

# Estimating the Distributed Generation Unit Sizing and Its Effects on the Distribution System by Using Machine Learning Methods

Mikail Purlu\*, Belgin Emre Turkay

*Department of Electrical Engineering, Faculty of Electrical and Electronics Engineering,  
Istanbul Technical University,  
34469 Maslak, Istanbul, Turkey  
purlu@itu.edu.tr*

**Abstract**—Many approaches about the planning and operation of power systems, such as network reconfiguration and distributed generation (DG), have been proposed to overcome the challenges caused by the increase in electricity consumption. Besides the positive effects on the grid, contributions on environmental pollution and other advantages, the rapid developments in renewable energy technologies have made the DG resources an important issue, however, improper DG allocation may result in network damages. A lot of studies have been practised with analytical and heuristic methods based on load flow for optimal DG integration to the network. This novel method based on estimation is proposed to determine the size of DG and its effects on the network to get rid of the coercive and time-consuming load flow techniques. Machine learning algorithms, such as Linear Regression, Artificial Neural Network, Support Vector Regression, K-Nearest Neighbor, and Decision Tree, have been used for the estimations and have been applied to well-known test systems, such as IEEE 12-bus, 33-bus, and 69-bus distribution systems. The accuracy of the proposed estimation methods has been verified with R-squared and mean absolute percentage error. Results show that the proposed DG allocation method is effective, applicable, and flexible.

**Index Terms**—Distributed generation; Distribution system planning; DG sitting; Dispersed generation; Forecasting; Machine learning; Weka.

## I. INTRODUCTION

Since installing new central power plants and transmission lines to meet the increase in electricity consumption requires high cost, it is recommended to integrate smaller production units close to consumption areas. The small powerful generating units are known as distributed generation (DG). Interest in DG has increased thanks to various benefits, such as reducing system losses, improving the voltage profile, reducing pollutant emissions, and increasing system reliability. Incorrect DG allocation to the distribution system (DS) does not benefit, but on the contrary, it hurts the power system. To ensure effective DG allocation, various optimization studies have been reported in the literature. Since a predictive method is used in this

study, literature studies on estimations related to power system also follow.

In [1], DG placement and network reconfiguration by using heuristic algorithms, such as genetic algorithm (GA), particle swarm optimization (PSO), differential evolution (DE), and artificial bee colony (ABC), have been carried out to decrease losses and increase system efficiency. In [2], to minimize the annual energy losses, a technique based on mixed integer nonlinear programming (MILP) for various renewable DG placements has been developed and applied on rural DS. Size of DG units, feeder capacity, voltage limits, and penetration limits have been chosen as the constraints of the problem. In [3], it has been aimed to integrate DG with minimum power as possible and minimize active power losses by using various PSO and DE algorithms for IEEE 33-bus and 69-bus radial DSs. In [4], DGs have been placed in the weakest buses to reduce system losses and improve voltage magnitude and stability. Whereas sensitivity indices and quadratic curve fitting technique have been used for single DG allocation, loss enhancement index and power loss reduction index have been used for multiple DGs. In [5], whale optimization algorithm has been used to determine the optimal size and location of four different renewable DG types in four distribution systems, such as IEEE 15-bus, 33-bus, 69-bus, and 85-bus test systems. Thus, the authors have ensured reducing the system losses, enhancing the voltage magnitude, and increasing reliability. In addition, the results obtained have been compared with other studies in the literature. In [6], a bacterial foraging optimization algorithm has been used to find the optimum DG size. Whereas minimizing network losses and operating costs and increasing the voltage stability have been determined as the objective function, the current carrying capacities of the lines have been as a constraint. In [7], three analytical approaches have been proposed to determine the best location, size, and power factor of DG unit considering energy loss minimization. In [8], an analytical method has been used to find the sizing and sitting of DGs in balanced radial DS for ensuring minimum network loss. In [9], the optimal location and size of solar photovoltaic DG have been found by using biogeography-based optimization to

minimize losses considering the voltage profile and voltage harmonic distortion limits. In [10], an analytical method for the allocation of DGs in radial DS has been introduced to reduce active and reactive power losses. In [11], a hybrid method based on tabu search and GA has been proposed to integrate DG and capacitor banks to DS for enhancing system performance. In [12], a modified jaya algorithm-based multi-objective function for allocation of DG with high penetration has been produced to reduce active losses and improve the voltage profile and has been carried out on standard IEEE 33-bus DS. In [13], an optimization technique based on hybrid shuffled frog leap teaching and learning algorithm has been proposed and applied for optimum location and size of electric vehicle fast charging stations and DG units to DS. In [14], fault current limiters (FCLs) in series with DGs have been used to minimize the negative effects of DG allocation to reduce power system losses. Nondominated sorting GAs have been used to determine DG location, DG size, and FCLs size and performed on two test systems. In [15], an optimization technique based on PSO for determining the size of a solar photovoltaic system (SPV) has been proposed to minimize the cost of SPV integrated to grid. In [16], DG placement optimization problem has been solved in the practical test system of Korea using the optimal locator index to determine DG location and using Kalman Filter to determine DG size. In [17], the ant colony system algorithm has been used to increase system reliability by allocating DG and reclosers. In [18], PSO has been used to allocate DG for increasing the loadability of distribution systems. In [19], to minimize network losses and maximize voltage magnitude, the optimal DG placement has been performed by using the cuckoo search algorithm. Harmony search algorithm [20], an analytical approach [21], ordinal optimization [22], and MINLP [23] have been used to allocate DG in DS.

In [24], a multiple linear regression (LR)-based methodology has been proposed for long-term load estimation using hourly resolution load data of 57 countries. In [25], by using meteorological data of Muzaffarabad city in Pakistan, electricity consumption has been estimated with the help of traditional machine learning (ML) algorithms. In [26], to estimate the electricity load of Jeju island, a GA-optimized approach consisting of ML algorithms, such as support vector regression (SVR), k-nearest neighbor (KNN), and XGBoost, has been proposed. In [27], the authors have aimed to increase the performance of intraday load forecasts by grouping according to customer behavior similarities and using smart meter data. In [28], in addition to short-term load and generation estimates for the Slovenia power system, active power losses have been estimated by using fuzzy logic decision. In [29], the power consumptions of the state of Maine, the region of New England, Singapore, and New South Wales of Australia have been estimated in the short term by using the method based on the second decision mechanism and cross multimodel. In [30], KNN has been used for short-term load forecasting (STLF) on Smart City Demo Aspern buildings. In [31], a locally weighed SVR has been used to forecast two real world electric loads. In [32], the fuzzy LR method has been used to predict electricity loads of holidays in the short term. In [33], multivariate LR

and feed-forward neural network (FFNN) have been suggested for load prediction of Delhi, India. In [34], a knowledge-based expert system has been implemented for annual load forecasting. In [35], recurrent artificial NN has been used to forecast mid-term daily peak load. In [36], a hybrid model for load forecasting has been created using data preprocessing technology, individual forecast algorithm, and weight determination theory. A method based on wavelet decomposition and quadratic gray NN combined with the enhanced Dickey-Fuller test [37], dynamic model selection based on Q-learning [38], boosting based multiple kernel learning method [39], a hybrid method consisting of convolutional NN and long short-term memory based deep learning (DL) [40], a DL method [41], and a hybrid model consisting of clustering and FFNN [42] have been proposed for STLF.

