIGURE 4. Graphical representation between mutual information and conditional entropy.

Conditional entropy is also observed in addition to entropy. Conditional entropy is the measurement of the first random variable's average uncertainty after the second random variable. To attain the mutual information among X, Y and MI(X,Y) a joint probability distribution function PXY(X,Y) is used since the entropy of the random variable is complexly linked to the theme of mutual information.

Let  $X = \{HOEP(t-1), HOEP(t-2), HOEP(t-3), \dots, HOEP(t-NL_{HOEP}), HOED(t), HOED(t-1), HOED(t-2), HOED(t-3), \dots, HOED(t-NL_{HOED})\}$  as a vector of all input features and  $Y = HOEP(t)$  as the target or output feature. In this study the demand of electricity and price are expressed by HOED(t) and HOEP(t) respectively. The computational equation of the mutual information (MI) between each individual input and output feature is shown in Fig. 3. For instance, the mutual information between the first lagged value of electricity demand and electricity price is calculated by MI {HOEP (t-1); HOED (t)}. After that, depending on the calculated value of MI, the input with all features is stored in descending order. A high MI value indicates that each input and output variable is highly dependent on one another. On the other hand, the value of MI lower than the threshold TH is deleted since it shows less significant on the output where a subset  $SX \subset X$  is formed by the residual input features.

In the first stage of filtering out irrelevant inputs is by computing the mutual information between every individual variable of input and feature of output using proposed feature selection technique according to [36]. The higher mutual information value implies a higher input and output dependency. Based on the computed mutual information, the input features are sorted in descending order. Input features that have a lower value of mutual information than the relevancy threshold TH and have a less significant influence on the output are eliminated. The relevancy criterion of TH = 0.46 is used to filter out redundant features. A total of 68 features have been determined to be the most relevant attributes after the filtering process was completed.

Feature selection approach is the most suitable approach for identifying variable of input. Through feature selection, the accomplishments of the final predictor variables (HOED and HOEP in preceding hours) in the best SVM models were assessed. It is worth noting that these factors were discovered by developing and manipulating multiple models with various input variables. While the model is being built, the embedded method is figured out which is the subset of features and has contributed the most to its accuracy. The feature selection and training processes in embedded methods are indistinguishable because feature selection and model construction is carried out simultaneously. The major downside of these methods is that the chosen features are responsive to the layout of the underlying model so that embedded methods are typically unique to their learning algorithms, even though the embedded methods are less computationally intensive than wrapper methods. Classification trees, random woodlands and regularization

approaches are multiple type of embedded system. The most common types of embedded feature selection methods are regularization approaches often known as penalization approaches. The penalization approaches apply more constraint to the model building procedure by penalizing the model for higher complexity caused the model bias toward the simplicity.

An embedded technique is developed for feature selection using SVMs. The reason for using this strategy is that through optimizing the Kerenl function, the classification performance can be improved by avoiding the features which influence on classifier's generalization. The main concept is to utilize a gradient descent approximation for Kernel optimization and feature deletion to penalize the usage of features in the dual formulation of SVMs [37]. The proposed approach combines the parameters of generalization (using the 2-norm), goodness of fit, and feature selection (using a 0-norm estimation) to identify the most appropriate RBF-type kernel function for each problem with a minimum dimension.

In the proposed technique, the anisotropic Gaussian Kernel function is utilized as follows:

$$K(A_i, A_s) = \exp - \sum_{j=1}^n \frac{(A_{ij} - A_{sj})^2}{2\sigma_j^2} \quad (36)$$

Where,  $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_n]$  is the kernel shape for n number of variables. Different widths in various dimensions are considered to find out the importance of feature  $\sigma_j$ . For a large value of  $\sigma_j$ , the importance of j is less while small value of  $\sigma_j$  increases the importance of j.

To convert the feature selection process into a minimization of problem  $\sigma = [1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_n]$  the changing of variables as follows:

$$K(A_i, A_s, \nu) = \exp\left(\frac{\|\nu * A_i - \nu * A_s\|^2}{2}\right) \quad (37)$$

Where, \* indicates the component wise vector product operator.

The proposed method incorporates the feature selection in the dual formulation of SVMs. In the formulation process, a penalization function  $f(\nu)$  based on "0" norm estimation and modification of the Gaussian Kernel applying anisotropic width vector as a decision variable. The penalization of the feature should be negative as the dual SVM is a maximization problem. The formulation of SVMs for feature selection is initially proposed as follows:

$$\max \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,s=1}^m \alpha_i \alpha_s Y_i Y_s K(A_i, A_s, \nu) - C_2 f(\nu)$$

Subject to,

$$\sum_{i=1}^m \alpha_i Y_i = 0 \quad (38)$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, m,$$

$$V_j \geq 0, j = 1, \dots, n$$

Since the equation 38 is a non-convex, an iterative algorithm is developed as an estimation of this formulation. Thus, a two-step methodology is proposed where traditional formulation of SVM for a fixed kernel width  $\nu$  is determined in the first step.

$$\max \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,s=1}^m \alpha_i \alpha_s Y_i Y_s K(A_i, A_s, \nu) - C_2 f(\nu)$$

