IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 19, 2021, accepted August 30, 2021, date of publication September 7, 2021, date of current version September 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3110774

Machine Learning-Based Predictive Techno-Economic Analysis of Power System

HEPHZIBAH JOSE QUEEN[1], JAYAKUMAR J.[1], DEEPIKA T. J.[2], K. VICTOR SAM MOSES BABU[3], (Member, IEEE), AND SURYA PRAKASH THOTA[4], (Senior Member, IEEE)

[1]Department of Electrical and Electronics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu 641114, India
[2]Department of Electrical and Electronics Engineering, Sona College of Technology, Salem, Tamil Nadu 636005, India
[3]Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani–Hyderabad Campus, Hyderabad, Telangana 500078, India
[4]Department of Electrical and Electronics Engineering, G. Narayanamma Institute of Technology and Science (A), Hyderabad, Telangana 500008, India

Corresponding author: Hephzibah Jose Queen (hephzijose@gmail.com)

ABSTRACT This paper proposes a predictive techno-economic analysis in terms of voltage stability and cost using regression-based machine learning (ML) models and effectiveness of the analysis is validated. Predictive analysis of a power system is proposed to address the need for faster and accurate analyses that would aid in the operation and control of modern power system. Several methods of analyses including metaheuristic optimization algorithms, artificial intelligence techniques and machine learning algorithms are being developed and used. Predictive ML models for two modified IEEE 14-bus and IEEE-30 bus systems, integrated with renewable energy sources (solar and wind) and reactive power compensative device (STATCOM) are proposed and developed with features that include hour of the day, solar irradiation, wind velocity, dynamic grid price and system load. An hour-wise input database for the model development is generated from monthly average data and hour-wise daily curves with normally distributed standard deviations. The data feasibility tests and output database generation is performed using MATLAB. Linear and higher order polynomial regression models are developed for the 8760hr database using Python 3.0 in JupyterLab and a best-fit predictive ML model is identified by analysing the coefficients of determination. The voltage stability and cost predictive ML models were tested for a 24hr input profile. The results obtained and the comparison with the expected values are furnished. Prediction of the outputs for the test data validate the accuracy of the developed model.

INDEX TERMS Cost analysis, machine learning, predictive analysis, renewable generation, STATCOM, voltage stability.

NOMENCLATURE

ANN	Artificial Neural Network
CV	Cross Validation
DT	Decision Tree
DG	Distributed Generation
ELM	Extreme Learning Machine
FACTS	Flexible Ac Transmission System
FVSI	Fast Voltage Stability Index
LQP	Line Stability Factor
LTVS	Long Term Voltage Stability
ML	Machine Learning
MLT	Machine Learning Techniques
NR	Newton Raphson
PCA	Principal Component Analysis
PV	Photo Voltaic
RES	Renewable Energy Sources
RMSE	Root Mean Square Error
STATCOM	Static Synchronous Compensator
STVS	Short Term Voltage Stability
SVM	Support Vector Machine
TEA	Techno Economic Analysis
VCPI	Voltage Collapse Point Indicators

The associate editor coordinating the review of this manuscript and approving it for publication was Ehab Elsayed Elattar.

I. INTRODUCTION

Power Systems have largely expanded to integrate and rely on small-scale renewable energy generation to meet the escalating demand with good quality power at reasonable rates. In order to get better efficiency and economic operation, grid has to be operated near to its physical limits [1]. Voltage stability is defined as "the ability of a power system to maintain acceptable voltage in the system both under normal conditions and also after being subjected to a disturbance". Hence, it should be maintained within safe limits in a good power grid. Various methods for evaluating voltage stability are L-index, P-V curve, V/V0 index, modal analysis, line stability index (L_{mn}), fast voltage stability index (FVSI), line stability factor (LQP), voltage collapse point indicators (VCPI) etc. [2]. The stability and reliability of the power system can be improved by providing decision making support in real time.

Involvement of machine learning techniques (MLTs) in complex applications has been increased vastly in present times. Especially in power systems various MLTs such as artificial neural network (ANN), decision tree (DT) and principal component analysis (PCA) have proven their capabilities in stability and reliability assessments. ML approach is based on the training of database model from unexposed measurements for accurate predictions by its generalization capability.

In recent years, the assessment of voltage stability has been investigated using data mining and artificial intelligence. Voltage stability assessment can be divided into Long-Term Voltage Stability (LTVS) and Short-Term Voltage Stability (STVS). For online assessment of voltage stability, different MLTs were used such as Support Vector Machine (SVM) [3], [4], ANN [5] and DT [6]–[8], for LTVS, and Extreme Learning Machine (ELM) technique for STVS [9], [10]. But the validation and testing are lacking for the improvement of real-time performance. Increased computational capabilities and efficient automation tools resulted rapid growth in MLTs for analysis of advanced power systems.

The real-time scheduling of intermittent renewable energy sources (RES) to meet the highly varying load depends to a great extent on the forecast and analysis of the system. Predictive analysis forms the basis for several aspects of a power system viz., optimal operation, maintenance scheduling, generation dispatch, load shedding etc. For predictive analysis, computational intelligence has been widely used in recent years to analyze system stability. Artificial Neural Network (ANN) based online long-term voltage stability analysis was proposed by using phasor values of system voltages [11]. Comprehensive research has been carried out in the field of power system analysis with a techno-economic analysis (TEA) considering the crucial technical and economic parameters of a power system integrated with RES [12]. Here, the incorporation of STATCOM and the analysis of voltage stability and cost for IEEE standard systems were also discussed. Several other types of research are carried out in the areas of extreme learning machine, support vector machine, multiple voltage stability indices and multiple machine learning models [13].

This wide pool of research on the predictive analysis of voltage stability in terms of long-term or short-term stability margin, loadability margin etc. using computational intelligence has served as a basis for carrying out the presented research work. This paper proposes a regression-based machine learning approach to predict both voltage stability and cost of a power system with integrated renewable energy sources and reactive power compensation devices. Major contributions of the work include:

1) Predictive techno-economic analysis of renewable energy integrated system, without actual prediction of renewable energy output based on climatic conditions.

2) Prediction of voltage stability of a system and the cost of energy purchased from the grid for faster and precise analysis and control.

3) Development, validation and comparison of Linear and Polynomial regression-based ML models.

The proposed method is described in detail in the next section.

II. PROPOSED METHOD

In the proposed method, RES integrated IEEE standard 14-bus and 30-bus systems with STATCOM are considered for the prediction of voltage stability and cost based on techno-economic analysis using machine learning. Developing a Machine Learning model is done in three major stages as shown in Fig. 1.

Most crucial and sensitive part of the model development is to create a valid and accurate dataset. The generated input-output database is then used to develop an ML model by a process of training and testing. The accuracy of the

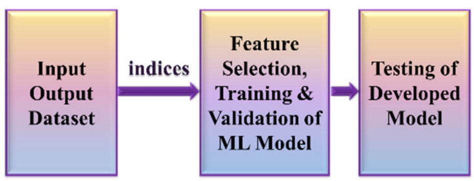

FIGURE 1. General block diagram of machine learning.