In [43], regression-based analysis has been used to estimate the bus voltages of IEEE 12-bus DS. In [44], wind speed estimation has been made for Batman province in Turkey by using ANN.

Works in [1]–[23] are studies on DG allocation, and the works in [24]–[44] are on predictions related to the power system. As can be seen from the literature review, the load is mostly accepted as constant and DG output is controllable when allocating DG. In practice, the loads and DG output constantly vary. Calculating losses and other factors using power flow-based algorithms is difficult and time-consuming.

The contribution of this study is the estimation of DG size, network active losses, reactive losses, and minimum bus voltages without of power flow calculations. Machine learning algorithms, such as Linear Regression (LR), Artificial Neural Network (ANN), Support Vector Regression (SVR), K-Nearest Neighbor (KNN), and Decision Tree (DT), are used for five estimation cases. The cases are implied to IEEE 12-bus, 33-bus, and 69-bus standard test systems and the obtained results are compared.

Remaining of the paper is organized as follows. The methods used for estimation are summarized in Section II. Estimation error and performance evaluation methods are explained in Section III. In Section IV, brief information is given about the data collection and the programs used in this study. Section V presents the results of five cases created for estimations. Results are discussed in Section VI. Finally, the conclusions are located in Section VII.

## II. METHODS USED FOR THE PROPOSED APPROACH

Machine learning is the computer modelling of systems that make predictions by making inferences from data with mathematical and statistical operations. It has become more popular as computers have become more powerful in recent years. In this paper, machine learning algorithms are proposed for the estimation process and they are as follows.

### A. Linear Regression

LR is one of the prediction methods that can determine the relationship between two or more variables that have a cause-effect relationship and make predictions from the future unknown about that subject by using this relationship.

In this method, a mathematical model is used to explain the relationship between two or more variables and this

model is called “linear regression model”. There are two types of regression models. These are simple LR with one independent variable and multivariate LR analysis with more than one independent variable. The linear equation can be written as

$$y = w_o + \sum w_i \times x_i + \varepsilon, \quad (1)$$

where  $y$  is the dependent variable (estimated value),  $w_i$  is the coefficient of weight,  $x_i$  is the independent variable, and  $\varepsilon$  is the error value. In this study, the independent variables are factors, such as load change, DG size, and location, and the dependent variable is system losses affected by these factors.

### B. Artificial Neural Network

ANN, first modelled in 1943 by neurophysiologist Warren McCulloch and mathematician Walter Pitts, is one of the artificial intelligence algorithms and has a wide range of uses [45]. Since it is modelled by imitating the human brain, ANN has the features of generating, forming, and interpreting new information, as well as learning information. It is a good classifier algorithm that gives successful results in the analysis of repeatedly measured data sets. ANN has varieties, such as feed-forward, back propagation, single layer, and multilayer. Figure 1 shows a multilayer network structure with three inputs.

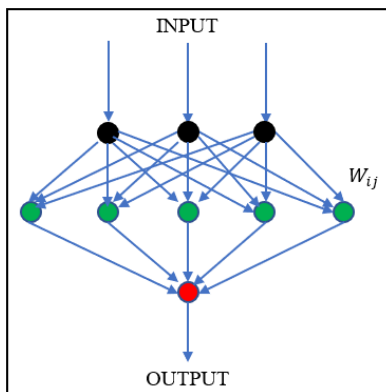


Fig. 1. Structure of a three-layered perceptron type ANN.

Although ANN has many advantages, there are also some disadvantages of them. These drawbacks can be given as follows [46]:

- It is not possible to know what is in the system;
- Stability analysis is problematic, except for some networks;
- It can be difficult to apply to different systems.

### C. Support Vector Regression

SVR was modelled by Drucker, Burges, Kaufman, Smola, and Vapnik using support vector machines [47]. It is a kernel-based machine learning algorithm used for classification and regression. This method has much better performance and ability to solve nonlinear problems compared to other traditional learning methods.

Examples of linear and nonlinear SVR are shown in Fig. 2 and Fig. 3, respectively [48]. The error tolerance is zero for the data inside the area defined as  $\varepsilon$ . Notably, the

variable  $\zeta$  represents the fault tolerance outside the sensitive area and is also referred to as training error in the literature. In the region within the sensitive area, the value of  $\zeta$  is zero. The fault tolerance can be expressed as in (2)

$$L_\varepsilon(x, y) = \begin{cases} |x - y| - \varepsilon, & |x - y| > \varepsilon, \\ 0, & \text{others.} \end{cases} \quad (2)$$

If it is within the defined range, the error is equal to zero. The term  $C$  is a constant that provides the balance between the experimental error expressed as the cost and the weight vector. It is important to determine these values for high accuracy support vector regression modelling.

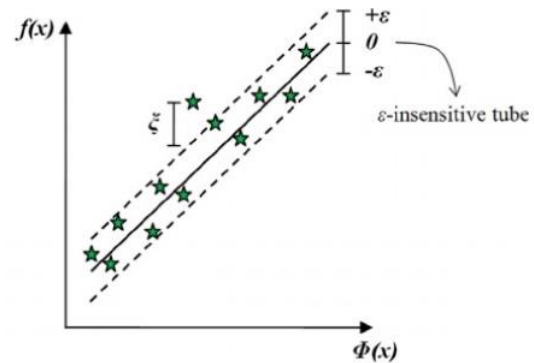


Fig. 2. Example of Linear Support Vector Regression [48].

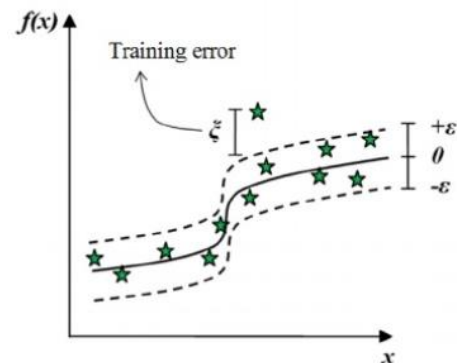


Fig. 3. Example of Non-linear Support Vector Regression [48].