Subject to,

$$\sum_{i=1}^m \alpha_i Y_i = 0 \quad (39)$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, m,$$

In the second step, for a certain value of  $\alpha$ , the algorithm is solved according to the following equation:

$$\min_{\alpha} F(\nu) = \alpha_i \alpha_s Y_i Y_s K(A_i, A_s, \nu) - C_2 f(\nu)$$

Subject to,

$$V_j \geq 0, j = 1, \dots, n \quad (40)$$

Among the 68 selected candidates, 30 features with most influential effect on electricity price are selected by developing KP- SVM algorithms as inputs for the next forecasting process. The optimal subsets of input variables selected by KP- SVM approach are presented as follows:

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

## V. NUMERICAL RESULTS

The electricity price forecasting process and development were coded in MATLAB (R2021b) and run on a personal computer, with a corei7 (4 generation processor of 2.70 GHz clock speed) and 8GB RAM. In the next process, SVR-BSA has been employed into this study to enhance the precision of EPF in Ontario mainland which considered as one of the most unstable and volatile market for electricity. The most effective input variables are selected by proposed feature selection method by developing MI-SVM as an input for electricity price prediction. In addition, SVR-BSA is compared to different machine learning algorithms for the short-term of EPF accuracy purposes. The process for the

proposed short-term electricity price forecasting (EPF) methodology is displayed in Figure. 5. Sequential steps to acquire the suitable models for short-term EP forecasting are implementing as follows:

**Step 1:** To establish the relationship between the electricity price and demands, an hourly interval-based time series can be demonstrated that may be seen in a dynamic electricity market. Literally, electricity price is a function of the electricity demand that relies on the present price of electricity and demand value for hour t and hour t-1 as specified at Eq. (34). The specific data from the Ontario electricity market in 2020 (HOEP and HOED) obtained using the suggested feature selection approach are treated as variables that are independent of any other variable and the HOEP of dependent variable is calculated. Due to seasonal impacts, the effectiveness of the employed approaches has been evaluated for EPF in various seasons by including one month from each season (i.e., winter (February), spring (May), summer (August), and autumn (November)). Variables are divided into two categories: independent and dependent variables types. They are two subdivisions, with the first three weeks' hourly data used for design phase training and the last week of each month's hourly data used for testing phase of models created from the analysis.

**Step 2:** Development of training step includes deriving algorithms that bind input variables with the output variables and in order to expedite the learning process, the formulation in Eq. (35) is used to normalize the input and output variables. A data set beyond the training data set known as the test data set is used to determine the predicted price. By using the values in the testing phase, the reliability of the produced model can be established.

**Step 3:** For a precise prediction of the electricity price, SVM-BSA is deployed by minimizing the cost function described as follows:

$$F = \sum_{t=1}^N |HOEP(t)_{Actual\ value} - HOEP(t)_{Estimated\ value}| \quad (41)$$

**Step 4:** Various assessment parameters, such as root mean square error (RMSE), mean absolute percentage error (MAPE), and Thiel's inequality coefficient (U - statistic), are deployed to indicate the efficacy of prediction models:

$$MAPE\% = \frac{1}{N} \sum_{t=1}^N \frac{|HOEP(t)_{Actual\ value} - HOEP(t)_{Estimated\ value}|}{HOEP(t)_{Actual\ value}} \times 100 \quad (42)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (HOEP(t)_{Actual\ value} - HOEP(t)_{Estimated\ value})^2} \quad (43)$$

$$U = \frac{RMSE}{\sqrt{\frac{1}{N} \sum_{t=1}^N (HOEP(t)_{Actual\ value})^2 + \frac{1}{N} \sum_{t=1}^N (HOEP(t)_{Estimated\ value})^2}} \quad (44)$$