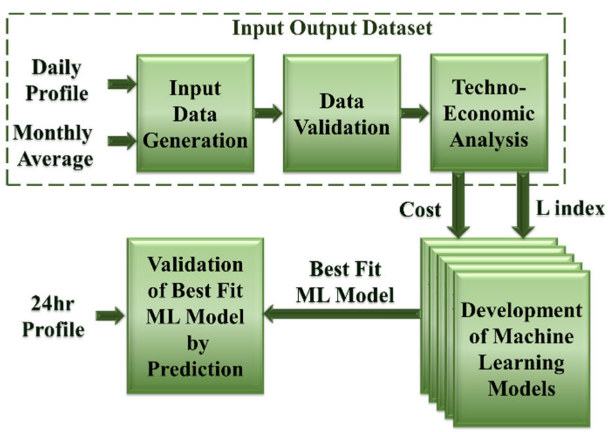

FIGURE 2. Block diagram of proposed method.

developed model and its prediction is also greatly affected by the features (input variables) chosen for training a model.

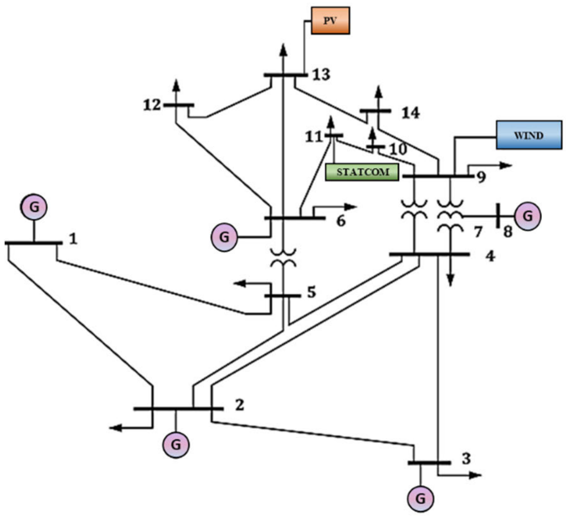

FIGURE 3. Single-line diagram of Modified IEEE 14 bus system with RES and STATCOM.

Correlation between the features and the output is analyzed to select the features. A considerable part of the updated database (based on feature selection) is used to train an ML model, while the remaining data is used to test the developed model. Training the model involves the development of an equation (the model) that relates the input and output variables. Testing, on the other hand, involves computation of the output for the test input data using the developed model. Accuracy of the model is determined by comparing the predicted output with the actual output of the test dataset. A well trained and accurately developed ML model can be used for predicting the output for any futuristic system input condition.

A predictive techno-economic analysis of modern power system is proposed. The entire framework, as presented in Fig. 2, can be broadly split into two parts viz., database generation and ML model development. Database generation involves the development of raw input data from the available profiles, followed by validation of the developed input data points. The valid input dataset for a year consisting of 8760 hourly datapoints is fed to the Techno-economic analysis codes developed in MATLAB to obtain the corresponding output (L-index and cost) datapoints. The 8760hr input-output database generated is then used to develop ML models through a series of steps which include feature selection, ML model training, testing, and validation. Four ML models namely multi-linear regression and polynomial regression models of order 3, 4 and 5 are developed, and a best-fit model is chosen for the proposed predictive analysis.

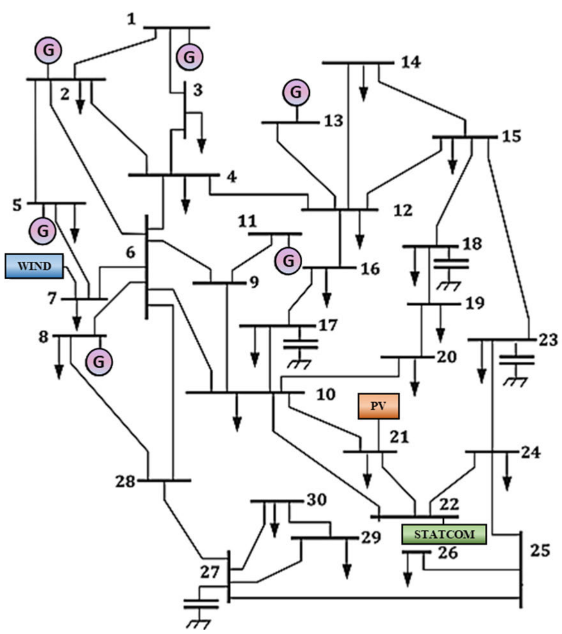

FIGURE 4. Single-line diagram of modified IEEE 30 bus system with RES and STATCOM.

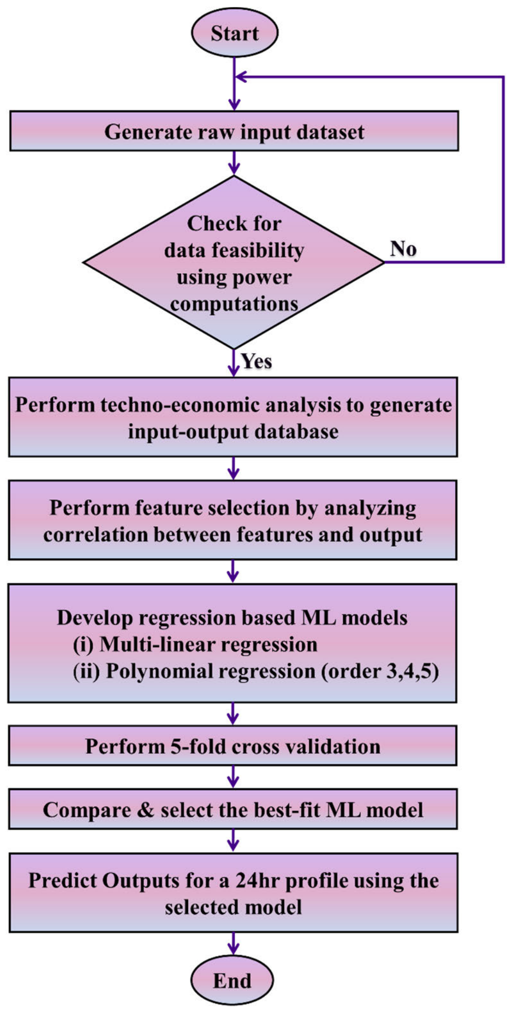

FIGURE 5. Flow diagram of proposed methodology.

The final ML model is then validated by predicting the output for a 24-hour profile. Two IEEE standard power system models (14-bus and 30-bus) are considered in this proposed work.

The IEEE-14 bus system is modified with the incorporation of Renewable Energy sources (PV and Wind) and a FACTS device (STATCOM) as shown in Fig. 3. The renewable energy sources, PV and Wind are added to Bus no 13 and 9 respectively. The location of PV and windfarm is decided considering two factors. The first factor is the L-index of the bus which indicates the voltage stability of the local bus and the second factor is the load at a particular bus. The FACTS device, STATCOM is a reactive power compensation device when added to the system improves the voltage of the system. In the IEEE 14-bus system, STATCOM is added to bus no 11 based on the voltage profile obtained from the Newton Raphson load flow analysis. Similarly, the IEEE 30-bus system is modified with the incorporation of PV, wind and STATCOM as shown in Fig.4. The renewable Energy sources, PV and wind are added to Bus no 21 and 7 respectively. The STATCOM is connected to bus no 22.

As described earlier, a predictive techno-economic analysis is carried out in this work. The entire process is presented with the help of flow diagram as represented in Fig. 5. Initially, raw input data set is generated and its feasibility is checked using power computations. Based on the feasibility input data set is modified. For generation of input-output database, TEA is performed in MATLAB. By analyzing the correlation between features and output, feature selection is performed. Here, four regression-based

TABLE 1. Monthly average input data.