### D. K-Nearest Neighbor

KNN, first introduced by Fix and Hodges in 1951, is the learning algorithm that finds the closest neighbor among the variables [49]. KNN is also known as the lazy student in the literature and it is often preferred because of its high performance in very wide areas of use. The most important point here is the distance between data points and the  $k$  value.  $K$  is an important parameter used to determine distances and is chosen as an odd number. Manhattan, Minkowski, Mahalanobis and Euclidean distance measures are generally used to calculate the distance between data [50].

Figure 4 shows a schematic diagram of KNN. When  $k$  is taken as 3, the sample classification result is prismatic. When  $k = 7$  and  $k = 11$ , the results are triangle and prismatic, respectively. It is clear from Fig. 4 that the classification results change significantly when  $k$  takes different values. Depending on the distance calculation method used, different nearest neighbors can be found, which affects the classification results [51].

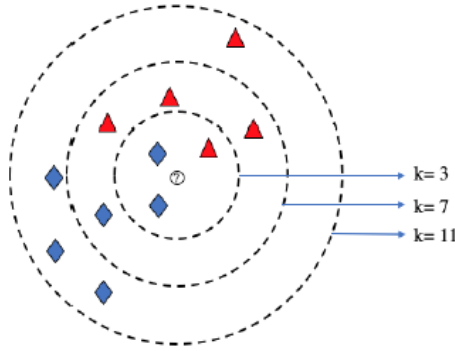


Fig. 4. K-Nearest Neighbor diagram [51].

### E. Decision Tree

DT is a classification algorithm in the form of a tree structure consisting of leaves, branches, and root nodes. DT works similarly to other ML algorithms. The algorithm generates decision trees using information gain and entropy.

A decision tree classification example can be seen in Fig. 5. In this example, the decision to go out is classified as yes or no, depending on the weather. Outlook is defined as the root node. Weather probabilities are primarily classified as sunny, overcast, and rainy, and these are branch node values. The second branch node values are the variables of humidity and wind. In cases the weather is sunny and the humidity is high, or when it is rainy and the wind is strong, the decision to go out is determined as no, and in other cases, it is classified as yes.

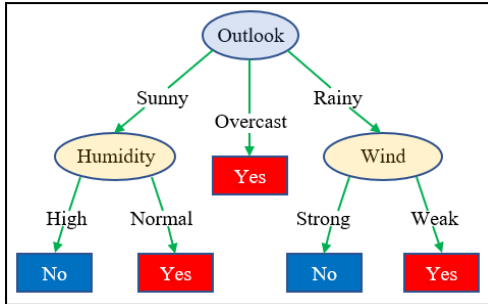


Fig. 5. An example of Decision Tree classification.

## III. ESTIMATION ERROR AND PERFORMANCE EVALUATION METHODS

### A. R-Squared

Actual values are used to test the accuracy of the estimation models. An R-squared error-based analysis is used to predict the performance and reliability of the model [24]. The R-squared error is calculated as follows

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}, \quad (3)$$

where  $y_i$  is the actual value,  $\hat{y}_i$  is the predicted value,  $\bar{y}_i$  is the mean of the actual values, and  $N$  is the number of samples. R-squared takes values between zero and one ( $0 \leq R^2 \leq 1$ ).

### B. Mean Absolute Percentage Error (MAPE)

The ratio between absolute prediction errors and the real values is defined as MAPE. It is used to measure the performance of predictions and evaluate the results [27].

MAPE is calculated as follows

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \hat{y}_i)}{y_i}. \quad (4)$$

## IV. DATA COLLECTION AND USED PROGRAMS

In this study, a data consisting of normalized load variation, DG power injections, DG locations, active power losses, reactive power losses, and minimum busbar voltages have been used for the three test systems, such as IEEE 12-bus, 33-bus, and 69-bus distribution systems. The normalized load variation data changing according to season and time of day is given in [2] and it is shown graphically in Fig. 6. These data are provided on the Github repository [52].

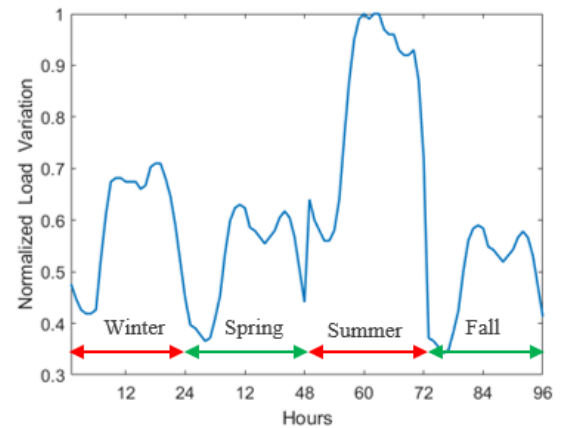


Fig. 6. The change of normalized load variation according to season and time of day.

Developed prediction models in this study have been created by WEKA, which consist of the initial of Waikato Environment for Knowledge Analysis. It is a free data mining software that includes many machine learning algorithms. While making predictions, 75 % of the data have been used for training and 25 % - for testing. The algorithms selected in WEKA for ML methods are given in Table I.

TABLE I. WEKA ALGORITHMS FOR ML METHODS.

ML Methods	WEKA Algorithms
LR	LinearRegression
ANN	MultilayerPerceptron
SVR	SMOreg
KNN	IBk
DT	M5P

Predictions have been made by optimizing the necessary parameters of ML methods. Related graphics and error performance analysis have been obtained as a result of processing the outputs of WEKA through the MATLAB software.

## V. RESULTS OF PREDICTION MODELS

### A. Case 1

In this case, DG sizing is estimated using normalized load variation (NLV). The five estimation methods mentioned in Section II are applied for 33-bus and 69-bus DSs.

The regression model equations of LR for DG sizing estimation are given in Table II. Estimation results and

actual values (AV) are shown in Fig. 7 for 33-bus and in Fig. 8 for 69-bus DSs. Calculated error and performance evaluations are compared in Table III.

TABLE II. REGRESSION MODEL EQUATIONS FOR DG SIZING ESTIMATION.

