

Long Term Solar Power Generation Prediction using Adaboost as a Hybrid of Linear and Non-linear Machine Learning Model

Sana Mohsin Babbar, Chee Yong Lau, Ka Fei Thang

School of Engineering, Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur 57000, Malaysia

Abstract—The usage of renewable energy sources has increased manifolds in terms of electric utilities. From other non-conventional sources, solar energy has been seen as a promising and convenient source used around the globe. In terms of meeting the daily requirements, solar energy has huge potential to fulfil the world's demand. However, firstly the characteristic of solar energy is very unpredictable and intermittent due to variation of weather. Secondly, the optimization and the planning of smart grid effect the operation of PV system. Thus, prediction on the long horizon is needed to address this problem. Nevertheless, long term forecasting of solar power generation is deliberated as a challenging problem. Therefore, this paper proposes a 10 day ahead solar power forecasting using combination of linear and non-linear machine learning models. At first, the outputs are generated from Recurrent Neural Network (RNN), Support Vector Machine (SVM) and Autoregressive with exogenous variable (ARX). Later on, these three outputs are combined and are made as a strong classifier with the Adaptive boost (Adaboost) algorithm. The simulations were conducted on the data obtained from real PV plant. By the experimental results and discussion, it was endogenously concluded that the combination of all techniques with the Adaboost have increased the performances and showing the high accuracy as compare to the individual machine learning models. The hybrid Adaboost shows %MAPE 8.88, which proven high accuracy. While on the other hand, for the individual technique, RNN shows 10.88, SVM reveals 11.78 and ARX gives 13.00 of percentage MAPE. The improvement proves that combination of techniques performs better than individual models and proclaims the high accuracy.

Keywords—*Recurrent neural network (RNN); support vector machine (SVM); autoregressive with exogenous variable (ARX); adaptive boosting (Adaboost) and photovoltaic system (PV)*

I. INTRODUCTION

In recent years, the attention has been drawn towards the renewable energies. The importance of wind and solar energies are at glance around the globe. Development of clean and green energy is highly considered by the researches in the present-day [1]. Wind turbines and photovoltaic array (PV) are vital technologies for harvesting energy from solar irradiance, module temperature and wind speed etc. Due to the rapid increase in population and the economy, the demand of electricity has been raised up in the past decade. Consequently, solar power systems and other renewable energy sources are widely used as a solution [2]. The solar

energy is eco-friendly non-conventional sources which is why it is adopted widely. Industrial, commercial and residential applications can be operated by the solar energy sufficiently [3]. While contrastingly, the nature of solar energy is sporadic and unpredictable due to which errors can be caused to the power grid. It can greatly limits the large-scale integration of PV power generation system to the power or smart grid. Forecasting on the large horizon is most promising key solution to the problems of intermittency and volatility. Forecasting energy is used for mitigating the unresolved challenges in the resources. In the research community, solar power forecasting is growing rapidly.

The collection of solar power data is itself a field. The nature of the solar power data set depends upon the model from which it has been collected. Solar power production has become the limelight of sustainability and planning of electricity generation and there are few models of energy generation of PV system. The observed values of solar power or other renewable energies are usually fetched from physical or statistical model. Generally in the past researches, forecasting is done by three models. First one and extensively used are physical models. Physical models usually contain the natural parameters from Numerical Weather Prediction (NWP) models. The forecasted data conventionally is solar irradiance, wind speed, orography, humidity, wind power and solar power, etc. [4]. The second one is statistical model which contains the historical data. And the third one is hybrid model, which is the ensemble of physical and statistical model [5]. In this paper, the input data is taken from physical model.

On the other side, forecasting also depends upon the time horizon and the usage of source. There are three types of forecasting based on the range of time period: (1) Short term forecasting (2) Medium-term forecasting and (3) long term-forecasting. Numerous studies have been carried out for short and medium forecasting which ranges from 3 to 72 hour ahead [6]. Less amount of work has been done for having a long term forecasting. The maximum range of prediction is days ahead. This paper aims to bring originality and predicts 10 days ahead solar power generation. This research also highlights the usefulness of solar power forecasting for the integration of electricity generation. In actual fact, if the prediction is done accurately, the integration and utilization of electricity can be of better service.

The paper is organized in a way that summarizes all aspects of the originality of this research work. The paper is structured as follows: in Sections 2 and 3 brief review of the machine learning models and problem statement has been discussed. In Section 4, proposed methodology is described step by step. In Section 5, the proposed models with implementations are discussed extensively. Whereas, Section 6 shows the 10 days ahead solar power generation prediction using different machine learning algorithms, in terms of root mean square error, mean absolute error and mean absolute percentage error. Furthermore, Section 6 also compares and analyze the precision of all proposed models. In the end, conclusion and future work are stated.

II. LITERATURE REVIEW

Over the few decades, machine learning techniques are substantially used for prediction, forecasting, classification, image and pattern recognition etc. Machine learning and Artificial Intelligence (AI) plays a vital role in forecasting of solar power. Neural networks and various kinds are popular for prediction purposes. Feed-forward Neural Network (FFNN), Multi-layer perceptron (MLP), Recurrent Neural Network (RNN) and Elman Neural Networks are always in demand of researches and studies. Recently, Artificial Neural Networks (ANN) is widely chosen for electricity demand forecasting field [7-8]. On the other hand, Support Vector Machine (SVM) and related regression techniques are also performing well in the field of forecasting. Aforementioned studies showed that SVM performs equally and greater than neural networks and other statistical models [9]. SVM has non-linear features with fitting ability and it can easily predict solar radiation at its fine [10]. While on the other hand, linear models like regression model, Autoregressive model (AR), Autoregressive with exogenous variable (ARX), and Autoregressive integrated moving average (ARIMA) are also widely used for prediction purpose for its time-series analysis with variable co-efficient [11].