Month	Avg. Load (MW)		Average Electricity Price ($/MWh)	Avg. Solar Irradiation (kWh/m^2)	Avg. Wind Velocity (m/s)
	14 Bus	30 Bus			
Jan	215.08	234.23	46.172	1.67	3.5
Feb	180.14	187.53	40.6742	3.02	3.74
March	182.83	199.71	38.4965	4.75	4.7
April	190.55	208.14	40.1149	4.88	8.3
May	259.98	283.59	54.6091	5.33	9.5
June	244.57	267.14	51.4913	5.79	7.7
July	257.03	280.76	54.1093	6	6.62
August	217.26	237.32	45.7436	6.05	5.9
Sept	251.1	274.27	52.8598	5.9	4.94
Oct	190.55	208.14	40.1149	4.88	4.82
Nov	150.46	155.11	46.3267	2.52	4.7
Dec	260	284	54.74	1.46	4.7

TABLE 2. Hourly input profile.

Time (hr)	Solar Irradiation (kW/m^2)	Wind Velocity (m/s)	Dynamic Price ($/MWh)	Load 14 Bus (MW)	Load 30 Bus (MW)
1	0	5.1	39.38	166.4	181.2096
2	0	4.3	36.97	156	169.884
3	0	3.7	35.57	150.8	164.2212
4	0	3.6	35.27	145.6	158.5584
5	0	3.3	35.28	145.6	158.5584
6	0.02	3	37.73	150.8	164.2212
7	0.9	2	42.89	166.4	181.2096
8	2.16	1	51.72	197.6	215.1864
9	3.68	2	54.54	226.2	246.3318
10	5.01	4	53.69	247	268.983
11	5.81	4	53.43	257.4	280.3086
12	4.57	4	53.57	260	283.14
13	3.38	2	53.40	257.4	280.3086
14	3.14	5	51.28	254	276.606
15	2.67	7	47.49	254	276.606
16	3.69	5	46.33	252.2	274.6458
17	1.99	3	45.89	249.6	271.8144
18	0.81	3	48.03	249.6	271.8144
19	0	3	50.22	241.8	263.3202
20	0	2	52.05	239.2	260.4888
21	0	3	52.30	239.2	260.4888
22	0	2	46.89	241.8	263.3202
23	0	3	43.09	226.2	246.3318
24	0	3	41.07	187.2	203.8608

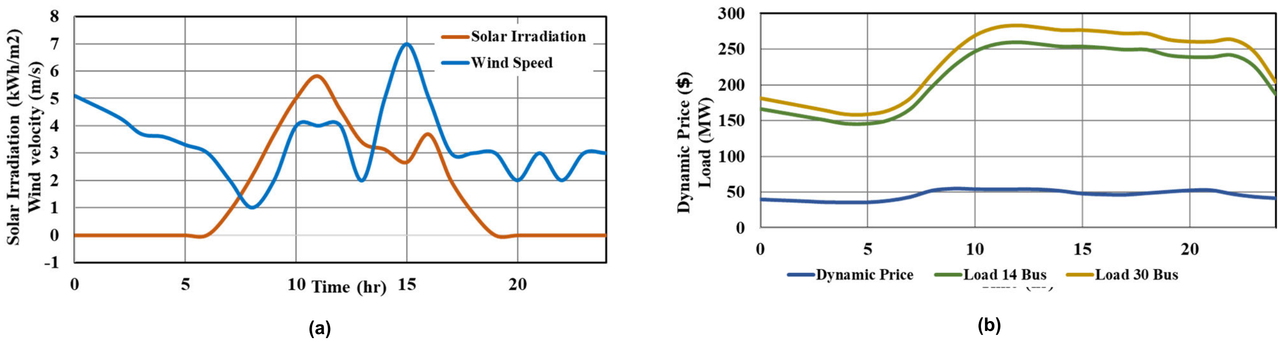

FIGURE 6. **Daily profiles – Input data, (a) Solar irradiation and wind velocity, (b) Dynamic price and load.**

ML models are developed and validated using five-fold cross validation technique. By comparison of the cross-validation (CV) results, the best-fit model is selected. This final model is validated by the prediction of voltage stability and cost for a futuristic 24-hour profile.

III. DATA-SET GENERATION FOR PREDICTIVE ANALYSIS

A comprehensive dataset for the predictive analysis is developed in three stages viz., random input data generation, checking for feasibility, and input-output dataset generation for the feasible inputs. The detailed dataset generation is described as follows.

A. INPUT DATA GENERATION

Inputs considered in the analysis are system load (for both IEEE 14-bus and 30-bus power systems with RES and STATCOM), electricity price, solar irradiation and wind velocity.

The monthly averages of each of these inputs for a year are considered as shown in Table 1. An hour-wise daily profile of load, electricity price, irradiation and wind velocity are shown in Fig. 6. The daily profile data provided in Table 2, is repeated for each day in a month and is fitted under the monthly averages. The obtained hour-wise yearly input profile is then modified by introducing normally distributed random deviations using (1). This resulted in an initial raw input profile for 8760hr which optimally covers the search space.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1}$$

B. DATA FEASIBILITY

The generated input dataset, comprising of the hourly profiles of system load, solar insolation, wind velocity and electricity prices for 8760hr, is checked for feasibility to ensure that the data generated is within the practical scope of the variable. The detailed modelling and analysis of the system, along with the procedure for feasibility tests are described.

1) MODELLING OF SOLAR PV

For modelling of solar PV system, only steady-state behavior is considered. In the integration of PV into the grid, reactive component of power is considered as zero and only the active power injection into the grid is taken into account [14] as described in (2) and (3).

$$P = max\left\{\eta P_{pv}(V)\right\} \tag{2}$$

$$P_{pv}(V) = VI = N_p I_{SC} V \left\{1 - \mathrm{m}\left[exp\left(\frac{n}{N_s}V\right) - 1\right]\right\} \tag{3}$$

IEEE 14-bus and 30-bus test systems are connected to the solar PV farm of 25MW. The solar irradiance at standard test condition is set as 1000W/m^2.

2) MODELLING OF WIND TURBINE

The active power output of the wind turbine [15] is expressed in (4).

$$P_{WG} = P_m = \frac{1}{2}\rho\left(\pi r^2\right) v_w^3 C_p(\lambda, \beta) \tag{4}$$

where

P_m – mechanical power
ρ – density of the wind (kg/m^3)
r – rotor blade swept area (m)
v_{w} – wind speed (m/s)
$\mathrm{C_p}$ – performance coefficient which is the function of tip speed ratio(λ) and pitch angle (β)

For IEEE 14-bus and 30-bus systems, the Wind farm of 50 MW is connected. It consists of 25 turbines, each with a rated power of 2MW. In all the 25 WT, the nominal wind speed of 10 m/s, the cut-in speed of 2.7 m/s and cut out speed of 25 m/s are considered.

3) MODELLING OF STATCOM

STATCOM controls the magnitude of the voltage by adjusting reactive power (Q) injected into or consumed from the system [16]. For low system voltage, the STATCOM produces reactive power (capacitive). On the other hand, high system voltage consumes reactive power (inductive). This shunt controller is modelled as follows to regulate voltage. Active and reactive powers are expressed in (5) and (6).

$$P = \frac{E.V}{X}\sin\delta \tag{5}$$

$$Q = \frac{E^2}{X} - \frac{E.V}{X}\cos\delta \tag{6}$$

where
E – voltage of the transmission line
V – voltage source converter's generated voltage
X – interconnected transformer's equivalent reactance
δ – phase angle of E with respect to V.