Distribution systems	Regression model equations
33-bus	$P_{dg} = 2566.1833 \times LoadLevel - 53.5766$
69-bus	$P_{dg} = 1827.2793 \times LoadLevel - 4.711$

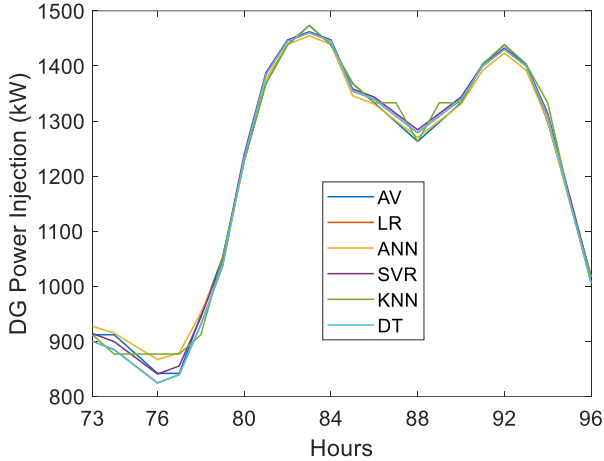


Fig. 7. Comparison of DG sizing predictions and actual values for 33-bus distribution system.

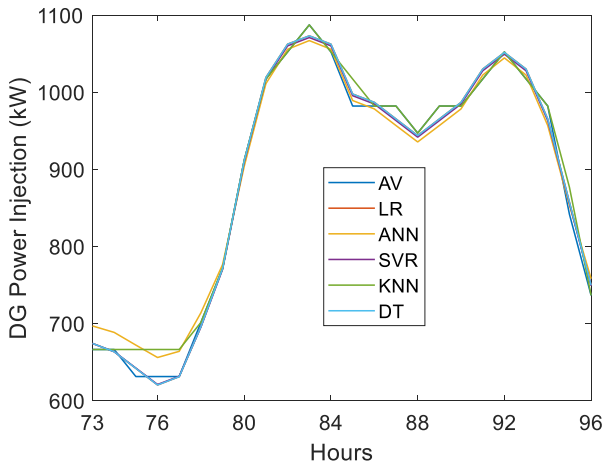


Fig. 8. Comparison of DG sizing predictions and actual values for 69-bus distribution system.

TABLE III. COMPARISON OF EVALUATIONS FOR DG SIZING ESTIMATION.

Distribution systems	Methods	R <sup>2</sup>	MAPE [%]
33-bus	LR	0.9965	0.9923
	ANN	0.9963	0.8630
	<b>SVR</b>	<b>0.9974</b>	<b>0.7560</b>
	KNN	0.9922	0.9997
	DT	0.9965	0.9923
	69-bus	LR	0.9956
ANN		0.9861	1.9299
<b>SVR</b>		<b>0.9956</b>	<b>0.9695</b>
KNN		0.9897	1.0168
DT		0.9956	0.9856

B. Case 2

In this case, the active power losses in 33-bus and 69-bus DSs are estimated using NLV through all mentioned methods.

The regression model equations of LR are given in Table IV. Relevant forecast values and actual values are shown in Fig. 9 and Fig. 10 for 33-bus and 69-bus DSs, respectively.

Comparison of calculated evaluation values are given in Table V.

TABLE IV. REGRESSION MODEL EQUATIONS FOR ESTIMATING ACTIVE POWER LOSS USING THE NLV.

Distribution systems	Regression model equations
33-bus	$P_L = 286.3641 \times LoadLevel - 97.0291$
69-bus	$P_L = 318.0325 \times LoadLevel - 108.6808$

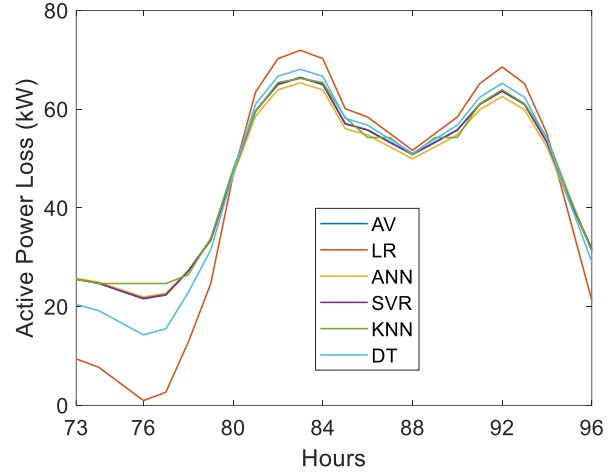


Fig. 9. Comparison of active power loss predictions using the normalized load variation and actual values for 33-bus distribution system.

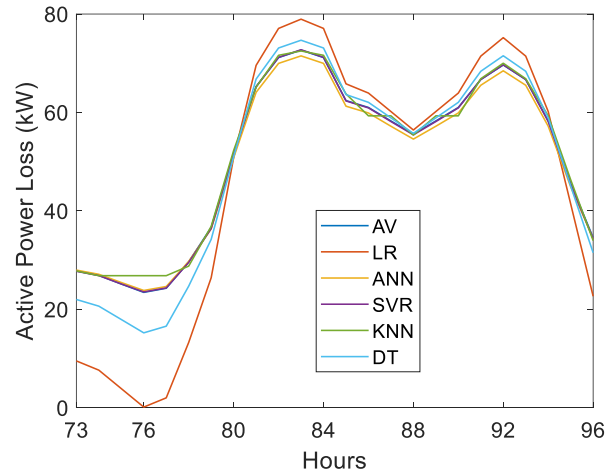


Fig. 10. Comparison of active power loss predictions using the normalized load variation and actual values for 69-bus distribution system.

TABLE V. COMPARISON OF EVALUATIONS FOR ACTIVE POWER LOSS ESTIMATION USING NLV.

Distribution systems	Methods	R <sup>2</sup>	MAPE [%]
33-bus	LR	0.6027	24.9303
	ANN	0.9972	1.4877
	<b>SVR</b>	<b>1</b>	<b>0.1683</b>
	KNN	0.9951	2.3198
	DT	0.9571	8.0833
	69-bus	LR	0.5789
ANN		0.9972	1.4793
<b>SVR</b>		<b>1</b>	<b>0.2203</b>
KNN		0.9951	2.3349
DT		0.9547	8.3852

C. Case 3

In this case, the reactive power losses are estimated using the normalized load variation for 33-bus and 69-bus DSs. Table VI shows the regression model equations of linear regression. Prediction values and actual values are shown in Fig. 11 for 33-bus and in Fig. 12 for 69-bus distribution system. Table VII compares the R-squared and mean absolute percentage error values.

TABLE VI. REGRESSION MODEL EQUATIONS FOR REACTIVE POWER LOSS ESTIMATION.