There always been a comparison in linear and non-linear models with respect to prediction. It has been seen from the past studies that linear models mostly need time-series data for prediction while non-linear models like ANN, SVM and fuzzy logics does not depend on the historical data. They are usually depend upon the machine learning algorithms to maintain and create a relationship between Numerical Weather Prediction (NWP) data and solar power generation [12]. In [13], a new way of RNN has been proposed for short term solar forecasting. Non-linear features were extracted from data set and then RNN was applied. Previous PV data was taken as input data. Later on, output was compared with persistence model and other machine learning techniques. RNN proved to be better performer in terms of accuracy. RNN is very well known for solving the complex networks and architecture. It is also keenly observed in past studies that was using long-short term memory recurrent neural network (LSTM-RNN) for forecasting can also be applicable for efficacy. This study was evaluated using hourly dataset of almost one year. The results were compared with three PV forecasting methods and showed accurate and precise output. It was also concluded that this propose model is effective tool for controlling and planning of PV grid system [14] while SVM is also a very

promising technique in terms of classification and regression. In [15], SVM with least square has been applied to the historical data. The predicted output was atmospheric transmissivity and then it was converted into short term solar power. By the results, it was observed that SVM flaunted more refined outputs as compare to the conventional AR model and radial base neural network function. Using satellite's image data to predict solar power with the help of SVM has shown remarkable results and then the proposed model was compared with ANN and time-series model to observe the accuracy [16]. Contradictory to the non-linear model, linear models have got their own specifications and characteristics. ARX and ARIMA have been extensively used over the past few years. ARX works best on the time series data and historical data as well. A piece of research showed that ARX as compare to the persistence models can be useful for 24 hour ahead prediction and it can be fully improved in terms of forecasting precision [17]. Solar power were predicted using Gaussian Conditional Random fields (GCRF). After obtaining the results, the final solar power was compared with persistence and ARX model after several experiments with and without pre-processing of data. It was concluded by the RMSE that GCRF has attained the forecasting estimation accurately as compare to persistence and ARX [18].

According to the data type, the good predictive models are selected and then trained to have precise results. Ensemble approaches are gaining popularity in terms of best performance. In the past research, it has been seen at a higher rate that hybrid approach of linear or non-linear models brings the better accuracy and performance. The combination and contrast of linear and non-linear models brings explicit results [19]. Numerous works has been done in the hybrid machine learning approaches. A research showed well-refined results using combination of different linear and non-linear models. A data of wind speed was taken from NWP and then employed to the Feed-forward Neural Network (FFNN), SVM and regression model. 72 hour ahead prediction was made into the consideration. Quantifying measures showed that ensemble approaches of different algorithms can perform much better as compare to the individual ones [20]. In [21], different training models i.e. Levenberg-Marquardt (LM) and Bayesian Regularization (BR) embedded in neural networks are used for solar power forecasting. It was observed that BR outperformed and brought significant results. Hybrid approaches using machine learning tools are attractive proposed methodologies with respect to the performance. Different kind of adaptive boosting techniques have been used in the past. Adaboost is best known for its combination with the weak classifier to form a strong classifier. In [22], ensemble approach of sparse Adaboost with Echo State Network (ECN) has been implemented on the electrical consumption in the Hubei, China. Aforementioned studies depicts that ensemble approach is a good choice for Industrial Electrical Consumption (IEC) planning and controlling.

Lots of review studies have been executed for analyzing the hybrid behavior. Ensemble approaches are categorized into two main categories: (1) Competitive ensemble forecasting and (2) Cooperative ensemble forecasting. Competitive

ensemble forecast is about training the different kinds of predictor with same or different datasets while cooperative divides the task to the different predictors and finally sums up the outputs [23]. In this paper, competitive ensemble approach has been adopted due to the parameters diversity.

In this work, linear and non-linear models are combined with each other using Adaboost algorithm. It makes use of different inputs obtained by the physical model i.e. solar irradiance, module temperature and solar power with per minute resolution. Non-linear models like RNN and SVM, while linear models like ARX are employed to the input data individually. Later on, the output obtained from the individual models are trained and combined with the assistance of Adaboost to form a strong classifier. They are trained with accurate ratio of training and testing to have best forecast accuracy with minimal error.

III. PROBLEM STATEMENT

Due to higher rate of fluctuations and intermittency in solar energy, there is always a need of accurate predictive model. In this regard, this paper provides an extensive comparison and combination of different machine learning approaches. The predictive models used in this study are RNN, SVM and ARX. Later then, the outputs from these predictive models are fed to Adaboost. This boosting technique is used in this research as an ensemble approach. In context of combining the techniques, several problems have been noticed in the past research. Firstly, lack of quality data and pre-processing of data. Secondly, choosing the right parameters for predictive models chosen. Lastly, the capability of churning the load of bid data sets. These are the key problems which usually researchers face while making combining the different machine learning models [24]. Therefore, Adaboost is chosen as combination of RNN, SVM and ARX for addressing these problems related to the data set because this combiner understands how to resist from over-fitting which makes a better hybrid approach [25].

Major contributions of the solar power prediction generation differ from other methods shown in the literature review, are listed down in the following aspect:

1) A new approach of combination of linear and non-linear models with different inputs are considered. Hybrid approach is contemplated with Adaboost, which has brought originality in the work done. Adaboost has chosen for combining the techniques because it efficiently increases the accuracy of predictors. It constructs the strong classifier by combining all weak classifiers with the poor performance and help them to achieve high accuracy.

2) Long term solar power prediction is proposed with the ensemble approach. Least amount of work has been carried out in the past for the long term forecasting. The main objective was to predict 10 day ahead solar power for combating the problem of variation of weather climate. Trading of electric supply to the smart grid shows errors sometimes due to intermittent nature of solar energy. For combating this issue in controlling and planning, a system is designed and implemented.

By the literature review and several experiments, it has been observed keenly that individual model would not be enough to have sharp accuracy. It was a great challenge to foster the precision with individual models. For the sake of having precision, amalgamation of different machine learning models has been proposed.

IV. PROPOSED METHODOLOGY

The multi-stage solar power forecasting techniques are generically comprises of three main stages. Firstly, data set is obtained from physical model and is molded and sifted according to the desired output then the predictive models are picked as stated in the past researches. Secondly, the outputs taken from linear and non-models are taken as an input for the ensemble model. Lastly, the hybrid model is trained and exhibits the final predicted output. The block diagram is shown in Fig. 1, which demonstrates the long term solar power forecasting using amalgamation of different machine learning algorithms with Adaboost model. Each stage is described below briefly:

Stage 1: First and foremost step is to analyse the data set. The availability of data set is very important for training the predictive model. The data was obtained from a physical model. The data set contains solar power (MW), solar irradiance (W/m^2) and module temperature and these parameters are the inputs of the RNN, SVM and ARX model. The division of data set is decided by the hit and trail rule, 70% of the data is set for the training purpose while 30% for the testing and validation. For the training purpose, the data is divided into months and 12283 samples are taken per month.

Stage 2: After sifting the data according to the nature of predictive models, the input data (12283×3) is fed to the RNN model first. The same input data is designated to SVM and then ARX. Every model shows us the solar power as their outputs. These outputs are now taken as an input to the ensemble approach proposed in this methodology. The process of how each model has performed is shown in Fig. 2.