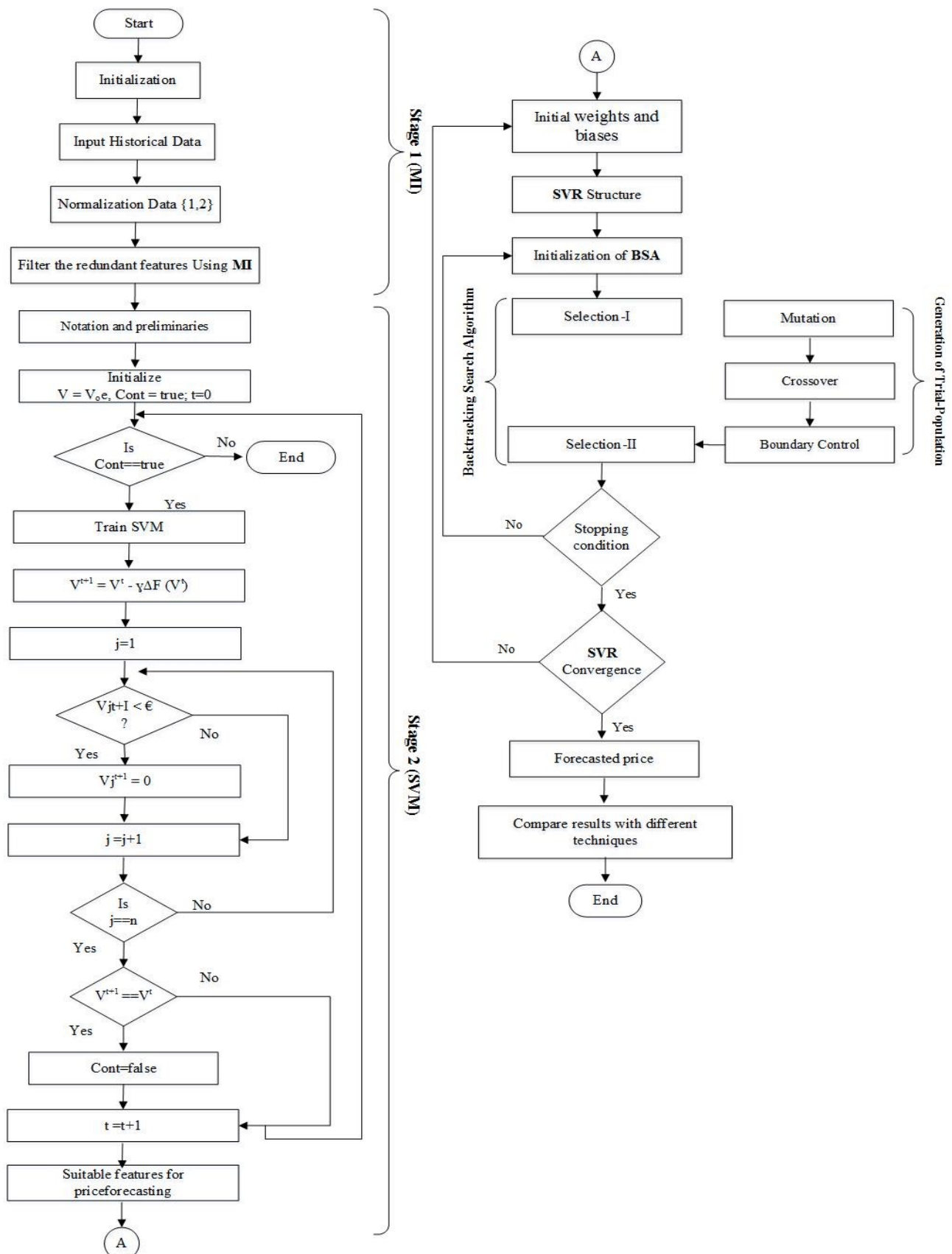


FIGURE 5. The procedure of developing SVR-BSA based method for short-term electricity price forecasting (EPF).

Binary results [0,1] frequently generated by **U-statistic** with value of zero is represented at more precise prediction, meanwhile value of one is depicted an approximation that is as inaccurate as a naive guess.

Because determining dependent parameters of AI-based methods is complex and there is no consensus on their optimal values, effective procedures in the literature of electricity price forecasting are used to characterize the control parameters of the simulated systems in this study. The parameter setting of the proposed techniques are summarized in Table 1.

TABLE I  
UNITS FOR MAGNETIC PROPERTIES

Applied methods	Parameters setting
SVM (SVR)	$\partial$ :1/6
	RBF kernel C:1
	$v$ :0.5
BSA	Number of individuals = 100 Control parameter (P)= 100% Swarm population (Np)=100
PSO	$w_{max} = 0.9, w_{min} = 0.4$ $c_1 = c_2 = 2$
CSA	Number of nests (Np) = 100
	Probability of an alien egg (Pa) = [0, 1] Distribution factor ( $\beta$ ) = 1.5

Definition of the appropriateness of the data series are which have been assured by the whiteness test obtained through the models, well-known as the Durbin-Watson test [38], after obtaining a confirmatory analysis. The goal of confirmatory analysis is to affirm the whiteness of the approximate residues ( $e(t)$ ), that indicates the non-correlation among them. To generate this calculation defined by the following equation, the Residual Autocorrelation Function (RACF) is

applied. RACF results vary from 0 to 1. The RACF goes beyond the confidence range if it varies considerably from zero, indicating a non-correlation (whiteness) of the residuals and suggesting that a significant independent variable is already omitted from the model under investigation.

$$RACF = \frac{\sum_{t=2}^N (e(t)e(t-1))}{\sum_{t=1}^N (e(t))^2} \quad (45)$$

As shown in Tables 2 and 3, the SVR-BSA performance for the EPF of the Ontario region in various months for the year 2020 are compared with multiple machine learning methods for the intention of further evaluation of solution methodology. The RACF values produced demonstrate the fact that the expected residuals of all models are uncorrelated and generated models are adequately designated the provided set of data. The forecasting accuracy of the employed approaches is ranked as SVR-BSA > SVR-PSO > SVR-CSA > SVR depicted at Tables 2 and 3. This ranking is based on multi-criteria decisions using the mean rank of the methods for each indicator (absolute error, RMSE, U-statistic and MAPE) in whole set. According to the findings from the comparison of studied methods for electricity market in Ontario region for different months in different seasons, SVR-BSA approach performs much better than other methods in terms of electricity forecasting since it has higher precision in terms of various indices such as U-statistic, MAPE (%), RMSE and absolute error. Figures 6,7,8, 9 depicts the SVR-BSA performance for forecasting the electricity price in February, May, August and November during the training and testing phase for Ontario region in 2020 year.