Distribution systems	Regression model equations
33-bus	$Q_L = 190.977 \times LoadLevel - 64.7527$
69-bus	$Q_L = 144.3074 \times LoadLevel - 49.1576$

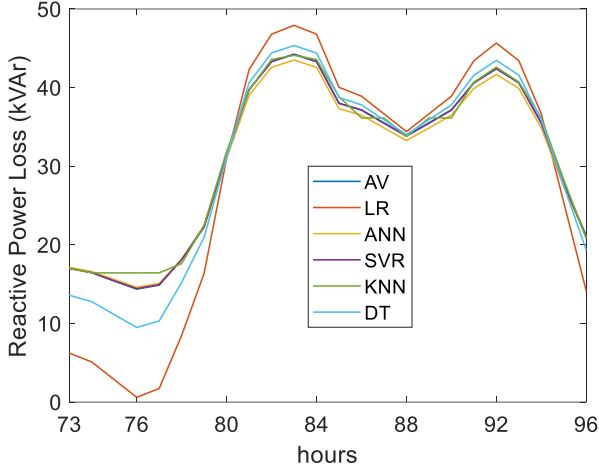


Fig. 11. Comparison of reactive power loss predictions and actual values for 33-bus distribution system.

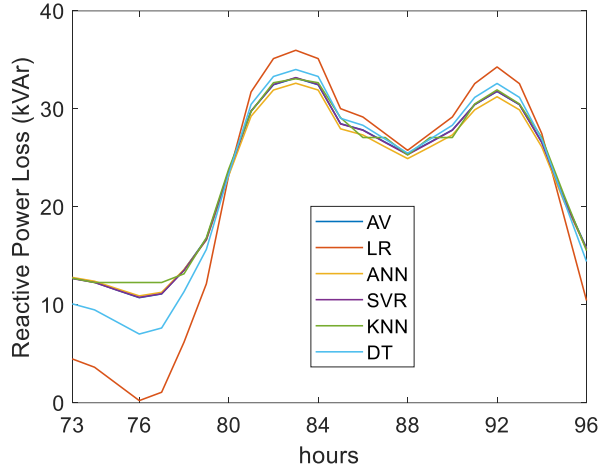


Fig. 12. Comparison of reactive power loss predictions and actual values for 69-bus distribution system.

TABLE VII. COMPARISON OF EVALUATIONS FOR REACTIVE POWER LOSS ESTIMATION.

Distribution systems	Methods	R <sup>2</sup>	MAPE [%]
33-bus	LR	0.6008	25.0064
	ANN	0.9972	1.4869
	<b>SVR</b>	<b>1</b>	<b>0.1716</b>
	KNN	0.9951	2.3208
	DT	0.9569	8.1057
69-bus	LR	0.5879	25.5357
	ANN	0.9972	1.4834
	<b>SVR</b>	<b>1</b>	<b>0.1905</b>
	KNN	0.9951	2.329
	DT	0.9557	8.2615

#### D. Case 4

In this case, the minimum or in other words worst busbar voltages are forecasted using NLV for 33-bus and 69-bus DSs. The regression model equations of LR are given in Table VIII for minimum voltage estimation. Estimation and actual values are shown in Fig. 13 and Fig. 14 for 33-bus and 69-bus DSs, respectively. Comparison of error and performance analysis for this case are given in Table IX.

TABLE VIII. REGRESSION MODEL EQUATIONS FOR MINIMUM BUSBAR VOLTAGE ESTIMATION.

Distribution systems	Regression model equations
33-bus	$V_{min} = -0.0896 \times LoadLevel + 1.0031$
69-bus	$V_{min} = -0.094 \times LoadLevel + 1.0037$

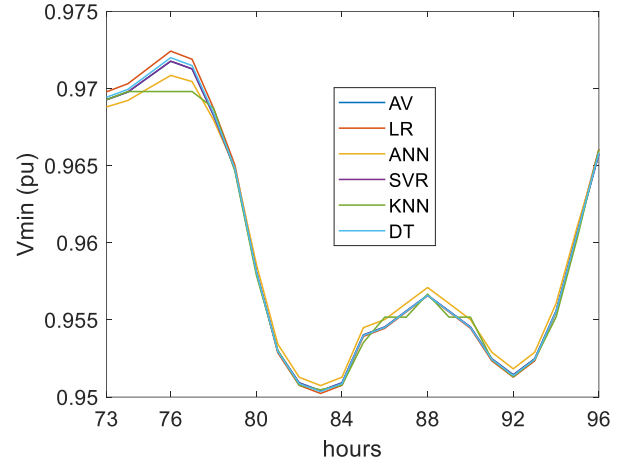


Fig. 13. Comparison of minimum busbar voltage predictions and actual values for 33-bus distribution system.

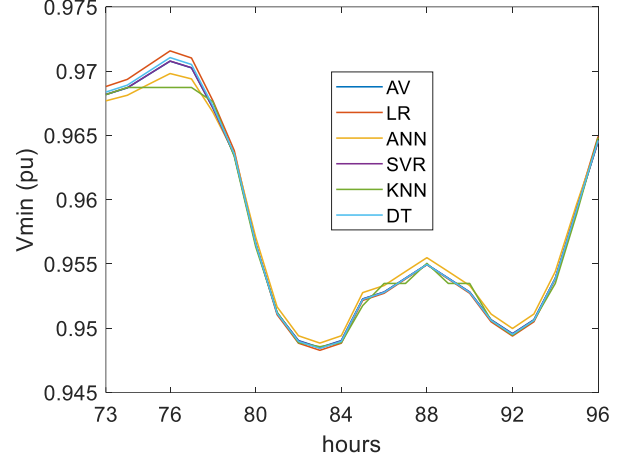


Fig. 14. Comparison of minimum busbar voltage predictions and actual values for 69-bus distribution system.

TABLE IX. COMPARISON OF EVALUATIONS FOR MINIMUM BUSBAR VOLTAGE ESTIMATION.

Distribution systems	Methods	R <sup>2</sup>	MAPE [%]
33-bus	LR	0.9982	0.0245
	ANN	0.9956	0.0476
	SVR	1	0.0010
	KNN	0.9932	0.0411
	DT	0.9998	0.0078
69-bus	LR	0.9976	0.0300
	ANN	0.9956	0.0498
	SVR	1	0.0017
	KNN	0.9932	0.0429
	DT	0.9997	0.0095

#### E. Case 5

In this case, the active power losses are estimated using normalized load variation, DG sizing, and location. Except for LR, the other four methods are applied on 12-bus, 33-bus, and 69-bus DSs. LR is not used in this case because it gives large errors well above the acceptable limit.

Results estimated by the four methods and actual values are shown in Fig. 15 for 12-bus, in Fig. 16 for 33-bus, and in

Fig. 17 for 69-bus DSs. Since the image is not very clear in Figs. 10–12, the image is enlarged in a certain range at the bottom of the table. Error and performance evaluations of these estimations are compared in Table X.