Stage 3: The outputs taken from RNN, SVM and ARX are actually serving as an input to the Adaboost algorithm. All the predictive tools used in the proposed model are the weak classifiers and they are trained again with new and updated weights. After having new weights, strong classifier is formed by applying signum function as shown in the Fig. 3. In the end, the strong classifier shows the final predicted solar power with precision and accuracy.

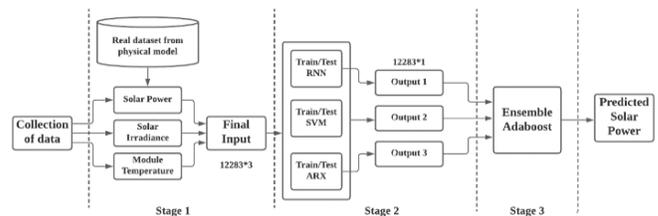


Fig. 1. Block Diagram of Proposed Methodology.

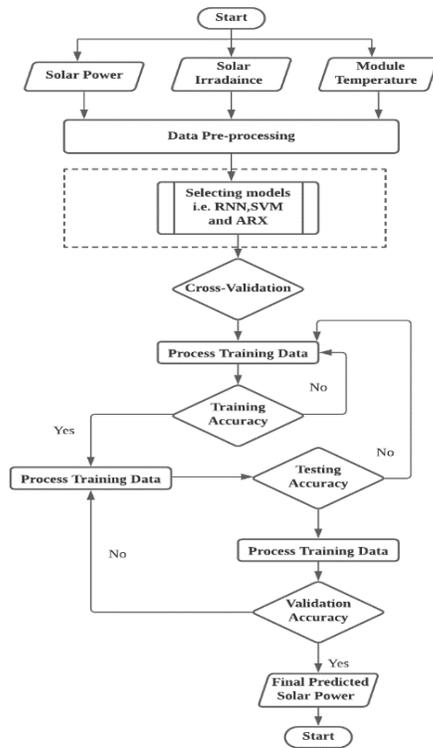


Fig. 2. Flow Chart of Performance of Stage 2 in Proposed Model.

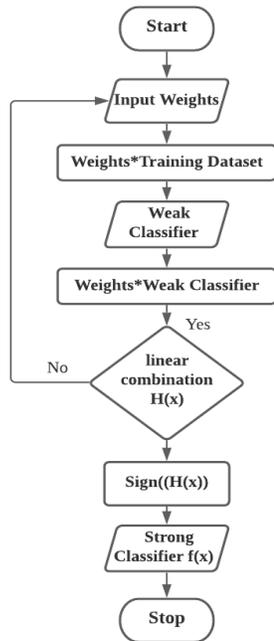


Fig. 3. Process of Adaboost Algorithm.

V. APPLICATIONS OF MODEL USED

The implementation and design of solar power forecasting has been enhanced with the methodology proposed in this paper. This research aims to determine the efficient use of solar power forecasting in terms of solar energy trading and management of electricity grid. It also emphasize in the planning and controlling of the PV power system. Due to the

variation in climate, difficulties occur in the generation of energy and trading the electrical energy to the power grid. Therefore, a multi-stage long term solar power generation has been predicted with the use of ensemble machine learning approaches [26].

Observations has been made by the past researches that combining models can bring efficacy in the predicted output. There are ample of advantages of using ensemble approaches. Combination of models brings more refined results as compare to the individual models. It also reveals good results with small datasets, as it removes all necessary outliers due to several iterations. Ensemble approach can easily be applicable to other domain of energy management systems. As described in the proposed methodology, RNN, SVM and ARX are the input generators for the hybrid approach used in this research. In this section, all the machine learning approaches are explained briefly with their working schematics. All the proposed models are using three basic parameters as an input i.e. solar power (MW), solar irradiance (W/m^2) and model temperature (K). In terms of predicting solar power, solar irradiance and power plays a vital role. Accurate solar power forecasting can help the grid operations to be optimized the electricity production.

A. Implementation of RNN

RNN is a different type of neural network with deals with a complex structure and data sets. It can deal with multiple hidden layers with a back propagation function. The training model used in this neural network is Levenberg-Marquadt (LM). LM is highly adaptive training model used in the neural networks. It has significant number of adaptive weights with respect to the training. It is also very well-known about its memory and operations [27]. RNN are also known as traditional kind of Elman neural networks. The output summation of RNN is stated in Equation 1.

$$v(t) = f(h(t - 1), x(t); b) \quad (1)$$

Where $v(t)$ denotes the output obtained after training. While $h(t)$ operates as a function which deals with the past sequence with respect to the input $x(t)$ and b shows the parameters chosen in the study. Now $v(t)$ is transferred through an activation function shown below in Equation 2.

$$y(t) = \emptyset(v(t)) \quad (2)$$

The reason that solar power generation being a positive number, sigmoid activation is considered for boosting up the training. It's a connection between output and hidden layer shown in Equation 3.

$$\emptyset = \frac{1}{(1 + \exp(-az))} \quad (3)$$

Where a is the slope and z is an input of activation function.

For having the optimal solution, the right parameters and the values should be chosen accordingly. Hit and trial method is always adopted to find out the number of hidden layers between input and output. After several trials, 30 layers were chosen to quantify the output. Supervised machine learning technique is introduced for input and output pairing. Main

objective of training the model is to minimize the biases and errors at its best. The process of having solar power as an output is shown in Fig. 4 with the assistance of flow chart.

B. Implementation of SVM

SVM is a popular supervised machine learning model which is a leading classifier [28]. By the past researches, SVM has gained the attention due to its high accuracy. Every so often, it shows accuracy more than any neural network and sometimes gives equal to it [29]. For forecasting purpose, SVM trains the response values at low-dimensions and shows good results. In SVM, the data is separated by hyper plane and defines classes. More the data is closed to the hyper plane, more it is incline towards high accuracy [30]. The expression of the SVM is given below in Equation 4.

$$y = \sum_{i=1}^n w K(\cdot) + b_0 + \varepsilon \quad (4)$$

Where y is the corresponding output, n is the training samples, w is the value of weights, $K(\cdot)$ is the kernel function, b_0 is the bias while ε is the error.

There are several types of kernel function used in the past studies. It is very critical to choose kernel function which should be operational to both linear and non-linear models. Radial base kernel function has been chosen for this paper. This function is also selected after several trials. The main reason behind choosing Radial base kernel function is to minimize the distance and ranges between input points and the hyper-plane. Appropriate kernel function is always obtained after the continuous iterative process until the square of the error gets decrease [31].