If $V_{sh} = V_{sh}\angle\theta_{sh}$ and $V_i = V_i\angle\theta_i$, then P and Q constraints of STATCOM are represented in (7) and (8).

$$P_{sh} = V_i^2 g_{sh} - V_i V_{sh} (g_{sh}\cos(\theta_i - \theta_{sh}) + b_{sh}\sin(\theta_i - \theta_{sh})) \tag{7}$$

$$Q_{sh} = -V_i^2 b_{sh} - V_i V_{sh}(g_{sh}\sin(\theta_i - \theta_{sh}) - b_{sh}\cos(\theta_i - \theta_{sh})) \tag{8}$$

where
$g_{sh} + jb_{sh} = \frac{1}{Z_{sh}}$
V_i – voltage at bus i,
V_{sh} – STATCOM voltage
P_{sh} – shunt converter branch active power flow
Q_{sh} – shunt converter branch reactive power flow
Z_{sh}– equivalent STATCOM shunt coupling transformer impedance.

4) LOAD FLOW ANALYSIS – NEWTON RAPHSON METHOD

The load flow analysis of the system, which includes computation of its line and node parameters in terms of voltage, current, power etc., is done using the Newton-Raphson (NR) method of power flow analysis [17], to have accurate analysis with simple control.

5) DATA FEASIBILITY TEST

The data feasibility test of the power system load is done by checking the convergence of load flow, and that of solar and wind input data by comparing the power output with the rated power of the installed farm. The pseudocodes for both the tests are shown below.

The input profiles for 8760hr are updated according to the feasibility tests conducted in MATLAB under the developed pseudocode.

```
Pseudo Code – Load Data Feasibility
FOR each hour
  PERFORM Load flow using NR method
   IF (it > it_limit) THEN
     IF (NR not converged) THEN
         MODIFY Load Data
Pseudo Code – DG Data Feasibility
FOR each hour
  COMPUTE Solar and wind power
   IF (power >= rating) THEN
     UPDATE power = efficiency * rating
          COMPUTE radiation or speed
```

C. INPUT-OUTPUT DATASET GENERATION

A machine learning-based model is developed to predict the voltage stability and cost of the system. As the ML models are developed by training and testing a base model with a huge input-output dataset, the two output parameters under consideration are computed for each hourly generated input data.

1) VOLTAGE STABILITY

Voltage stability within the safe limit should be maintained for successful power system operation [18]. One of the basic and efficient analytical techniques to inspect voltage stability is L-index [19]–[21], and is expressed in (9).

$$L_j = \left|1 - \sum_{i=1}^{g} F_{ji}\frac{V_i}{V_j}\right| \tag{9}$$

The term F_{ji} can be calculated using the admittance matrix calculation using (10).

$$\left[V_l\ I_g\right] = \left[Z_{ll}\ F_{lg}\ K_{gl}\ Y_{gg}\right]\left[I_l\ V_g\right] \tag{10}$$

where
$\left[F_{lg}\right] = -\left[Y_{ll}\right]^{-1}\left[Y_{lg}\right]$
Y_{ll} is self-admittance at the node l.
Y_{lg} is mutual admittance between node l and g.

L-index denotes the level of voltage stability. Its value can range from 0 to 1. If there is no load in a load bus, the value of the L-index is 0. An L-index value of 1 indicates a voltage collapse at the particular bus. So, the lower the L-index value, the higher will be the stability of the system.

2) COST OF ENERGY PRODUCED

The hourly cost of the system is computed for analysis. The hourly operation and maintenance costs of renewable energy sources [22] are minimal and can be neglected when compared to the price of electricity procured from the grid. The costs of electricity procured from the power grid are considered under a dynamic or time-varying pricing system. The hourly power consumption from the grid is computed and multiplied with the corresponding dynamic rates of electricity to obtain the operating costs. Aiming at the reduction of computational time, the outputs have been computed as an extension of the power flow executed during the feasibility cycle. The entire input-output dataset generated is stored in the database and are used in the ML model development.

IV. PREDICTIVE ML MODEL DEVELOPMENT

It is aimed to achieve high accuracies in prediction of the two distinct output parameters, and two separate models are developed.

A. FEATURE SELECTION

Feature selection is a key factor that greatly affects the accuracy of the developed ML model. Extensively used broad classes of feature selection are the correlation-based

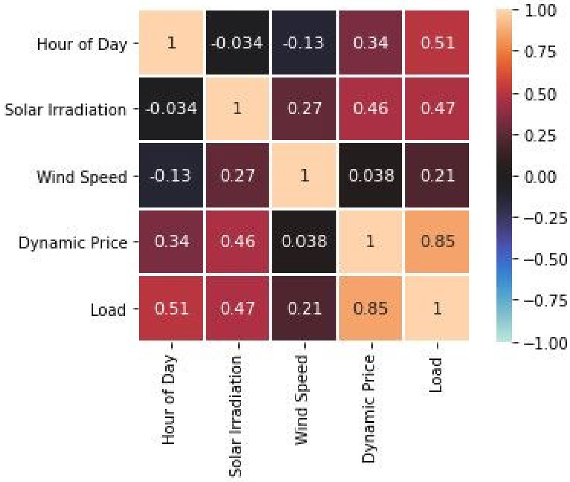

FIGURE 7. Heatmap of correlation among the input variables for 14 bus.

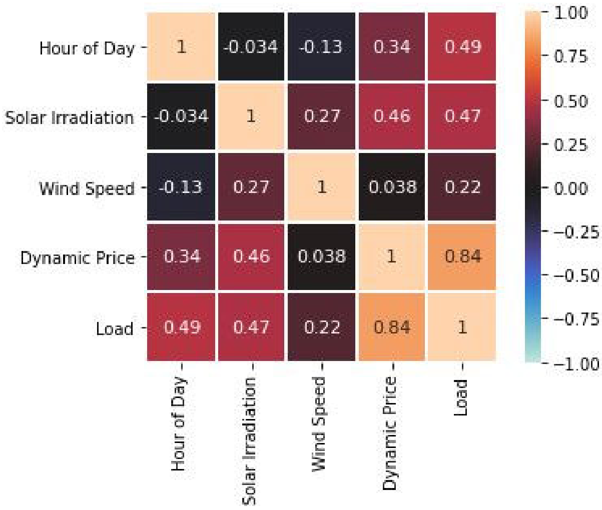

FIGURE 8. Heatmap of correlation among the input variables for 30 bus.

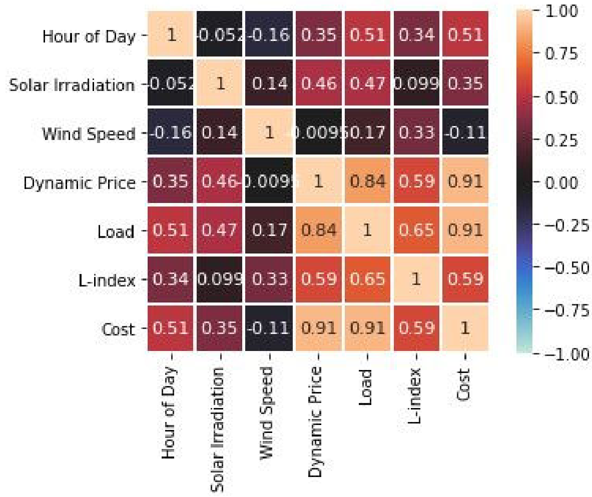

FIGURE 9. Heatmap of Spearman correlation – Voltage stability and cost for 14 bus.