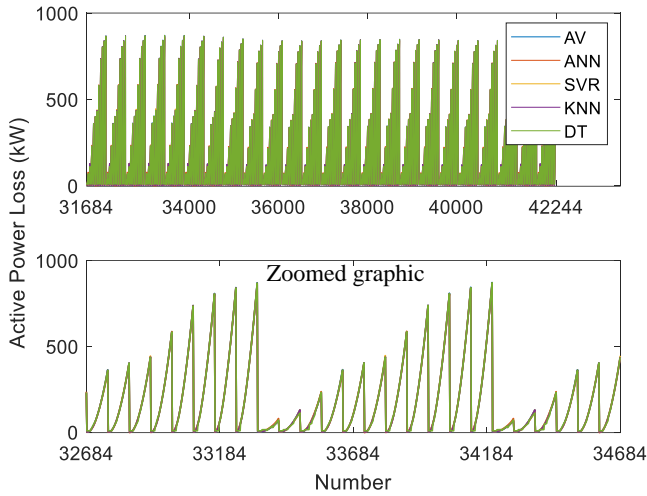


Fig. 15. Comparison of actual value and active power loss predictions using the normalized load variation, DG sizing, and DG location for 12-bus distribution system.

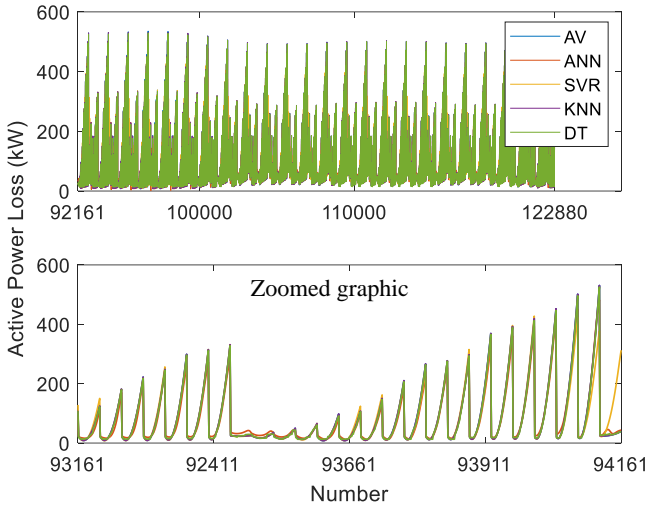


Fig. 16. Comparison of actual value and active power loss predictions using the normalized load variation, DG sizing, and DG location for 33-bus distribution system.

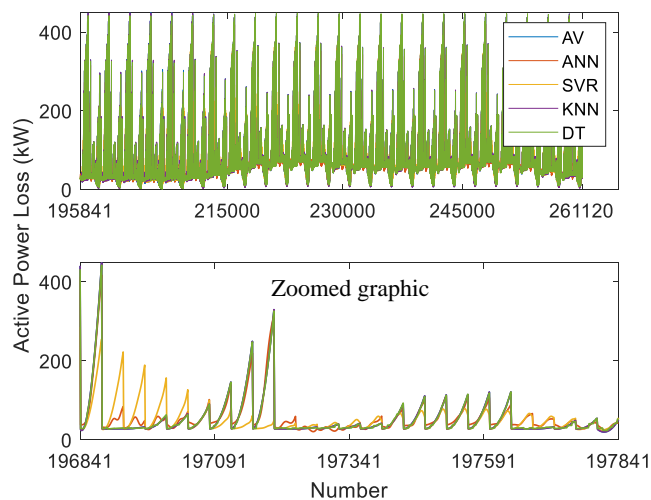


Fig. 17. Comparison of actual value and active power loss predictions using the normalized load variation, DG sizing, and DG location for 69-bus distribution system.

TABLE X. COMPARISON OF EVALUATIONS FOR ACTIVE POWER LOSS ESTIMATION USING NLV, DG SIZE, AND LOCATION.

Distribution systems	Methods	R <sup>2</sup>	MAPE [%]
12-bus	ANN	0.9997	19.3755
	SVR	1	4.7413
	<b>KNN</b>	<b>1</b>	<b>0.4473</b>
	DT	0.9994	44.2325
33-bus	ANN	0.9861	13.9403
	SVR	0.8876	14.0762
	<b>KNN</b>	<b>0.9999</b>	<b>1.1118</b>
	DT	0.9989	5.4546
69-bus	ANN	0.9743	13.5462
	SVR	0.7977	14.4224
	<b>KNN</b>	<b>0.9998</b>	<b>1.4433</b>
	DT	0.9983	5.1876

## VI. DISCUSSION

The following inferences can be made by looking at all tables and figures in Section V. In case 1, the estimates made with all proposed methods are very close to the real values and the best results (0.756 % and 0.9695 %) are obtained by SVR. In the second and third cases, while the predictions made with the other four methods except LR are very successful, LR gives high MAPE (25 %) and low R-squared (0.6) value and SVR gives the best results ( $\cong 0.2$  %). In Section IV, all methods make successful predictions with nearly zero ( $< 0.05$  %) errors and the best (0.001 %) is SVR. In case 5, the best results ( $< 1.5$  %) are obtained by KNN.

## VII. CONCLUSIONS

In this paper, a new DG allocation approach based on estimation is proposed. In the first four cases, DG size, active losses, reactive losses, and worst voltages are estimated using the normalized load variation (single input), while in the last case, only active losses using normalized load level, DG location, and DG size (three inputs) are estimated. Machine learning algorithms, such as LR, ANN, SVR, KNN on WEKA, and DT, for single input predictions are applied on IEEE 33-bus and 69-bus test systems. Mentioned algorithms, except LR for the last case, are applied on 12-bus, 33-bus, and 69-bus distribution systems. SVR gives the best results for single input estimations (cases 1–4), whereas the best results are obtained with KNN for multi-input predictions. Whereas SVR can be used to obtain the best predicted output, such as DG size, system losses, and minimum voltages, in the systems before DG integration, KNN can be used for estimation after DG integration. If there is no linear relationship between the input and output data, it is not appropriate to make predictions with LR. The results of the analysis demonstrate that the proposed approach can be adequate to determine DG size and its effects on DS.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## REFERENCES

- [1] A. Dogan and M. Alci, "Simultaneous optimization of network reconfiguration and DG installation using heuristic algorithms", *Elektronika ir Elektrotechnika*, vol. 25, no. 1, pp. 8–13, 2019. DOI: 10.5755/j01.eie.25.1.22729.
- [2] Y. M. Atwa, E. F. El-Saadany, M. M. A. Salama, and R. Seethapathy,