TABLE II  
COMPARISON BETWEEN FORECASTING ACCURACY OF STUDIED METHODS FOR EPF OF ONTARIO IN FEBRUARY AND MAY IN 2020.

Performance Indexes	February				May				
	SVM	SVM-CSA	SVM-PSO	SVM-BSA	SVM	SVM-CSA	SVM-PSO	SVM-BSA	
RACF	Training	0.385	0.784	0.967	0.953	0.021	0.036	0.058	0.961
	Testing	0.928	0.927	0.957	0.943	0.966	0.982	0.981	0.982
Absolute error	Training	0.971	0.913	0.897	<b>0.884</b>	1.644	1.456	1.277	<b>0.819</b>
	Testing	0.839	0.846	0.883	<b>0.855</b>	0.838	0.855	0.851	<b>0.841</b>
RMSE	Training	1.573	1.029	0.911	<b>0.905</b>	10.13	6.808	5.207	<b>0.833</b>
	Testing	0.872	0.988	0.899	<b>0.884</b>	0.848	0.856	0.856	<b>0.832</b>
U-statistic	Training	0.042	0.028	0.024	<b>0.024</b>	0.216	0.139	0.104	<b>0.016</b>
	Testing	0.024	0.024	0.025	<b>0.024</b>	0.025	0.025	0.025	<b>0.022</b>
MAPE (%)	Training	6.563	6.353	6.343	<b>6.252</b>	7.763	7.681	7.633	<b>7.077</b>
	Testing	5.789	5.784	5.767	<b>5.716</b>	6.322	6.279	6.269	<b>6.175</b>

TABLE III  
COMPARISON BETWEEN FORECASTING ACCURACY OF STUDIED METHODS FOR EPF OF ONTARIO IN AUGUST AND NOVEMBER 2020.

Performance Indexes		August				November			
		SVM	SVM-CSA	SVM-PSO	SVM-BSA	SVM	SVM-CSA	SVM-PSO	SVM-BSA
RACF	Training	0.232	0.475	0.487	0.943	0.003	0.004	0.022	0.058
	Testing	0.925	0.965	0.959	0.963	0.726	0.917	0.962	0.965
Absolute error	Training	0.934	0.752	0.749	<b>0.663</b>	4.793	3.963	1.911	<b>1.281</b>
	Testing	0.683	0.671	0.668	<b>0.672</b>	1.141	0.957	0.953	<b>0.921</b>
RMSE	Training	2	1.108	1.089	<b>0.069</b>	49.834	42.728	11.221	<b>5.185</b>
	Testing	0.715	0.683	0.682	<b>0.671</b>	1.306	0.983	0.971	<b>0.953</b>
U-statistic	Training	0.041	0.023	0.022	<b>0.014</b>	0.636	0.521	0.103	<b>0.044</b>
	Testing	0.019	0.018	0.018	<b>0.018</b>	0.051	0.039	0.038	<b>0.037</b>
MAPE (%)	Training	4.816	4.196	4.185	<b>3.829</b>	6.372	5.373	5.058	<b>4.889</b>
	Testing	4.349	4.241	4.228	<b>4.219</b>	4.822	4.236	4.225	<b>3.439</b>

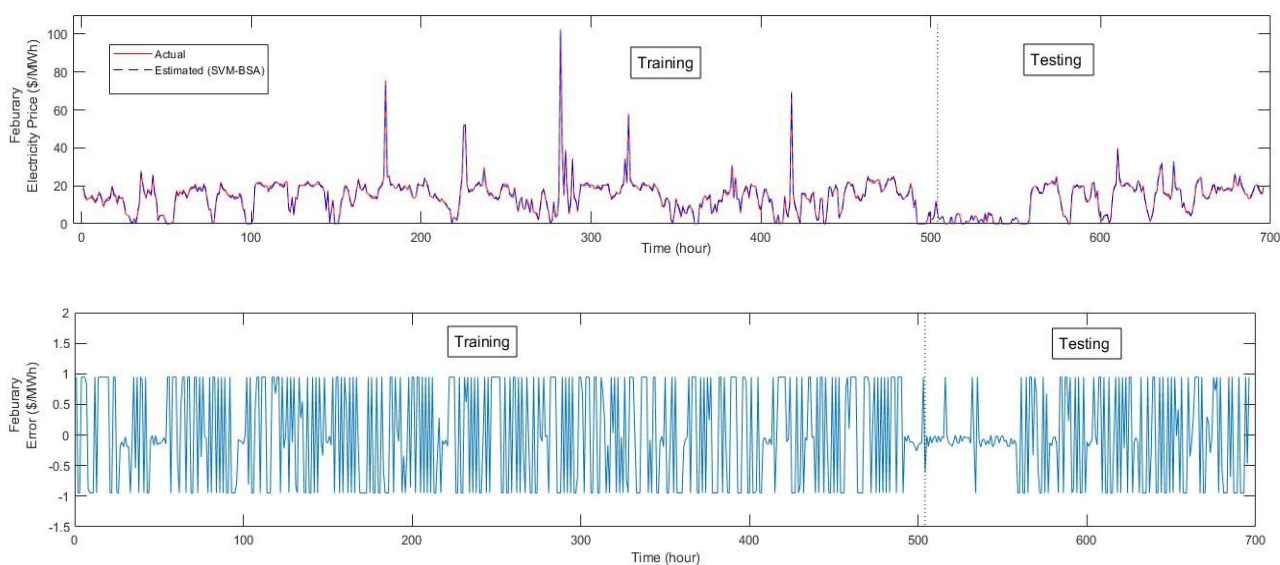


FIGURE 6. The performance of SVM(SVR)-BSA during training of design phase and testing phase and its corresponding error for EPF of Ontario in February 2020.