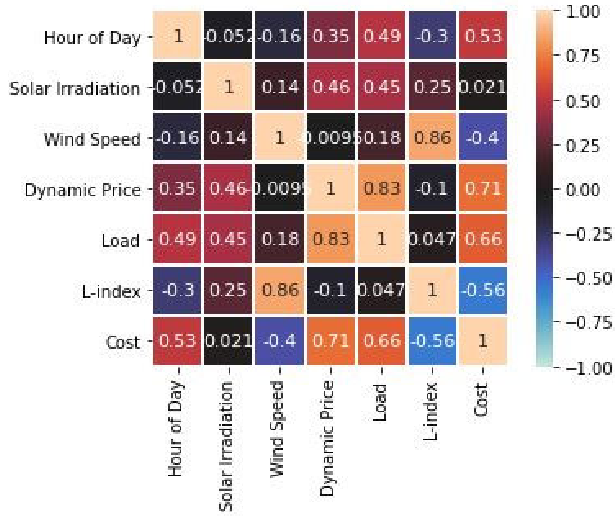

FIGURE 10. Heatmap of Spearman correlation – Voltage stability and cost for 30 bus.

and probability-based selection. In this proposed method, correlation-based feature selection is incorporated.

Correlation defines the extent of the relationship between two variables. A higher correlation between two input variables indicates that both the variables have similar characteristics and hence similar impact on output. So, one of the two variables can be considered for the analysis. Various features proposed here for the predictive analysis are hour of the day, hourly load, solar irradiation, wind velocity and dynamic pricing. The heatmap showing the correlation between the input variables is plotted using seaborn library for 14-bus and 30-bus systems database.

The obtained heatmaps are shown in Fig. 7 and Fig. 8. A value of $+/-1$ indicates perfect correlation, while zero indicates no correlation at all. From the heatmaps, it is clear that no two features have a strong correlation of 0.9 or greater. So, all the five proposed features are considered in training and testing the ML models for predictive analysis. Further, the impact of each variable on the output is also computed and analysed using Spearman's rank correlation.

Spearman's rank correlation coefficient ρ (rho), which indicates the monotonicity of the relationship between two variables, is computed for the entire input-output data set. This correlation is depicted as heatmaps in Fig. 9 and Fig. 10. A correlation coefficient of $+/-1$ indicates a monotonic relationship, while zero signifies a non-monotonic relationship. Monotonicity in a function indicates that the function is either continuously increasing or continuously decreasing with a positive first order derivative.

It is observed from the heatmap that the input and output variables are non-monotonic, with poor correlation coefficient values less than 0.9. This can be attributed to the complementary profiles of solar, wind and load profiles. It is thus, proposed to develop higher order polynomial regression models, as linear regression models are highly unlikely to fit the entire non-monotonic dataset.

B. ML MODEL DEVELOPMENT

An ML model is developed through a process of training and testing using the generated database. ML algorithms in developing a model can be broadly classified into supervised and unsupervised learning algorithms. Supervised ML algorithms are used when there is sufficient database to train and test the model. As sufficient database is available, supervised regression-based ML model is considered. For best suitability of the generated database, a higher order multi-variable polynomial regression model is developed. A linear model is also developed and compared with the proposed polynomial model for validation.

1) TRAINING

Training is the most essential stage of ML model development where both input and output data is fed to the ML algorithm to train and develop the model. The entire dataset is split into two parts in the ratio of 4:1 for training and testing, where 80% of the data is used for training the model, and 20% is used for testing.

Linear regression model and polynomial regression models of order 3, 4 and 5 are developed using the training

TABLE 3. ML model validation for voltage stability.

System	Increase in Load	R^2 values				
		Linear Regression	Polynomial Regression deg(3)	Polynomial Regression deg(4)	Polynomial Regression deg(5)	5-fold Cross Validation (Mean)
IEEE 14 bus	0 %	0.9638	0.9948	0.9976	0.9989	0.978 (poly 3) 0.946 (poly 4) -5.199 (poly 5)
	10 %	0.9679	0.9956	0.99	0.9989	0.973 (poly 3) 0.962 (poly 4) -2.438 (poly 5)
	20 %	0.971	0.9962	0.9982	0.9989	0.972 (poly 3) 0.964 (poly 4) -1.873 (poly 5)
IEEE 30 bus	0 %	0.9154	0.979	0.9932	0.9985	0.924 (poly 3) 0.845 (poly 4) -10.46 (poly 5)
	10 %	0.922	0.98	0.9861	0.998	0.942 (poly 3) 0.862 (poly 4) -7.745 (poly 5)
	20 %	0.9305	0.9815	0.9942	0.9986	0.935 (poly 3) 0.893 (poly 4) -6.376 (poly 5)

TABLE 4. ML model validation for cost of energy purchased.

System	Increase in Load	R^2 values				
		Linear Regression	Polynomial Regression deg(3)	Polynomial Regression deg(4)	Polynomial Regression deg(5)	5-fold Cross Validation (Mean)
IEEE 14 bus	0 %	0.9648	0.9958	0.9988	0.9999	0.988 (poly 3) 0.966 (poly 4) -5.181 (poly 5)
	10 %	0.9689	0.9966	0.991	0.9999	0.989 (poly 3) 0.979 (poly 4) -2.42 (poly 5)
	20 %	0.972	0.9972	0.9992	0.9999	0.99 (poly 3) 0.982 (poly 4) -1.853 (poly 5)
IEEE 30 bus	0 %	0.9164	0.98	0.9949	0.9995	0.945 (poly 3) 0.865 (poly 4) -11.56 (poly 5)
	10 %	0.923	0.981	0.9871	0.999	0.947 (poly 3) 0.886 (poly 4) -8.745 (poly 5)
	20 %	0.9315	0.9825	0.9952	0.9996	0.958 (poly 3) 0.912 (poly 4) -5.368 (poly 5)

TABLE 5. Comparison of developed models (base load case).

		R^2 values	
		IEEE 14-bus	IEEE 30-bus
Voltage Stability	Linear Reg.	0.9638	0.9154
	4^{th} deg. Polynomial Reg.	0.9976	0.9932
	5-fold CV	0.9460	0.8450
Cost	Linear Reg.	0.9648	0.9164
	4^{th} deg. Polynomial Reg.	0.9988	0.9949
	5-fold CV	0.9660	0.8650

data. Polynomial regression models are developed through a process of transformation (normalization and polynomial transformation) and prediction (using linear regression). All of these stages can be combined using pipeline function, and thereby reducing the computational burden and time.

2) TESTING

Testing is the stage of ML model development that validates or evaluates the developed model. The inputs of the test data (remaining 20%) are used to predict and compute the corresponding outputs according to the developed model. The developed models can be analysed in terms of two factors namely the root-mean-square error (RMSE) and the R^2 scores. The R^2 values indicate the fraction of accurate predictions that the developed model can make. A value of 1 indicates perfect fit. Hence, R^2 values are used to analyse the fitness of the developed model with the test data.