The process of working of SVM in this study is illustrated in Fig. 5. The operation shows that after selecting the input parameters i.e. solar power, solar irradiance and module temperature; the functional parameters are set for training the SVM model. After successive training, converging fitness model is calculated for model's accuracy. If solar power is achieved with minimal percentage error as output then SVM model will get stop else it will again select the weights, kernel function and train itself again.

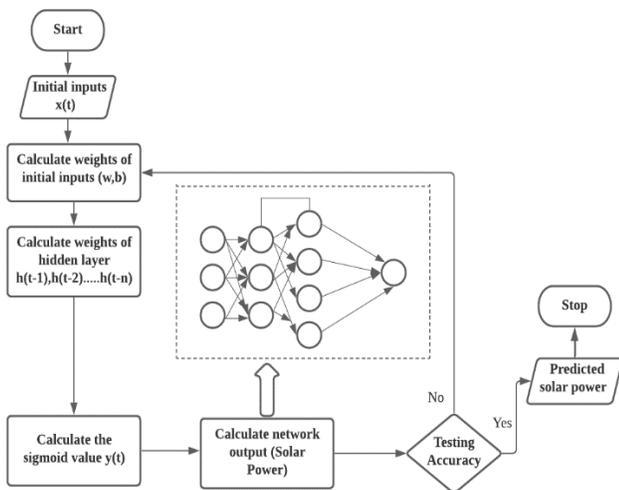


Fig. 4. Flow Chart showing Process of RNN.

C. Implementation of ARX

ARX is used as a linear machine learning model in the proposed method. ARX is one of the widely adopted linear technique in the past studies [32]. ARX is basically a time-series linear model which creates a relationship between dependent and independent variable. The basic criterion in ARX is choosing the exogenous variable based on priori analysis. In this study, module temperature and solar irradiance are the dependent variables for having the solar power at the output [33]. Accuracy of the model is calculated based on the co-efficient and training phase. The ARX model is demonstrated in the Equation 5 below [34]:

$$A(z)y(t) = B(z)x(t) + e(t) \quad (5)$$

Where $y(t)$ is the desired output i.e. solar power (MW), $x(t)$ is an input, $e(t)$ is the white noise produced in the system. While $A(z)$ and $B(z)$ are the coefficients varies with the time delay. Fig. 6 shows the process of ARX model which is implemented in this study. Operation of ARX model is started with initializing input i.e. solar power (MW). After introducing the input, the estimator is chosen according to the complexity of dataset. Cascaded feed-forward neural network is selected with 20 hidden layers to have better accuracy. Later then, ARX model is applied with a delay of 0.2 seconds with coefficients $A(z)$ and $B(z)$. In this model, forecasting is estimated with the help of Mean Square Error (MSE). Uncertainly, if model is giving good accuracy then ARX model is ready to show the predicted solar power (MW). Otherwise, again the estimator is selected and training will be held.

In past researches, ARX is constructed for predicting the solar power with good accuracy. It was also observed that ARX model or other regression techniques gives more accuracy when combine with other machine learning techniques [35].

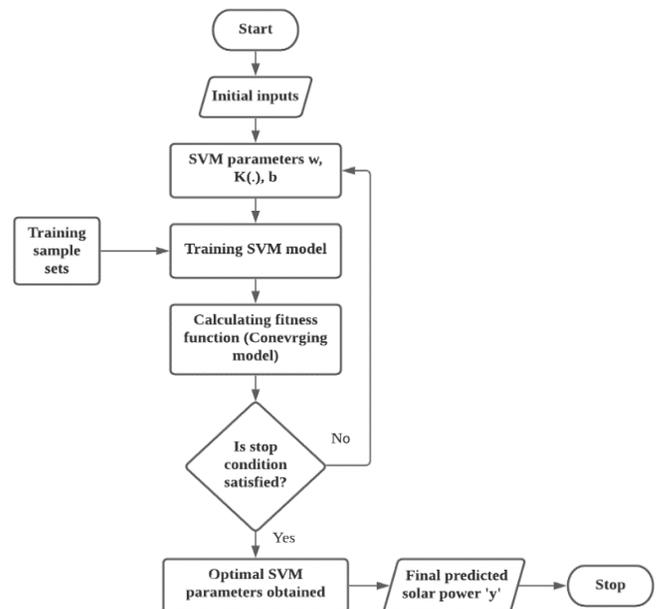


Fig. 5. Flow Chart showing the Process of SVM.

D. Hybrid of all Models using AdaBoost

In this study, the predictive models are combined by the Adaboost technique. Adaboost plays an important and vital role in this research. The main purpose of Adaboost is to ensemble machine learning models and improves the accuracy and precision. Adaboost has been extensively adopted in the recent years [36]. In the mechanism of this model, weights are generated at each step of algorithm for better classification and prediction. In this paper, outputs taken from RNN, SVM and ARX are fed to Adaboost to form a strong classifier as shown in Fig. 7 [37].

The figure shows the generic schematics of Adaboost model. The expression below depicts the function of n=3 variables used as an input.

$$f(x_1, x_2, x_3) = y \tag{6}$$

Where x_1, x_2, x_3 are the outputs taken from the RNN, SVM and ARX while y is the final predicted solar power.

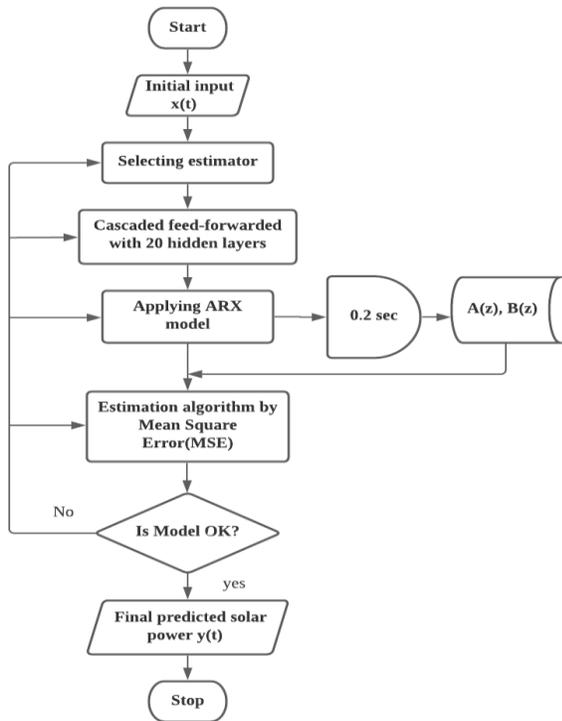


Fig. 6. Flowchart showing the Process of ARX.