- “Optimal renewable resources mix for distribution system energy loss minimization”, *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 360–370, 2010. DOI: 10.1109/TPWRS.2009.2030276.
- [3] H. Manafi, N. Ghadimi, M. Ojaroudi, and P. Farhadi, “Optimal placement of distributed generations in radial distribution systems using various PSO and DE algorithms”, *Elektronika ir Elektrotechnika*, vol. 19, no. 10, pp. 53–57, 2013. DOI: 10.5755/j01.eee.19.10.1941.
- [4] S. Essallah, A. Bouallegue, and A. Khedher, “Optimal sizing and placement of DG units in radial distribution system”, *International Journal of Renewable Energy Research (IJRER)*, vol. 8, no. 1, pp. 166–177, 2018.
- [5] P. D. P. Reddy, V. C. V. Reddy, and T. G. Manohar, “Whale optimization algorithm for optimal sizing of renewable resources for loss reduction in distribution systems”, *Renewables: Wind, Water, and Solar*, vol. 4, no. 1, article no. 3, 2017. DOI: 10.1186/s40807-017-0040-1.
- [6] A. Mohamed Imran and M. Kowsalya, “Optimal size and siting of multiple distributed generators in distribution system using bacterial foraging optimization”, *Swarm and Evolutionary Computation*, vol. 15, pp. 58–65, 2014. DOI: 10.1016/j.swevo.2013.12.001.
- [7] D. Q. Hung, N. Mithulananthan, and R. C. Bansal, “Analytical strategies for renewable distributed generation integration considering energy loss minimization”, *Applied Energy*, vol. 105, pp. 75–85, 2013. DOI: 10.1016/j.apenergy.2012.12.023.
- [8] R. Viral and D. K. Khatod, “An analytical approach for sizing and siting of DGs in balanced radial distribution networks for loss minimization”, *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 191–201, 2015. DOI: 10.1016/j.ijepes.2014.11.017.
- [9] M. Q. Duong, T. D. Pham, T. T. Nguyen, A. T. Doan, and H. V. Tran, “Determination of optimal location and sizing of solar photovoltaic distribution generation units in radial distribution systems”, *Energies*, vol. 12, no. 1, p. 174, 2019. DOI: 10.3390/en12010174.
- [10] S. N. G. Naik, D. K. Khatod, and M. P. Sharma, “Analytical approach for optimal siting and sizing of distributed generation in radial distribution networks”, *IET Generation, Transmission & Distribution*, vol. 9, no. 3, pp. 209–220, 2015. DOI: 10.1049/iet-gtd.2014.0603.
- [11] B. R. Pereira, G. R. M. da Costa, J. Contreras, and J. R. S. Mantovani, “Optimal distributed generation and reactive power allocation in electrical distribution systems”, *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 975–984, 2016. DOI: 10.1109/TSTE.2015.2512819.
- [12] M. D. Hraiz, J. A. M. García, R. J. Castañeda, and H. Muhsen, “Optimal PV size and location to reduce active power losses while achieving very high penetration level with improvement in voltage profile using modified Jaya algorithm”, *IEEE Journal of Photovoltaics*, vol. 10, no. 4, pp. 1166–1174, 2020. DOI: 10.1109/JPHOTOV.2020.2995580.
- [13] G. Battapothula, C. Yammani, and S. Maheswarapu, “Multi-objective optimal planning of FCSs and DGs in distribution system with future EV load enhancement”, *IET Electrical Systems in Transportation*, vol. 9, no. 3, pp. 128–139, 2019. DOI: 10.1049/iet-est.2018.5066.
- [14] M. E. Hamidi and R. M. Chabanloo, “Optimal allocation of distributed generation with optimal sizing of fault current limiter to reduce the impact on distribution networks using NSGA-II”, *IEEE Systems Journal*, vol. 13, no. 2, pp. 1714–1724, 2019. DOI: 10.1109/JSYST.2018.2867910.
- [15] A. Khare and S. Rangnekar, “Optimal sizing of a grid integrated solar photovoltaic system”, *IET Renewable Power Generation*, vol. 8, no. 1, pp. 67–75, 2014. DOI: 10.1049/iet-rpg.2012.0382.
- [16] S. H. Lee and J.-W. Park, “Optimal placement and sizing of multiple DGs in a practical distribution system by considering power loss”, *IEEE Transactions on Industry Applications*, vol. 49, no. 5, pp. 2262–2270, 2013. DOI: 10.1109/TIA.2013.2260117.
- [17] L. Wang and C. Singh, “Reliability-constrained optimum placement of reclosers and distributed generators in distribution networks using an ant colony system algorithm”, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 6, pp. 757–764, 2008. DOI: 10.1109/TSMCC.2008.2001573.
- [18] N. C. Hien, N. Mithulananthan, and R. C. Bansal, “Location and sizing of distributed generation units for loadability enhancement in primary feeder”, *IEEE Systems Journal*, vol. 7, no. 4, pp. 797–806, 2013. DOI: 10.1109/JSYST.2012.2234396.
- [19] Z. Moravej and A. Akhlaghi, “A novel approach based on cuckoo search for DG allocation in distribution network”, *International Journal of Electrical Power & Energy Systems*, vol. 44, no. 1, pp. 672–679, 2013. DOI: 10.1016/j.ijepes.2012.08.009.
- [20] K. Nekooei, M. M. Farsangi, H. Nezamabadi-Pour, and K. Y. Lee, “An improved multi-objective harmony search for optimal placement of DGs in distribution systems”, *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 557–567, 2013. DOI: 10.1109/TSG.2012.2237420.
- [21] S. Elsaiah, M. Benidris, and J. Mitra, “Analytical approach for placement and sizing of distributed generation on distribution systems”, in *Proc. of 2014 Clemson University Power Systems Conference*, 2014, pp. 1–7. DOI: 10.1109/PSC.2014.6808097.
- [22] R. A. Jabr and B. C. Pal, “Ordinal optimisation approach for locating and sizing of distributed generation”, *IET generation, transmission & distribution*, vol. 3, no. 8, pp. 713–723, 2009. DOI: 10.1049/iet-gtd.2009.0019.
- [23] A. M. Pasha, H. H. Zeineldin, E. F. El-Saadany, and S. S. Alkaabi, “Optimal allocation of distributed generation for planning master-slave controlled microgrids”, *IET Generation, Transmission & Distribution*, vol. 13, no. 16, pp. 3704–3712, 2019. DOI: 10.1049/iet-gtd.2018.5872.
- [24] A. Toktarova, L. Gruber, M. Hlusiak, D. Bogdanov, and C. Breyer, “Long term load projection in high resolution for all countries globally”, *International Journal of Electrical Power & Energy Systems*, vol. 111, pp. 160–181, 2019. DOI: 10.1016/j.ijepes.2019.03.055.
- [25] M. Jawad, M. S. A. Nadeem, S.-O. Shim, I. R. Khan, A. Shaheen, N. Habib, L. Hussain, and W. Aziz, “Machine learning based cost effective electricity load forecasting model using correlated meteorological parameters”, *IEEE Access*, vol. 8, pp. 146847–146864, 2020. DOI: 10.1109/ACCESS.2020.3014086.
- [26] P. W. Khan and Y.-C. Byun, “Genetic algorithm based optimized feature engineering and hybrid machine learning for effective energy consumption prediction”, *IEEE Access*, vol. 8, pp. 196274–196286, 2020. DOI: 10.1109/ACCESS.2020.3034101.
- [27] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, “Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities”, *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2015. DOI: 10.1109/TSG.2014.2364233.
- [28] M. Rejc and M. Pantos, “Short-term transmission-loss forecast for the slovenian transmission power system based on a fuzzy-logic decision approach”, *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1511–1521, 2011. DOI: 10.1109/TPWRS.2010.2096829.
- [29] P. Zeng, M. Jin, and Md. F. Elahe, “Short-term power load forecasting based on cross multi-model and second decision mechanism”, *IEEE Access*, vol. 8, pp. 184061–184072, 2020. DOI: 10.1109/ACCESS.2020.3028649.
- [30] O. Valgaev, F. Kupzog, and H. Schmeck, “Building power demand forecasting using K-nearest neighbours model—practical application in Smart City Demo Aspern project”, *CIRED-Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 1601–1604, 2017. DOI: 10.1049/oap-cired.2017.0419.
- [31] E. E. Elattar, J. Goulermas, and Q. H. Wu, “Electric load forecasting based on locally weighted support vector regression”, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 4, pp. 438–447, 2010. DOI: 10.1109/TSMCC.2010.2040176.
- [32] K.-B. Song, Y.-S. Baek, D. H. Hong, and G. Jang, “Short-term load forecasting for the holidays using fuzzy linear regression method”, *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 96–101, 2005. DOI: 10.1109/TPWRS.2004.835632.
- [33] M. V. Selvi and S. Mishra, “Investigation of performance of electric load power forecasting in multiple time horizons with new architecture realized in multivariate linear regression and feed-forward neural network techniques”, *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5603–5612, 2020. DOI: 10.1109/TIA.2020.3009313.
- [34] M. S. Kandil, S. M. El-Debeiky, and N. E. Hasanien, “Long-term load forecasting for fast developing utility using a knowledge-based expert system”, *IEEE transactions on Power Systems*, vol. 17, no. 2, pp. 491–496, 2002. DOI: 10.1109/TPWRS.2002.1007923.
- [35] S.-M. Baek, “Mid-term load pattern forecasting with recurrent artificial neural network”, *IEEE Access*, vol. 7, pp. 172830–172838, 2019. DOI: 10.1109/ACCESS.2019.2957072.
- [36] H. Bo, Y. Nie, and J. Wang, “Electric load forecasting use a novelty hybrid model on the basic of data preprocessing technique and multi-objective optimization algorithm”, *IEEE Access*, vol. 8, pp. 13858–13874, 2020. DOI: 10.1109/ACCESS.2020.2966641.
- [37] B. Li, J. Zhang, Y. He, and Y. Wang, “Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test”, *IEEE Access*, vol. 5, pp. 16324–16331, 2017. DOI: 10.1109/ACCESS.2017.2738029.
- [38] C. Feng, M. Sun, and J. Zhang, “Reinforced deterministic and