To obtain the maximum accuracy of the developed model, it is compared with the results obtained by performing a 5-fold cross validation. In this validation method, the entire data is split into 5-folds and different combinations of 4:1 train-test data is chosen from the folds. The combination that gives the best fit is chosen in the cross validation. The model that has the best fit in both single hold-out train-test split and the cross validation is chosen as the final model. The result and selection of the best-fit model among the developed models are discussed in the next section.

C. BEST-FIT MODEL SELECTION

All the ML models for the predictive analysis are developed using the 8760hr database generated for the study. Several training methods such as multi-variable linear regression, polynomial regression and 5-fold cross-validation are used and their corresponding R^2 scores are computed.

For the prediction of voltage stability, L-index is considered as one of the output variables to be predicted and analysed. The model developed for this analysis considered the L-index of the entire system which is the maximum of the L-indices computed for each load bus. R^2 values of the models developed for predicting the voltage stability is given in Table 3. The best model selected by comparing R^2 scores can be used to predict the system's voltage stability for any futuristic input data.

For predicting cost of energy purchased from the grid for the best-fit model is chosen by comparing the R^2 values of the developed models viz., multi-variable linear regression, polynomial regression and 5-fold cross-validation. The R^2 values of the models developed for predicting the cost is given in Table 4. The operating cost of any futuristic input data can be predicted using the selected model.

TABLE 6. Validation and testing of 4^{th} deg. polynomial Reg. model – 24 hr test data.

		R^2 values	
		IEEE 14-bus	IEEE 30-bus
Voltage Stability	Model Development	0.9976	0.9932
	Validation – Test Data	0.9932	0.9794
Cost	Model Development	0.9988	0.9949
	Validation – Test Data	0.9934	0.9651

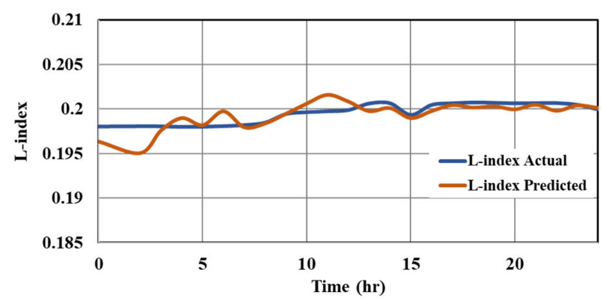

FIGURE 11. L-index prediction for 14-bus system.

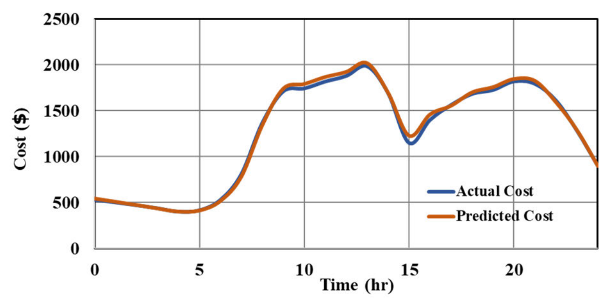

FIGURE 12. Cost prediction for 14-bus system.

Both voltage stability and cost prediction require a best-fit ML model. As the models of different loading conditions show similar R^2 values, base load case is considered to select the best fit model. It is also observed from Table 3 and Table 4 that the 4^{th} degree polynomial regression model exhibits a better fit among the other polynomial models.

The comparison of R^2 scores obtained in the linear regression, 4^{th} degree polynomial regression and 5-fold cross-validation of the 4^{th} degree polynomial regression for base load case is tabulated in Table 5. It can be observed from the scores that the polynomial regression models of degree 4 are considered most suitable, for the prediction of both cost and voltage stability in a power system.

V. RESULT ANALYSIS

The developed best-fit ML model for predictive analysis of cost and voltage stability of 14-bus and 30-bus systems is tested for a 24hr profile. The predicted values, actual values and the coefficient of determination for each of the prediction are analysed. The results are discussed in four cases formed with the combination of the two outputs and the two systems considered. The R^2 values obtained in the prediction of these four cases are tabulated in Table 6. The detailed predictive analysis of each case is discussed. The voltage stability (L-index) prediction and cost prediction for IEEE 14-bus and 30-bus systems are presented in Fig. 11 to Fig. 14 respectively.

TABLE 7. Voltage stability and cost prediction for IEEE 14-bus system.

Hr of	L - Index			Cost		
Day	Actual	Predicted	Error (%)	Actual ($)	Predicted ($)	Error (%)
1	0.198051	0.196356	0.86	532.2	545.4	2.47
2	0.198079	0.195069	1.52	471.0	474.2	0.68
3	0.19807	0.197608	0.23	438.0	437.2	0.19
4	0.198014	0.19899	0.49	401.1	402.8	0.41
5	0.198028	0.198186	0.08	415.6	414.2	0.33
6	0.198095	0.199744	0.83	528.4	517.7	2.02
7	0.198209	0.197952	0.13	808.8	783.7	3.10
8	0.198477	0.19838	0.05	1366.6	1345.3	1.56
9	0.199429	0.199469	0.02	1711.2	1743.6	1.90
10	0.199634	0.2006	0.48	1742.8	1794.6	2.97
11	0.199727	0.201586	0.93	1817.8	1870.7	2.91
12	0.199868	0.200826	0.48	1878.5	1926.2	2.54
13	0.200588	0.199758	0.41	1984.1	2021.5	1.89
14	0.200618	0.200104	0.26	1690.2	1697.6	0.44
15	0.199332	0.198983	0.18	1148.8	1229.9	7.06
16	0.200436	0.199773	0.33	1400.1	1462.9	4.48
17	0.200614	0.200429	0.09	1562.1	1557.3	0.30
18	0.200711	0.200162	0.27	1682.5	1703.0	1.22
19	0.200668	0.200303	0.18	1725.9	1759.9	1.97
20	0.200605	0.199965	0.32	1818.9	1849.0	1.66
21	0.200631	0.200477	0.08	1792.7	1828.3	1.98
22	0.200641	0.199805	0.42	1611.7	1595.9	0.98
23	0.200449	0.200401	0.02	1283.9	1292.4	0.67
24	0.199933	0.200121	0.09	899.8	900.3	0.05

TABLE 8. Voltage stability and cost prediction for IEEE 30-bus system.

Hr of	L - Index			Cost		
Day	Actual	Predicted	Error (%)	Actual ($)	Predicted ($)	Error (%)
1	0.317424	0.316247	0.37	923.4	894.7	3.11
2	0.313473	0.313666	0.06	964.9	870.9	9.74
3	0.311397	0.312388	0.32	984.1	886.3	9.94
4	0.311151	0.312238	0.35	967.9	977.7	1.02
5	0.310382	0.311707	0.43	989.0	1003.1	1.42
6	0.309679	0.309794	0.04	1062.7	1071.3	0.80
7	0.308198	0.309431	0.40	1237.2	1233.8	0.28
8	0.307299	0.307486	0.06	1549.4	1529.8	1.27
9	0.308581	0.307604	0.32	1710.3	1797.3	5.09
10	0.313302	0.312143	0.37	1607.4	1746.7	8.67
11	0.313246	0.311652	0.51	1639.3	1730.3	5.55
12	0.313421	0.312579	0.27	1693.6	1769.2	4.46
13	0.30994	0.309062	0.28	1875.1	1955.7	4.30
14	0.320795	0.320243	0.17	1474.9	1510.3	2.40
15	0.339876	0.33894	0.28	1077.9	971.3	9.89
16	0.32077	0.318072	0.84	1357.9	1345.1	0.94
17	0.311668	0.312202	0.17	1628.9	1521.1	6.62
18	0.312334	0.307559	1.53	1712.4	1692.9	1.14
19	0.312246	0.30856	1.18	1747.1	1766.6	1.11
20	0.310473	0.306125	1.40	1832.8	1965.9	7.26
21	0.312286	0.309042	1.04	1784.0	1924.6	7.88
22	0.31042	0.307272	1.01	1713.9	1746.6	1.91
23	0.312489	0.309585	0.93	1496.9	1531.0	2.28
24	0.313067	0.309435	1.16	1394.7	1483.1	6.34