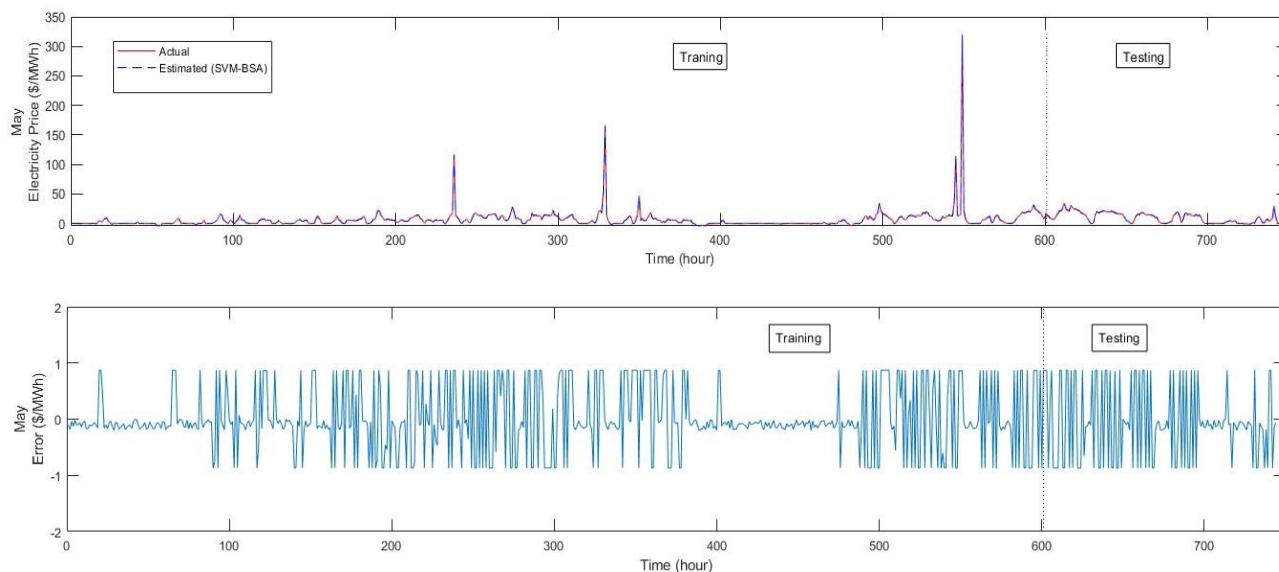
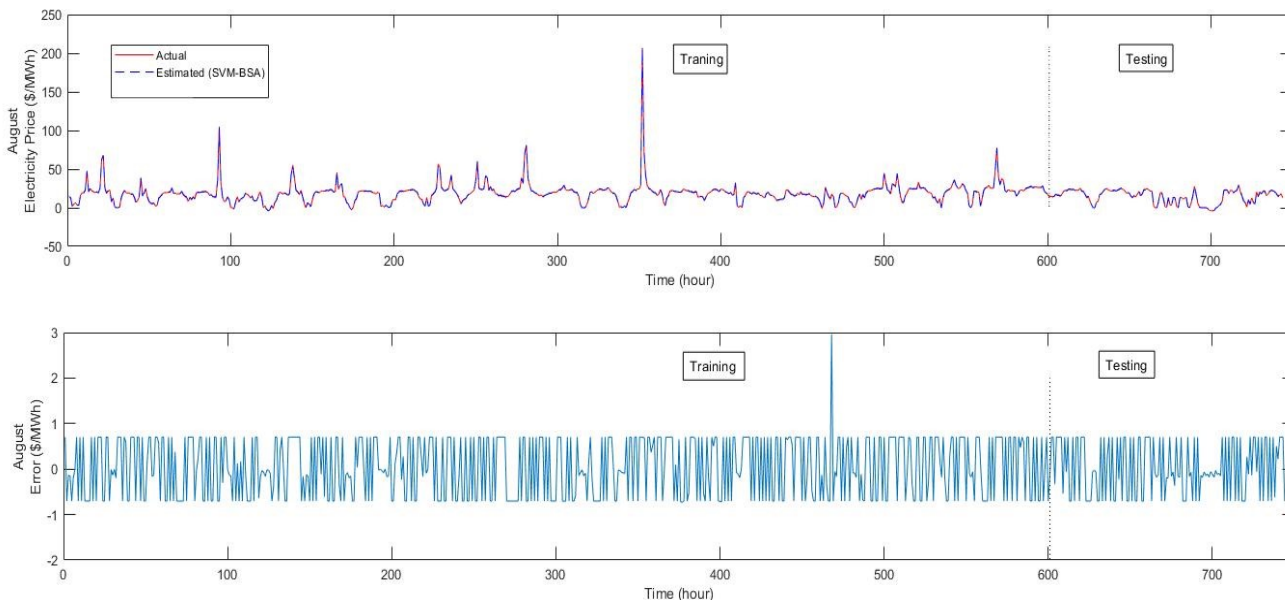
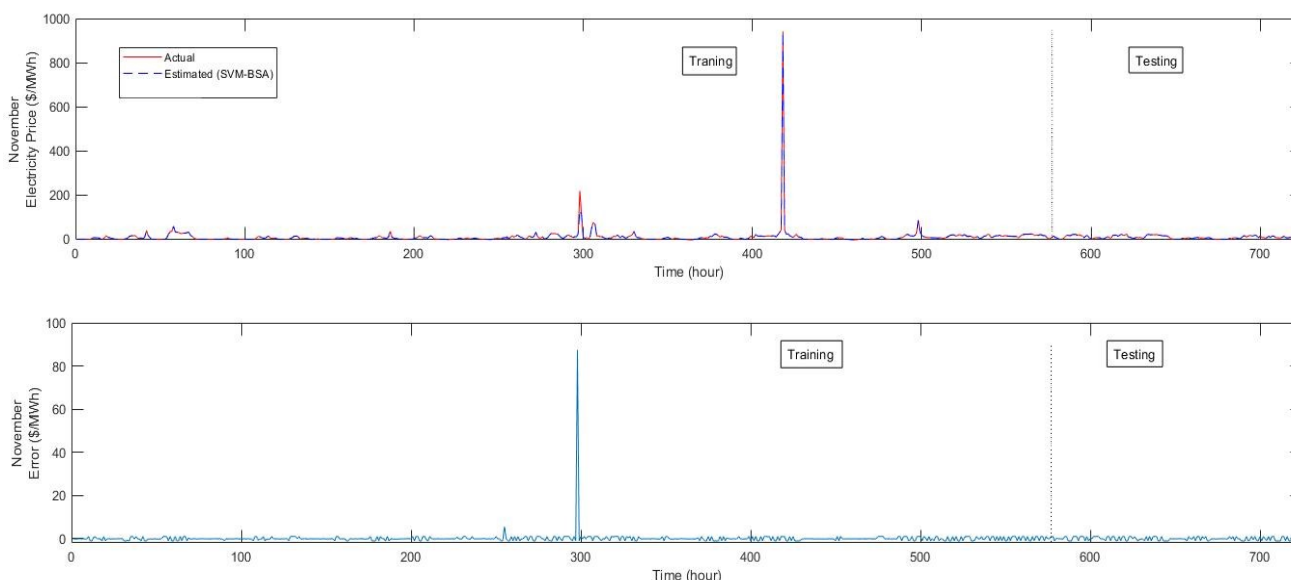