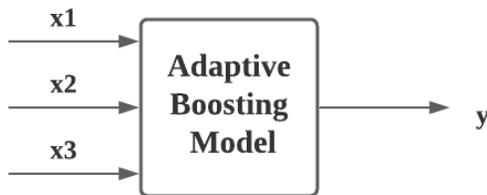


Fig. 7. Block Diagram of Output Fed to the Adaboost.

There are mainly two parts consists in the Adaptive boosting technique: First is step-by-step forward model and second is additional model [38]. In this paper, additional model is adopted for the linear combination of weak classifiers. The algorithm prepared for this research is demonstrated below in the Table I. The Pseudo code below describes well the whole process of Adaboost. In this paper, additional model Adaboost type is used. The expression of additional model is shown below in the Equation 7:

$$H(x) = \sum_{t=1}^T a_t h_t(x) \tag{7}$$

In the above equation, $H(x)$ shows the linear combination of weak classifier, whereas $h(x)$ is a weak classifier itself with a product of at weights generated of every step. On the other hand, there is also an expression found to calculate the weighted strategy, which is depicted below in Equation 8:

$$a_t = 0.5 \ln \left(\frac{1 - \text{total error}}{\text{total error}} \right) \tag{8}$$

Where total error is the classification error at the nth iteration while training. Lastly, the strong classifier is formed by the linear combination of weak classifier with the assistance of signum function as shown in Equation 9:

$$f(x) = \text{sign}(H(x)) \tag{9}$$

However, sign function is a symbolic function which converts the prediction results into the desired output. The block diagram of strong classifier is demonstrated in the Fig. 8.

TABLE I. ALGORITHM OF ADAPTIVE BOOSTING LEARNING MODEL

Pseudo code of the Adaptive boost algorithm	
Input:	The data A; The number of weak classifier T
Output:	The strong classifier $f(x)$
1:	Initializing sample weights a_1, a_2, \dots, a_t
2:	for $t=1; t < T; t++$ do
3:	Training weak classifier h_t based on a_t and A.
4:	Calculating the total classification error
5:	Calculating the weight a_t of the weak classifier h_t , $a_t = 0.5 \ln(1 - \text{total error} / \text{total error})$
6:	Updating new sample weights = $w_t * e^{a_t}$
7:	Forming a linear combination of weak classifiers, $H(x) = \sum_{t=1}^T a_t h_t(x)$
8: end for	
9:	Forming a strong classifier, $f(x) = \text{sign}(H(x))$
10: return	$f(x)$

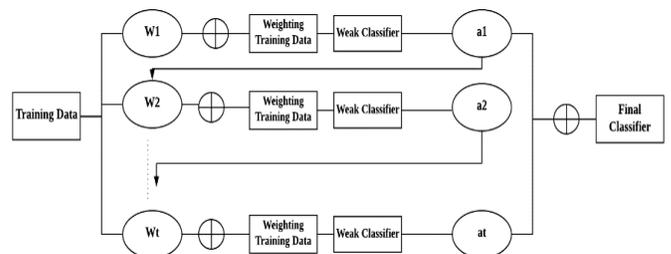


Fig. 8. The Schematics of Adaptive Boost Model.

VI. RESULT AND DISCUSSION

Training period of data set is 9 months from July-2019 to Mar-2020, solar power and solar irradiance varies according to the weather changes. Noticeably, the data set was maintained in per minute resolution according to each day. Solar irradiance plays an important role in solar power forecasting. For having a smooth power supply and balance in energy trading, it is necessary to have predicted solar power along solar irradiance. The nature of solar energy is sporadic in nature due to so many climate variations and deviation in the solar radiation. The solar power can be varied minute to minute as shown in Fig. 9. The characteristic of solar power below in the plot shows the behaviour of solar power from 8:00 am to 5:30 pm respectively. It can be observed easily that every day, every minute even every second matters in terms of forecasting. Therefore, an effective method of prediction is proposed to address this issue for generating solar power using different machine learning techniques.

In this section, individual and combined behaviour of all proposed model is discussed briefly in order. Comparison of all the models is also performed with the Adaboost to observe the performance of each technique with the performance indices. There are different kind of quantifying measures used in the past researches for observing the accuracy and precision. In this paper, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Percentage Mean Absolute Error (%MAPE) are chosen as performance evaluators and are expressed below:

$$RMSE = \frac{\sqrt{\sum_{i=1}^N (y - y')^2}}{N} \tag{10}$$

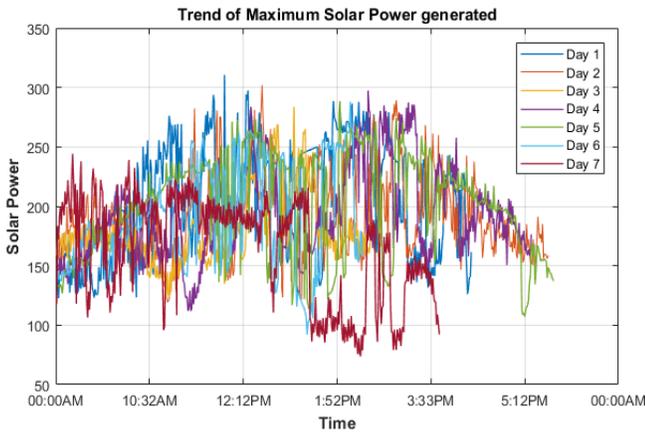


Fig. 9. Trend of Solar Power.

$$\%MAPE = \frac{1}{N} \sum_{i=0}^n \left| \frac{y - y'}{y} \right| \times 100 \tag{11}$$

$$MAE = \frac{\sum_{i=0}^n |y - y'|}{N} \tag{12}$$

Where y is the observed value and y' is the predicted value. For a comparative analysis, these values are contrasted with each model to check the efficacy and precision. While on the other hand, the ranges of good and worst accuracy have also been setup by checking percentage MAPE. Table II shows the estimation of accurate forecasting.

TABLE II. ESTIMATION OF FORECASTING ACCURACY BY THE %MAPE CRITERION

%MAPE	Forecasting Evaluation
≤ 10%	High accuracy
% 10 to 20%	Good accuracy
20% to 50%	Reasonable accuracy
≥ 50%	Inaccurate accuracy

A. Comparison of RNN, SVM and ARX

In the first stage, the outputs are taken from RNN, SVM and ARX individually. The outputs are matched with the response taken as a catalyst in the proposed predictive models. Each model is trained separately with the data set of different months. The process of obtaining output is already shown in Fig. 4.