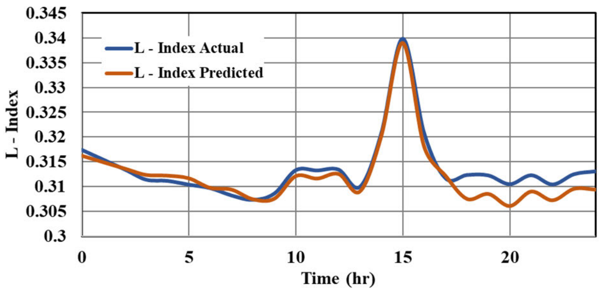

FIGURE 13. L-index prediction for 30-bus system.

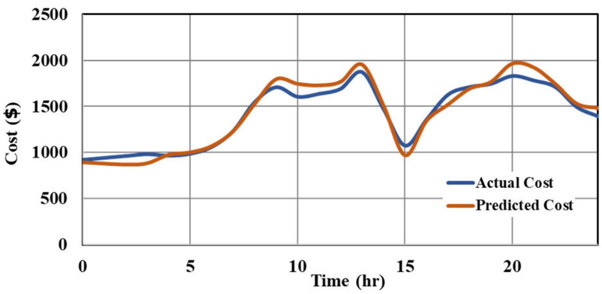

FIGURE 14. Cost prediction for 30-bus system.

A. IEEE 14-BUS VOLTAGE STABILITY PREDICTION

The voltage stability of the 14-bus system predicted by the developed ML model is shown in Table 7 along with the actual values. It is observed that the maximum percentage error in the prediction is as low as 1%. The graph depicting the deviations between the expected and predicted values is presented in Fig. 11.

On further analysing the prediction error percentage, it can be observed that a very few values are closer to 1, while the others are closer to zero. This validates the higher value (0.9932) of R^2 obtained for this prediction data.

B. IEEE 30-BUS VOLTAGE STABILITY PREDICTION

Voltage stability of 30-bus system predicted by the developed ML model is shown in Table 8 along with the actual values. It is evident that the maximum percentage error in the prediction is as low as 1.5%. The graph depicting the deviations between the expected and predicted values is presented in Fig. 13. Further analysis of the prediction error percentage shows that the number of values close to 1 is a little higher compared to the results of the 14-bus system. The same is also reflected in the R^2 value of 0.9794 given in Table 6.

C. IEEE 14-BUS COST PREDICTION

Cost of energy purchased from the grid is predicted for the 24hr profile of 14-bus system using the developed ML model. A graph depicting the deviations between the actual and predicted values is presented in Fig. 12. The predicted cost is furnished in Table 7 along with the actual values. It is observed that the maximum percentage error in the prediction is around 7%. However, most of the values are less than 5%, and an error of greater than 5% in prediction is observed for very few hours. This validates the higher value (0.9934) of R^2 obtained for this prediction data.

TABLE 9. Computational time of IEEE 14-bus and 30-bus systems.

Method of Analysis	Execution Time (s) for IEEE 14-bus System	Execution Time (s) for IEEE 30-bus System
Conventional Load Flow Analysis (Newton-Raphson Method)	0.876339	4.44219
Proposed ML based Predictive Analysis	0.0160	0.0216

D. IEEE 30-BUS COST PREDICTION

Cost of energy purchased from the grid predicted for 30-bus system, using the developed ML model is presented in Table 8 along with the actual values. It is evident that the maximum percentage error in the prediction is less than 10%. The graph depicting the deviations between the actual and predicted values is presented in Fig. 14. On further analysing the prediction error percentage, it can be observed that a larger portion of the prediction error is greater than 5%. The same is also reflected in the R^2 value of 0.9651 shown in Table 6.

E. COMPUTATIONAL TIME COMPARISON

A comparison table of computational times for both IEEE 14-bus and 30-bus systems with conventional and proposed methods for analysis (24-hour validation data) is provided in Table 9.

Conventional load flow analysis is carried out using Newton-Raphson Method in MATLAB and the proposed ML based predictive analysis is carried out in JupyterLab. It can be observed that ML based algorithms have greater computational efficiency over traditional computational approaches. Thus, it is evident that the proposed ML based predictive analysis is much faster and can be used for real-time analysis and control.

VI. CONCLUSION

Predictive analysis of a power system in terms of voltage stability and cost has been carried out. Two modified power systems, IEEE 14-bus and IEEE 30-bus systems, with integrated renewable generation and STATCOM devices, were considered for the analytical study. Analysing the heatmaps of correlations suggested that no two features have a strong correlation, so all the proposed features were suggested for developing the predictive ML models. It was also observed that the relationship between inputs and outputs is non-monotonic, suggesting that a polynomial model is best suited for the database. ML models were developed using several methods including multi-variable linear regression, polynomial regression (of 3rd, 4th and 5th degrees) and 5-fold cross-validation. These models were analyzed in terms of their R^2 scores, which suggested that 4th order polynomial regression model with R^2 values greater than 9 is optimally suited for all the systems in the prediction of both voltage stability in terms of systems' L-index and cost in terms of energy purchased from the grid.

The validation tests for the predictive voltage stability and cost analysis were performed for both IEEE 14-bus and 30-bus systems for a 24hr test data. R^2 values greater than 0.9 suggest the effectiveness of the developed ML models. This effectiveness was further validated by performing a detailed comparative study of each of the four cases of predictive analysis. The comparison of predicted and actual values of all the cases were furnished and analyzed in terms of deviation and percentage error.

The validation tests conducted for predicting the voltage stability and cost of the considered systems for a 24hr test data confirm the effectiveness of the developed ML models.

The proposed method can be extended to predict voltage stability in case of other disturbances or sequence of events (short circuits, tripping of network elements, etc.) by including additional features like bus voltage, line current, line voltage drop etc. to the model development. This can be done by merely adding the values of bus voltage, line current etc. as additional inputs (dimensions) to each datapoint while generating the input-output dataset using MATLAB.

The proposed ML model can also be applied for any larger power network with the availability of dataset. As both MATLAB and Machine Learning codes are generalized, i.e., they can interchangeably be used for any system by replacing the input data in the developed MATLAB codes for dataset generation and replacing the modified dataset to train the ML models.

REFERENCES

[1] A. Bose, "Power system stability: New opportunities for control," in *Stability and Control of Dynamical Systems with Applications*. New York, NY, USA: Springer, 2003, ch. 16, pp. 315–330.

[2] J. Modarresi, E. Gholipour, and A. Khodabakhshian, "A comprehensive review of the voltage stability indices," *Renew. Sustain. Energy Rev.*, vol. 63, pp. 1–12, Sep. 2016.