FIGURE 7. The performance of SVM(SVR)-BSA during training of design phase and testing phase and its corresponding error for EPF of Ontario in May 2020.



**FIGURE 8.** The performance of SVM(SVR)-BSA during training of design phase and testing phase and its corresponding error for EPF of Ontario in August 2020.



**FIGURE 9.** The performance of SVM(SVR)-BSA during training of design phase and testing phase and its corresponding error for EPF of Ontario in November 2020.

Computational time of the applied method with different methods has been listed in table IV

TABLE IV

COMPUTATIONAL TIME OF DIFFERENT METHODS

months	Computational time of applied methods(s)		
	SVM-CSA	SVM-PSO	SVM-BSA
February	6105.5	5106.4	<b>300.2451</b>
May	4001.2	4941.3	<b>72.5930</b>
August	3589.4	1147.0	<b>87.7186</b>
November	2698.7	1945.3	<b>58.9205</b>

Validity of mathematical models developed by SVM (SVR)-BSA is verified through application of different statistical methods as external validation. Following attributes are recommended to assess the performance of the developed model [39]:

- i- For a model that generates  $|R| > 0.8$ , a strong correlation emerges between the predicted and observed values.
- ii- For a model that generates  $0.2 < |R| < 0.8$ , a correlation emerges between the predicted and observed values.

iii- For a model that generates  $|R| < 0.2$ , a weak correlation emerges between the predicted and observed values.

Various months of Ontario mainland are being considered for computing the statistical factors of the forecasting model.

Table V shows that the developed model meets all of the statistical requirements. For de-regulated electricity market, the findings specify that the created model is a good potential and enthusiastic approach to be executed for forecasting the future electricity price.

TABLE V  
STATISTICAL FACTORS OF THE SVM(SVR)-BSA MODEL FOR EPF OF ONTARIO IN FEBRUARY, MAY, AUGUST AND NOVEMBER 2020.