Firstly, RNN was trained and the output was validated until it shows a refine output with minimal errors as shown in the Fig. 10. In all the figures discussed in this section represents time on the horizontal axis while solar power (MW) on the vertical axis. The trend between output and response is plotted in the graph. The blue colored segment illustrates the predicted solar power obtained from RNN while red trend shows the target chosen for training purpose. The plot is constructed on the test indices. The solar power characteristic illustrates that there is serrated relation between output and the target. It is seen that the RMSE, MAE and %MAPE by the RNN model for the month of August, 2019 is 30.18, 37.40 and & 11.29%, respectively. According to the Table II, the %MAPE lies in the category of good accuracy. It is also concluded that RNN removes up to 70 percent of the errors between actual and predicted values.

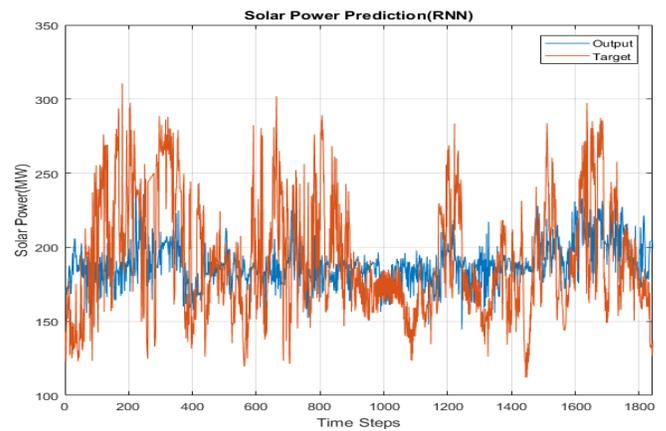


Fig. 10. Solar Power Characteristic Obtained by RNN.

After the training of RNN, same dataset is applied to the SVM for the same purpose. In the past studies, it was seen that SVM outperforms better than any other neural network [39]. In this proposed methodology, RNN and SVM performed approximately equal to each other with a minor difference. According to the performance indicators, SVM has attained 32.25 of RMSE, 40.55 of MAE while 11.77 percent of %MAPE. The output gained by the SVM is demonstrated in the Fig. 11. In the mentioned figure below, the red peaks

shows target while the blue highlights the output i.e. solar power (MW). The relationship between target and the output attained by the SVM is quite level-headed. By the criterion of percentage MAPE, SVM has showed good accuracy slightly equal as compare to the RNN.

In line with, the implementation of non-linear model is more adaptive as compare to the linear models in aforementioned studies. It depends on the data set that which model is working well. In the perspective of statistical analysis, regression models like ARX, ARMA and ARIMA performs well on the skewed and historical data while non-linear predictive tools such as neural networks, support vector machines and fuzzy logics performs excellent on the real data [40].

Together with non-linear models, ARX is applied to the same data set as this research heads to competitive ensemble approach. As the data used in this study was real data, ARX showed good accuracy as depicted in Fig. 12. In the graph below, the blue crests and troughs shows target while red ones are showing the output obtained after the training by ARX. It attained good accuracy with respect to the MAPE criterion showing 13.81 percentage MAPE. While on the other hand, ARX obtained 35.09 of RMSE and 44.29 of MAE. Conclusively, it is observed that RNN outperformed well as compare to the SVM and ARX in this research. The order of the performance is shown below in the expression:

$$RNN > SVM > ARX$$

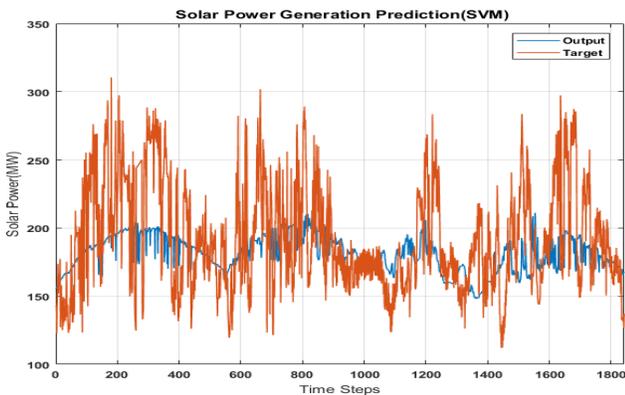


Fig. 11. The Solar Power Characteristic Obtained by SVM.

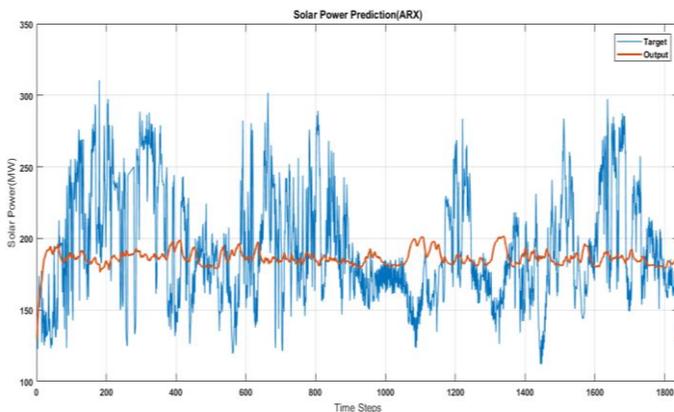


Fig. 12. The Solar Power Characteristic Obtained by ARX.

B. Comparison of Techniques with Adaboost

Adaptive boosting algorithm has been seen as a hallmark in terms of combining classifiers [41]. In terms of prediction and forecasting, any kind of boosting technique would perform well and gives suitable accuracy and precision as compare to other individual machine learning techniques [42]. In this paper, RNN, SVM and ARX are combined together and formed a strong classifier with the help of Adaboost algorithm. By the results and simulations, it was keenly observed that Adaboost is showing some major decrease in the performance indicators as shown in Table III. The statistical figures in the table depicts that Adaboost has served as strong classifier and has obtained the high performance among all other models proposed. By looking at the Table III, it is obvious that hybrid of linear and non-linear models showed up with high accuracy of forecasting. Almost in all months from August 2019 to March 2020, the results are quite same but there are fluctuations and disparities due to the weather changes.