- probabilistic load forecasting via  $\mathcal{QS}$ -learning dynamic model selection”, *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1377–1386, 2020. DOI: 10.1109/TSG.2019.2937338.
- [39] D. Wu, B. Wang, D. Precup, and B. Boulet, “Multiple kernel learning-based transfer regression for electric load forecasting”, *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1183–1192, 2020. DOI: 10.1109/TSG.2019.2933413.
- [40] M. Alhussein, K. Aurangzeb, and S. I. Haider, “Hybrid CNN-LSTM model for short-term individual household load forecasting”, *IEEE Access*, vol. 8, pp. 180544–180557, 2020. DOI: 10.1109/ACCESS.2020.3028281.
- [41] Y. Hong, Y. Zhou, Q. Li, W. Xu, and X. Zheng, “A deep learning method for short-term residential load forecasting in smart grid”, *IEEE Access*, vol. 8, pp. 55785–55797, 2020. DOI: 10.1109/ACCESS.2020.2981817.
- [42] I. P. Panapakidis, N. Skiadopoulou, and G. C. Christoforidis, “Combined forecasting system for short-term bus load forecasting based on clustering and neural networks”, *IET Generation, Transmission & Distribution*, vol. 14, no. 18, pp. 3652–3664, 2020. DOI: 10.1049/iet-gtd.2019.1057.
- [43] Y. P. Soni and E. Fernandez, “Bus voltage estimation of IEEE-12 bus radial structured microgrid feeder: A regression-based approach”, in *Proc. of 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCon)*, 2020, pp. 619–622. DOI: 10.1109/GUCon48875.2020.9231231.
- [44] T. C. Akinci, “Short term wind speed forecasting with ANN in Batman, Turkey”, *Elektronika ir Elektrotehnika*, vol. 107, no. 1, pp. 41–45, 2011.
- [45] W. S. McCulloch and W. Pitts, “A logical calculus of ideas immanent in nervous activity”, *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943. DOI: 10.1007/BF02478259.
- [46] Yapay sinir ağları, Wikipedi, özgür ansiklopedi. [Online]. Available: [https://tr.wikipedia.org/wiki/Yapay\\_sinir\\_ağları](https://tr.wikipedia.org/wiki/Yapay_sinir_ağları)
- [47] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola and V. Vapnik, “Support vector regression machines”, in *Proc. of the 9th International Conference on Neural Information Processing Systems*, 1996, pp. 155–161.
- [48] S. Mahdevari, K. Shahriar, S. Yagiz, and M. A. Shirazi, “A support vector regression model for predicting tunnel boring machine penetration rates”, *International Journal of Rock Mechanics and Mining Sciences*, vol. 72, pp. 214–229, 2014. DOI: 10.1016/j.ijrmms.2014.09.012.
- [49] K-nearest neighbours algorithm, Wikipedia, the free encyclopedia. [Online]. Available: [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)
- [50] 4 Distance Measures for Machine Learning. [Online]. Available: <https://machinelearningmastery.com/distance-measures-for-machine-learning/>
- [51] W. Zhang, X. Chen, Y. Liu, and Q. Xi, “A distributed storage and computation k-nearest neighbor algorithm based cloud-edge computing for cyber-physical-social systems”, *IEEE Access*, vol. 8, pp. 50118–50130, 2020. DOI: 10.1109/ACCESS.2020.2974764.
- [52] Data repository. [Online]. Available: <https://github.com/VasuBalan/RA>



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) license (<http://creativecommons.org/licenses/by/4.0/>).