Item	Formula	Condition	SVR-BSA-based Method			
			February	May	August	November
1	$R$	$0.8 < R_0$	0.9988	0.9936	0.9854	0.9997
2	$K = \frac{\sum_{i=1}^n (h_i \times t_i)}{\sum_{i=1}^n h_i^2}$	$0.85 < k < 1.15$	0.9977	0.9986	0.9964	0.9972
3	$K' = \frac{\sum_{i=1}^n (h_i \times t_i)}{\sum_{i=1}^n t_i^2}$	$0.85 < k' < 1.15$	1.0032	1.0012	1.0022	1.0007
4	$m = \frac{R^2 - R_o^2}{R^2}$	$ m  < 0.1$	-0.0032	-0.0035	-0.0041	-0.0023
5	$n = \frac{R^2 - R_o^2}{R^2}$	$ n  < 0.1$	-0.0032	-0.0037	-0.0042	-0.0023
6	$R_m = R^2 \times \left(1 - \sqrt{ R^2 - R_o^2 }\right)$	$0.5 < R_m$	0.9973	0.9922	0.9866	0.9981
Where	$R_o^2 = 1 - \frac{\sum_{i=1}^n (t_i - h_i^o)^2}{\sum_{i=1}^n (t_i - \bar{t}_i)^2}, h_i^o = k \times t_i$	$0.8 < R_o^2 < 1$	0.9997	1.0000	1.0000	0.9995
	$R_o'^2 = 1 - \frac{\sum_{i=1}^n (h_i - t_i^o)^2}{\sum_{i=1}^n (h_i - \bar{h}_i)^2}, t_i^o = K' \times h_i$	$0.8 < R_o'^2 < 1$	0.9997	1.0000	1.0000	0.9995

## VII. CONCLUSION

In this research, SVR-BSA model for day ahead electricity price prediction in the Ontario electricity market based on historical data i.e. HOEP (t) and HOED (t) are satisfactorily developed.

- A key element of machine learning algorithm is feature selection. Proper feature selection, which includes selecting significant data from a large pool of incoherent and redundant information, can result in high prediction accuracy. A hybrid feature selection technique, with the combination of SVM and mutual information techniques is proposed in this work that selects optimal subset of features among a pool of 68 features and set as input for the prediction. The proposed method shows the robustness with the proficient suitable features selection and removing the redundant and irrelevant data.