C. Comparison of all Models used

By the brief analysis of results and simulations, it was concluded that all linear and non-linear models used in the proposed method are showing good accuracy while Adaboost is attaining high accuracy by the combination of RNN, SVM and ARX. Fig. 13 demonstrates the performance of each model and showing that Adaboost has lowest %MAPE well as compare to the other machine learning models. In Fig. 14, it is depicted that Adaboost outperforms well than RNN, SVM and ARX by the means of quantifying measures as shown in the expression below.

$$Adaboost > RNN > SVM > ARX$$

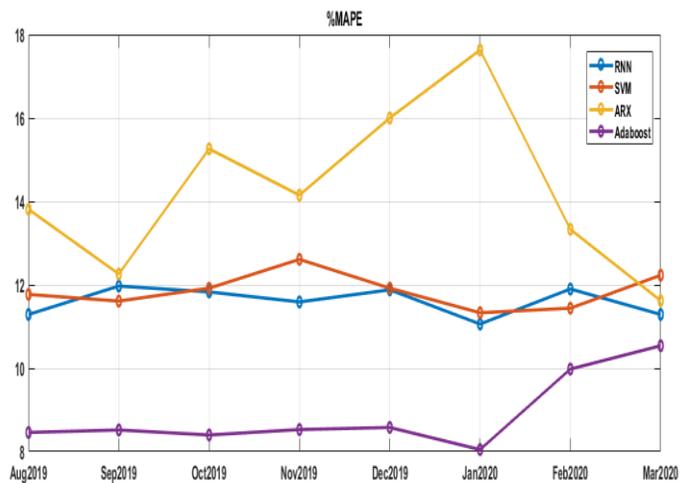


Fig. 13. %MAPE of Proposed Models.

In the aforementioned studies, the performance of ensemble approach is always better than individual models. Latently, the data set went through lots of variant processes and training sessions which finally show the precision in the prediction. In [26], simple neural networks are used and %MAPE has been calculated. It has bought 9.97% accuracy. On the other hand, the proposed model in this research has

shown more precision till 8.05. In this study also, the combination of techniques forming Adaboost gives major goals of attaining high accuracy and led to successful long term solar power generation prediction. Hence, it shows the combination of technique can achieve high accuracy and brings effective results.

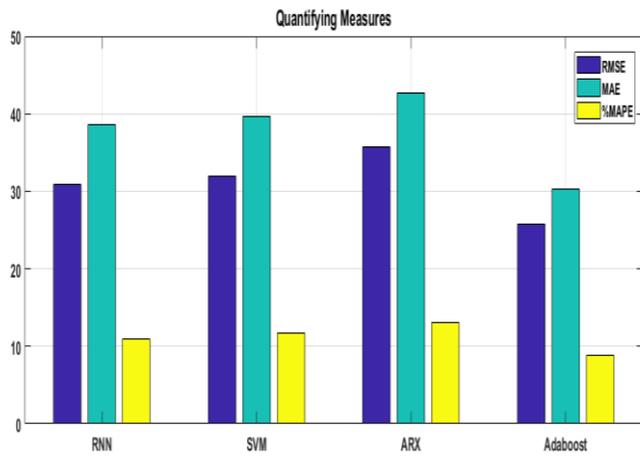


Fig. 14. Performance Indicator of Proposed Model.

TABLE III. QUANTIFYING MEASURES OBTAINED BY THE PROPOSED METHOD

Months	Errors	RNN	SVM	ARX	Adaboost
August 2019	RMSE	30.18	32.45	35.09	25.25
	MAE	37.40	40.35	44.29	29.56
	%MAPE	11.29	11.77	13.81	8.46
September 2019	RMSE	31.54	31.79	31.57	25.26
	MAE	39.22	39.10	37.03	29.6
	%MAPE	11.97	11.61	12.26	8.52
October 2019	RMSE	31.47	32.6	35.9	25.27
	MAE	39.18	40.72	43.06	29.7
	%MAPE	11.83	11.92	15.26	8.40
November 2019	RMSE	30.8	32.42	34.12	25.3
	MAE	38.2	40.36	40.25	29.5
	%MAPE	11.59	12.16	14.15	8.53
December 2019	RMSE	31.39	32.28	41.00	25.23
	MAE	39.5	40.48	38.88	29.6
	%MAPE	11.88	11.92	6.00	8.58
January 2020	RMSE	29.87	31.20	44.00	24.22
	MAE	39.6	41.32	57.20	28.4
	%MAPE	6.06	11.33	17.63	8.05
February 2020	RMSE	31.38	31.31	34.14	27.15
	MAE	38.6	37.89	44.24	31.86
	%MAPE	11.90	11.44	13.33	9.98
March 2020	RMSE	30.19	31.97	30.30	28.48
	MAE	37.40	36.5	36.40	34.03
	%MAPE	11.29	12.23	11.62	10.54

VII. CONCLUSION AND FUTURE WORK

The importance of renewable energy has been taken into the consideration since last two decades. Economic dispatch, integration to the power grid and mismanagement in the power management system due to the variability in the solar energy are the key issues to address. Therefore, to resolve these problems, an effective model is proposed to predict the

solar power generation for 10 days ahead using the hybrid approach. The ensemble technique utilizing RNN, SVM and ARX with the adaptive boosting classifier is presented in this paper. The results and discussion brought a significance stance that using adaptive boosting algorithm for combination of linear and non-linear models produces good results and shows high accuracy. From the output, it was seen that percentage MAPE of solar power characteristic lies between the ranges of 8 to 10 percent, which is a high accuracy for forecasting estimation. On the other hand, hybrid approach has obtained 25.77 of RMSE and 30.28 of MAE respectively. It is noteworthy that the optimizations and improvement brought in solar power generation prediction using combination of RNN, SVM and ARX with the Adaboost is significant and applicable.