[3] M. V. Suganyadevi and C. K. Babulal, "Support vector regression model for the prediction of loadability margin of a power system," *Appl. Soft Comput.*, vol. 24, pp. 304–315, Nov. 2014.

[4] K. S. Sajan, V. Kumar, and B. Tyagi, "Genetic algorithm based support vector machine for on-line voltage stability monitoring," *Int. J. Elect. Power Energy Syst.*, vol. 73, pp. 200–208, Dec. 2015.

[5] D. Q. Zhou, U. D. Annakkage, and A. D. Rajapakse, "Online monitoring of voltage stability margin using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1566–1574, Aug. 2010.

[6] H. Mohammadi and M. Dehghani, "PMU based voltage security assessment of power systems exploiting principal component analysis and decision trees," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 655–663, Jan. 2015.

[7] R. Diao, K. Sun, V. Vittal, R. J. O'Keefe, M. R. Richardson, N. Bhatt, D. Stradford, and S. K. Sarawgi, "Decision tree-based online voltage security assessment using PMU measurements," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 832–839, May 2009.

[8] H. Khoshkhoo and S. M. Shahrtash, "On-line dynamic voltage instability prediction based on decision tree supported by a wide-area measurement system," *IET Gener., Transmiss. Distrib.*, vol. 6, no. 11, pp. 1143–1152, Nov. 2012.

[9] C. Ren, Y. Xu, Y. Zhang, and R. Zhang, "A hybrid randomized learning system for temporal-adaptive voltage stability assessment of power systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3672–3684, Jun. 2020.

[10] Y. Zhang, Y. Xu, Z. Y. Dong, and R. Zhang, "A hierarchical self-adaptive data-analytics method for real-time power system short-term voltage stability assessment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 74–84, Jan. 2019.

[11] K. D. Dharmapala, A. Rajapakse, K. Narendra, and Y. Zhang, "Machine learning based real-time monitoring of long-term voltage stability using voltage stability indices," *IEEE Access*, vol. 8, pp. 222544–222555, 2020.

[12] H. J. Queen, J. Jayakumar, and T. J. Deepika, "Comparative techno-economic analysis of power system with and without renewable energy sources and STATCOM," *J. Green Eng.*, vol. 11, no. 2, pp. 1648–1667, Feb. 2021.

[13] O. A. Alimi, K. Ouahada, and A. M. Abu-Mahfouz, "A review of machine learning approaches to power system security and stability," *IEEE Access*, vol. 8, pp. 113512–113531, 2020.

[14] Q. Alsafasfeh, O. Saraereh, I. Khan, and S. Kim, "Solar PV grid power flow analysis," *Sustainability*, vol. 11, no. 6, p. 1744, Mar. 2019.

[15] V. Seshadri Sravan Kumar and D. Thukaram, "Accurate modeling of doubly fed induction generator based wind farms in load flow analysis," *Electr. Power Syst. Res.*, vol. 155, pp. 363–371, Feb. 2018.

[16] B. U. Musa and M. Mustapha, "Modelling and simulation of STATCOM for reactive power and voltage control," *J. Multidisciplinary Eng. Sci. Technol.*, vol. 2, no. 2, pp. 1–5, Feb. 2015.

[17] S. Chatterjee and S. Mandal, "A novel comparison of gauss-seidel and Newton-Raphson methods for load flow analysis," in *Proc. Int. Conf. Power Embedded Drive Control (ICPEDC)*, Mar. 2017, pp. 1–7.

[18] O. P. Rahi, A. K. Yadav, H. Malik, A. Azeem, and B. Kr, "Power system voltage stability assessment through artificial neural network," *Procedia Eng.*, vol. 30, pp. 53–60, Jan. 2012.

[19] B. Gao, G. K. Morison, and P. Kundur, "Voltage stability evaluation using modal analysis," *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1529–1542, Nov. 1992.

[20] C. Reis and F. P. M. Barbosa, "A comparison of voltage stability indices," in *Proc. MELECON*, May 2006, pp. 1007–1010.

[21] P. Kessel and H. Glavitsch, "Estimating the voltage stability of a power system," *IEEE Trans. Power Del.*, vol. PD-1, no. 3, pp. 346–354, Jul. 1986.

[22] C. Li, F. de Bosio, F. Chen, S. K. Chaudhary, J. C. Vasquez, and J. M. Guerrero, "Economic dispatch for operating cost minimization under real-time pricing in droop-controlled DC microgrid," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 5, no. 1, pp. 587–595, Mar. 2017.

HEPHZIBAH JOSE QUEEN received the B.E. degree in electrical and electronics engineering from Karunya University, Coimbatore, India, in 2011, and the M.E. degree in power systems engineering from SRM Valliammai Engineering College, Anna University, Chennai, in 2013. She is currently a Research Scholar with the Department of Electrical and Electronics Engineering, Karunya University. She has published many research papers in international journals and conferences. Her research interests include smart grid, power system analysis, machine learning, and improvement of power quality with FACTS devices.

JAYAKUMAR J. received the B.E. degree from Bharathiar University, Coimbatore, India, in 1999, the M.E. degree from Madurai Kamaraj University, Madurai, in 2002, and the Ph.D. degree from Anna University, Chennai, in 2010. He is currently working as an Associate Professor with the Department of Electrical and Electronics Engineering, Karunya University, Coimbatore. He has published many research papers in international journals and conferences. His research interests include multimedia systems, cloud computing, smart grid, power system analysis, generation and distribution, and control and modeling of the power system with flexible ac transmission systems (FACTS).

DEEPIKA T. J. received the B.E. degree in electrical and electronics engineering from Karunya University, Coimbatore, in 2011, the M.Tech. degree in power systems from SRM University, Chennai, in 2014, and the Ph.D. degree in electrical engineering from Anna University, Chennai, in 2019. She is currently working as an Assistant Professor with the Department of Electrical and Electronics Engineering, Sona College of Technology, Salem, India. Her research interests include power systems, solar photovoltaics, distribution systems, hybrid renewable distributed generation, smart grid, deregulated power market, soft computing, machine learning, and data analytics using python.

K. VICTOR SAM MOSES BABU (Member, IEEE) received the B.E. degree from Karunya University, Coimbatore, India, in 2011, and the M.E. degree in power electronics and drives from the Sona College of Technology, Salem, in 2014. He is currently pursuing the Ph.D. degree with the BITS Pilani–Hyderabad Campus, India. He has three years of teaching experience as an Assistant Professor and three years of research experience as a JRF/SRF in international collaborative research project between India and Sri Lanka funded by DST, Government of India. His research interests include hybrid renewable energy-based standalone and grid-tied systems, energy management systems, artificial intelligence-based controllers, multilevel inverters, special electrical machines, power electronic converters, and electric drives.

SURYA PRAKASH THOTA (Senior Member, IEEE) received the B.Tech. degree in electrical and electronics engineering from Nagarjuna University, Andhra Pradesh, India, in 2001, and the M.E. degree in electrical engineering with specialization in industrial drives and control and the Ph.D. degree in electrical engineering from Osmania University, Hyderabad, Telangana, India, in 2004 and 2021, respectively. He is currently working as an Assistant Professor with the Department of EEE, G. Narayanamma Institute of Technology and Science for Women, Hyderabad. He has more than 17 years of teaching and research experience. His research interests include power quality, renewable energy systems, smart grids, and electric vehicles. He is a Life Member of the Institute of Engineers India (IEI) and the Indian Society for Technical Education (ISTE).