- Results validate the robustness of developed model in Ontario electricity market, as it ensures a higher forecasting legibility and simplicity for electricity price forecasting in term of, MAPE = 5.72%, 6.17%, 4.22% and 3.44% in February, May, August, and November respectively with respect to other artificial intelligence (AI) models.
- It is attained that the generated models by SVR are sufficiently able to predict the short-term EP with low complexity using a simple computer. For the forthcoming sustainable and smart grid technologies, the proposed technique has a great advantage in electricity price forecasting in any competitive and deregulated power market. Hence, the proposed technique can be termed as an intelligent technique in energy market for the electricity suppliers in addition to the 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," *Applied Energy*, vol. 172, pp. 132-151, 2016.
- [2] 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," *Applied Energy*, vol. 190, pp. 291-305, 2017.
- [3] L. Sun, L. Li, B. Liu, and S. Saeedi, "A Novel Day-ahead Electricity Price Forecasting Using multi-modal combined Integration via Stacked Pruning Sparse Denoising Auto Encoder," *Energy Reports*, vol. 7, pp. 2201-2213, 2021.
- [4] Z. Deng, C. Liu, and Z. Zhu, "Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network," *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106499, 2021.
- [5] L. Hernandez *et al.*, "A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1460-1495, 2014.
- [6] A. Pourdaryaei *et al.*, "Recent Development in Electricity Price Forecasting Based on Computational Intelligence Techniques in Deregulated Power Market," *Energies*, vol. 14, no. 19, p. 6104, 2021.
- [7] F. Keynia, "A new feature selection algorithm and composite neural network for electricity price forecasting," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 8, pp. 1687-1697, 2012.
- [8] H. Shayeghi and A. Ghasemi, "Day-ahead electricity prices forecasting by a modified CGSA technique and hybrid WT in LSSVM based scheme," *Energy Conversion and Management*, vol. 74, pp. 482-491, 2013.
- [9] H. Shayeghi, A. Ghasemi, M. Moradzadeh, and M. Nooshyar, "Simultaneous day-ahead forecasting of electricity price and load in smart grids," *Energy conversion and management*, vol. 95, pp. 371-384, 2015.
- [10] N. Li-xiao, W. Zheng-fang, Z. Chuan-zhi, S. Wen-li, and Z. Sheng-shan, "Hybrid model based on wavelet transform and ARIMA for short-term electricity price forecasting," *Application Research of Computers/Jisuanji Yingyong Yanjiu*, vol. 31, no. 3, p. 688, 2014.
- [11] C. Hamzaçebi, "Improving artificial neural networks' performance in seasonal time series forecasting," *Information Sciences*, vol. 178, no. 23, pp. 4550-4559, 2008.
- [12] W.-M. Lin, H.-J. Gow, and M.-T. Tsai, "Electricity price forecasting using enhanced probability neural network," *Energy Conversion and Management*, vol. 51, no. 12, pp. 2707-2714, 2010.
- [13] 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*: Springer, 2006, pp. 880-888.
- [14] 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 *2008 Second International Symposium on Intelligent Information Technology Application*, 2008, vol. 1, pp. 85-89: IEEE.
- [15] H. Ebrahimiyan, S. Barmayoon, M. Mohammadi, and N. Ghadimi, "The price prediction for the energy market based on a new method," *Economic research-Ekonomiska istraživanja*, vol. 31, no. 1, pp. 313-337, 2018.
- [16] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong, "A hybrid approach for probabilistic forecasting of electricity price," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 463-470, 2013.
- [17] A. Baziar and A. Kavousi-Fard, "Short term load forecasting using a hybrid model based on support vector regression," *International journal of scientific & technology research*, vol. 4, no. 5, pp. 189-195, 2015.
- [18] W.-Y. Chang, "Short-term load forecasting using radial basis function neural network," *Journal of Computer and Communications*, vol. 3, no. 11, p. 40, 2015.
- [19] N. Amjady and A. Daraeepour, "Mixed price and load forecasting of electricity markets by a new iterative prediction method," *Electric power systems research*, vol. 79, no. 9, pp. 1329-1336, 2009.
- [20] L. Wu and M. Shahidehpour, "A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid," *IET Generation, Transmission & Distribution*, vol. 8, no. 12, pp. 1937-1950, 2014.
- [21] G. Memarzadeh and F. Keynia, "Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm," *Electric Power Systems Research*, vol. 192, p. 106995, 2021.
- [22] W. Gao, A. Darvishan, M. Toghiani, M. Mohammadi, O. Abedinia, and N. Ghadimi, "Different states of multi-block based forecast engine for price and load prediction," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 423-435, 2019.
- [23] O. Abedinia and N. Amjady, "Short-term load forecast of electrical power system by radial basis function neural network and new stochastic search algorithm," *International transactions on electrical energy systems*, vol. 26, no. 7, pp. 1511-1525, 2016.
- [24] O. Abedinia, N. Amjady, M. Shafie-Khah, and J. Catalão, "Electricity price forecast using combinatorial neural network trained by a new stochastic search method," *Energy Conversion and Management*, vol. 105, pp. 642-654, 2015.
- [25] T. I. E. S. O. (IESO), "<http://www.ieso.ca>," 2020-11-17.
- [26] H. Zareipour, K. Bhattacharya, and C. A. Cañizares, "Electricity market price volatility: The case of Ontario," *Energy Policy*, vol. 35, no. 9, pp. 4739-4748, 2007.
- [27] X.-Q. Bian, B. Han, Z.-M. Du, J.-N. Jaubert, and M.-J. Li, "Integrating support vector regression with genetic algorithm for CO<sub>2</sub>-oil minimum miscibility pressure (MMP) in pure and impure CO<sub>2</sub> streams," *Fuel*, vol. 182, pp. 550-557, 10/15/ 2016.
- [28] H. Wang and D. Hu, "Comparison of SVM and LS-SVM for regression," in *2005 International Conference on Neural Networks and Brain*, 2005, vol. 1, pp. 279-283: IEEE.
- [29] W.-C. Hong, "Application of chaotic ant swarm optimization in electric load forecasting," *Energy Policy*, vol. 38, no. 10, pp. 5830-5839, 10/ 2010.
- [30] A. K. Seyed Hamidreza, "Long-term electrical energy consumption: Formulating and forecasting via optimized gene expression programming/Seyed Hamidreza Aghay Kaboli," University of Malaya, 2018.
- [31] P. Civioglu, "Backtracking Search Optimization Algorithm for numerical optimization problems," *Applied Mathematics and Computation*, vol. 219, no. 15, pp. 8121-8144, 4/1/ 2013.
- [32] E. T. Renani, M. F. M. Elias, and N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," *Energy Conversion and Management*, vol. 118, pp. 193-203, 2016.
- [33] N. Amjady and M. Hemmati, "Energy price forecasting-problems and proposals for such predictions," *IEEE Power and Energy Magazine*, vol. 4, no. 2, pp. 20-29, 2006.
- [34] N. Amjady and A. Daraeepour, "Design of input vector for day-ahead price forecasting of electricity markets," *Expert Systems with Applications*, vol. 36, no. 10, pp. 12281-12294, 2009.
- [35] T. M. Cover and J. A. Thomas, "Elements of information theory 2nd edition," *Wiley-Interscience*: NJ, 2006.
- [36] A. Pourdaryaei, H. Mokhlis, H. A. Illias, S. H. A. Kaboli, S. Ahmad, and S. P. Ang, "Hybrid ANN and artificial cooperative search algorithm to forecast short-term electricity price in deregulated electricity market," *IEEE Access*, vol. 7, pp. 125369-125386, 2019.
- [37] S. Maldonado, R. Weber, and J. Basak, "Simultaneous feature selection and classification using kernel-penalized support vector machines," *Information Sciences*, vol. 181, no. 1, pp. 115-128, 2011.

- [38] M. AlRashidi and K. El-Naggar, "Long term electric load forecasting based on particle swarm optimization," *Applied Energy*, vol. 87, no. 1, pp. 320-326, 2010.
- [39] S. M. Mousavi, E. S. Mostafavi, and F. Hosseinpour, "Gene expression programming as a basis for new generation of electricity demand prediction models," *Computers & Industrial Engineering*, vol. 74, pp. 120-128, 2014.