REFERENCES

- [1] Nema, P., Nema, R. K., and Rangnekar, S. (2009) A current and future state of art development of hybrid energy system using wind and PV-solar: A review. *Renewable and Sustainable Energy Reviews*, 13(8), 2096-2103.
- [2] Perera, K. S., Aung, Z., and Woon, W. L. (2014) Machine learning techniques for supporting renewable energy generation and integration: a survey. In *International Workshop on Data Analytics for Renewable Energy Integration* (pp. 81-96). Springer, Cham.
- [3] Shaikh, M. R. S (2017). A review paper on electricity generation from solar energy.
- [4] Sun, X., and Zhang, T. Solar power prediction in smart grid based on NWP data and an improved boosting method (2017). In *2017 IEEE International Conference on Energy Internet (ICEI)* (pp. 89-94). IEEE.
- [5] Ozkan, M. B., and Karagoz, P. (2015) A novel wind power forecast model: Statistical hybrid wind power forecast technique (SHWIP). *IEEE Transactions on industrial informatics*, 11(2), 375-387.
- [6] Abdel-Nasser, M., and Mahmoud, K. (2019) Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing and Applications*, 31(7), 2727-2740.
- [7] Abuella, M., and Chowdhury, B.(2015) Solar power forecasting using artificial neural networks. In *2015 North American Power Symposium (NAPS)* (pp. 1-5). IEEE.
- [8] Mohsin, S., Ramli, S.N. and Imdad, M., (2021) Medium-Term Wind Speed Prediction using Bayesian Neural Network (BNN). *International Journal of Systematic Innovation*, 6(5), pp.11-20.
- [9] Buwei, W., Jianfeng, C., Bo, W., and Shuanglei, F. (2018) A solar power prediction using support vector machines based on multi-source data fusion. In *2018 International Conference on Power System Technology (POWERCON)* (pp. 4573-4577). IEEE.
- [10] Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., and Antonanzas-Torres, F. (2016) Review of photovoltaic power forecasting. *Solar Energy*, 136, 78-111.
- [11] Yang, C., and Xie. (2012) A novel ARX-based multi-scale spatio-temporal solar power forecast model. In *2012 North American Power Symposium (NAPS)* (pp. 1-6). IEEE.
- [12] Aggarwal, S. K., and Saini, L. M. (2014) Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013-14 Solar Energy Prediction Contest. *Energy*, 78, 247-256.
- [13] Li, G., Wang, H., Zhang, S., Xin, J., and Liu, H. (2019) Recurrent neural networks based photovoltaic power forecasting approach. *Energies*, 12(13), 2538.
- [14] Lee, D., and Kim, K. (2019) Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. *Energies*, 12(2), 215.
- [15] Zeng, J., and Qiao, W. (2013) Short-term solar power prediction using a support vector machine. *Renewable Energy*, 52, 118-127.

- [16] Jang, H. S., Bae, K. Y., Park, H. S., and Sung, D. K. (2016) Solar power prediction based on satellite images and support vector machine. *IEEE Transactions on Sustainable Energy*, 7(3), 1255-1263.
- [17] Duran, M. J., Cros, D., and Riquelme, J. (2007) Short-term wind power forecast based on ARX models. *Journal of Energy Engineering*, 133(3), 172-180.
- [18] Zhang, B., Dehghanian, P., and Kezunovic, M. (2016) Spatial-temporal solar power forecast through use of Gaussian conditional random fields. In 2016 IEEE Power and Energy Society General Meeting (PESGM) (pp. 1-5). IEEE.
- [19] AlKandari, M., and Ahmad, I. (2020) Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Applied Computing and Informatics*.
- [20] BABBAR, S. M., and LAU, C. Y. (2020) Medium Term Wind Speed Forecasting using Combination of Linear and Nonlinear Models. *Solid State Technology*, 63(1s), 874-882.
- [21] Younis, O., (2018). A predictive model for solar photovoltaic power using the Levenberg-Marquardt and Bayesian regularization algorithms and real-time weather data. *International Journal of Advanced Computer Science and Applications*, 9(1).
- [22] Wang, L., Lv, S. X., and Zeng, Y. R. (2018) Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. *Energy*, 155, 1013-1031.
- [23] Ren, Y., Suganthan, P. N., and Srikanth, N. (2015) Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 50, 82-91.
- [24] Schleder, G. R., Padilha, A. C., Acosta, C. M., Costa, M., and Fazzio, A. (2019) From DFT to machine learning: recent approaches to materials science—a review. *Journal of Physics: Materials*, 2(3), 032001.
- [25] Wang, F., Li, Z., He, F., Wang, R., Yu, W., and Nie, F. (2019) Feature learning viewpoint of AdaBoost and a new algorithm. *IEEE Access*, 7, 149890-149899.
- [26] Alanazi, M., Alanazi, A., and Khodaei, A. (2016) Long-term solar generation forecasting. In 2016 IEEE/PES transmission and distribution conference and exposition (T&D) (pp. 1-5). IEEE.
- [27] Lera, G., and Pinzolas, M. (2002) Neighborhood based Levenberg-Marquardt algorithm for neural network training. *IEEE transactions on neural networks*, 13(5), 1200-1203.
- [28] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., and Lopez, A. (2020) A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215.
- [29] Zhang, Y. (2012) Support vector machine classification algorithm and its application. In *International Conference on Information Computing and Applications* (pp. 179-186). Springer, Berlin, Heidelberg.
- [30] Eseye, A. T., Zhang, J., and Zheng, D. (2018) Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. *Renewable Energy*, 118, 357-367.
- [31] Hsu, C. W., Chang, C. C., and Lin, C. J. (2003) A practical guide to support vector classification.
- [32] Guo, Y., Nazarian, E., Ko, J., and Rajurkar, K. (2014) Hourly cooling load forecasting using time-indexed ARX models with two-stage weighted least squares regression. *Energy Conversion and Management*, 80, 46-53.
- [33] Mechaqrane, A., and Zouak, M. (2004) A comparison of linear and neural network ARX models applied to a prediction of the indoor temperature of a building. *Neural Computing & Applications*, 13(1), 32-37.
- [34] Shadab, A., Ahmad, S., and Said, S. (2020) Spatial forecasting of solar radiation using ARIMA model. *Remote Sensing Applications: Society and Environment*, 20, 100427.
- [35] Brownlee, J. (2016) A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning.
- [36] Potočník, P., Vidrih, B., Kitanovski, A., and Govekar, E. (2019) Neural network, ARX, and extreme learning machine models for the short-term prediction of temperature in buildings. In *Building Simulation* (Vol. 12, No. 6, pp. 1077-1093). Tsinghua University Press.
- [37] Tang, D., Tang, L., Dai, R., Chen, J., Li, X., and Rodrigues, J. J. (2020) MF-Adaboost: LDoS attack detection based on multi-features and improved Adaboost. *Future Generation Computer Systems*, 106, 347-359.
- [38] Javed, A., Kasi, B. K., and Khan, F. A. (2019) Predicting Solar Irradiance Using Machine Learning Techniques. In 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC) (pp. 1458-1462). IEEE.
- [39] Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., and Gupta, B. (2018) Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of computational science*, 27, 386-393.
- [40] Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P. (2016) Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy conversion and management*, 112, 115-124.
- [41] Ferreira, J. M., Pires, I. M., Marques, G., Garcia, N. M., Zdravevski, E., Lameski, P., and Spinsante, S. (2020) Identification of daily activities and environments based on the adaboost method using mobile device data: A systematic review. *Electronics*, 9(1), 192.
- [42] Kamalasri, D., Prasath, J. A., and Prabu, R. T. (2015) Solar Radiation Prediction using Adaboost Algorithm.