Supervised Learning Approach for State Estimation of Unmeasured Points of Distribution Network

GWANGPYO HONG AND YUN-SU KIM, (Member, IEEE)
School of Integrated Technology, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Yun-Su Kim (yunsukim@gist.ac.kr)

ABSTRACT This paper presents a new approach to state estimation (SE) of distribution networks, which becomes more complex when there is lack of monitoring. Several studies have been carried out on SE to compensate for the lack of monitoring; however, the observability of the distribution system is poor compared to the transmission system. In the proposed approach, the representative load profile and the electricity charges of consumers are required to obtain the load profile of each consumer. In addition, the uncertainty was considered owing to the poor accuracy of these obtained load profiles, and the results were analyzed according to the uncertainty. The obtained load profiles were used to calculate the voltage magnitudes and angles by power flow calculations, and the calculated voltage magnitudes and angles were used to train the used supervised learning algorithms including the feed-forward neural network (FFNN), linear regression (LR), and support vector machine (SVM). IEEE 13-, 34-, and 37-node test feeders were used to verify the proposed approach. The proposed approach is not applicable to a terminal bus; however, the voltage magnitudes and angles of consecutive unmeasured buses more than two can be estimated. In addition, the impact of input data on the results was analyzed for each algorithm, and the impact of measurement errors was also analyzed for FFNN and SVM.

INDEX TERMS Distribution system state estimation (DSSE), supervised learning, support vector machine (SVM), linear regression (LR), feed-forward neural network (FFNN).

I. INTRODUCTION

A distribution system that was built and designed to accommodate expected seasonal demand variations and growth can become more complex and difficult to control [1]. However, the measuring and monitoring of a distribution system are not well established unlike those of a transmission system. In particular, the measurement devices are often limited to substations [2] or are installed in a low number of buses in a low-voltage (LV) network [3]–[6]. It is practically and economically impossible to install measurement devices in all measurement locations because the number of measurement points increases exponentially from the substation to the end customers [8]. In addition, the reliability of the installed measurement devices is poor because the measured values of the two measurement devices installed under the same

transformer are different [9]. For this reason, it is difficult to determine the cause of the problem with an LV network. Therefore, measurement with high accuracy in distribution systems becomes essential. A number of studies have been carried out to improve upon the lack of monitoring, to supplement insufficient data, and to filter out bad data through smart meters, the distribution system state estimation (DSSE) method, etc.

Deploying smart meters was proposed in [3], [6] to obtain necessary information from buses through communication; however, installation of smart meters in every bus is technically impossible. It cannot be guaranteed that the gathered data are measured at the same time due to the limited bandwidth, infrequent delays and synchronization problems in communication. Furthermore, delay differs among smart meters, as synchronizing the measurements of huge number of devices is too expensive. Aggregators are required to manage vast amount of data, which also incurs costs. In addition,

The associate editor coordinating the review of this manuscript and approving it for publication was Gaetano Zizzo.

data missing still occurs due to unexpected device power off, communication failure, measurement error, or other unknown errors [7]. It is still not clear how the communications infrastructure for transmitting expected measurement data to a distribution management system (DMS) will change [6].

Pseudo-measurements are often used to compensate for the lack of data in monitoring. Generally, pseudo-measurements are performed before state estimation (SE) to improve the accuracy of estimated values and measured values [10]. One of the valuable functions of SE is detecting and identifying bad measurements. Because of these functions, in transmission systems, SE is an effective way to obtain the states of systems. However, state estimation of the resulting distribution systems may be unreliable owing to the poor accuracy and high variance of pseudo-measurements [5], [6], [10].

In [1], a new methodology using artificial neural networks (ANNs) was proposed for pseudo-measurement modeling. Load profiles and load flow simulation results were trained through the ANN, and the error was modeled through a Gaussian mixture model (GMM). The results of the GMM and ANN training were used as input data for DSSE to increase the accuracy. However, this requires much more input data than output data for ANN training to estimate the voltage magnitudes and voltage angles of the target buses. In addition, the approach with an ANN was proposed to estimate state variables in [11]; however, this requires measurements and cannot estimate the state variables of more than two buses at a time. The authors of [5], [10] introduced a technique for meter placement to improve the quality of voltage and angle estimates across a network. In [4], a novel probabilistic approach was presented to distribution network observability. Also proposed were a revised branch current-based distribution system state estimation algorithm for obtaining a snapshot of the distribution system as accurately as possible in [12], a new pseudo load profile determination approach for estimating states of unmeasured buses and improving the accuracy of pseudo-measurements in [2], [13], and a method for enhancing the observability of a distribution system for compensating of missing specifications on non-metered buses in [3]. In addition, several approaches for DSSE using machine learning have been proposed in [14]–[17]. In [14] and [15], methods were proposed to detect harmonic and to estimate voltage harmonic waveforms in distribution system using echo state network (ESN), respectively. In [16], a hybrid machine learning with optimization approach for determining the way of initializing Gauss–Newton through the shallow neural network using available measurements is proposed, and a nodal load estimation approach using relevance vector machine (RVM) was proposed in [17]. However, approaches were all based on measurements.

In this paper, a new approach is proposed that uses supervised learning algorithms to estimate the voltage magnitude and voltage angle of unmeasured consecutive buses. If the voltage and load information of adjacent buses of one unmeasured bus, which is called a "target bus", are known, then the

information of the target bus can be calculated. This is why we assumed the consecutive unmeasured buses more than two buses.

This approach requires the electricity charges of customers, a representative load profile (RLP), and the voltage data of adjacent buses of the target buses. The voltage magnitude and angle, which cannot be obtained from the RLP and the electricity charges, are obtained by power flow calculation for training and testing. In the training step, the active and reactive power of the substation, and the voltage magnitude and angle data of several buses, which is called "input buses", among adjacent buses of the target buses, which are most affected by the state variables of the target buses, are required.

Only a few measurements are required to estimate the consecutive unmeasured target buses. The proposed approach was conducted on three supervised learning algorithms including the feed-forward neural network (FFNN), linear regression (LR), and support vector machine (SVM). To verify the proposed approach, the voltage magnitude and angle of the assumed unmeasured buses were estimated for three networks of IEEE 13-, 34-, and 37-node test feeders [18]–[20]. The results through used three supervised learning algorithms for each network, when one of input buses is missing, were obtained to analyze the impact of each input bus. In addition, the error distribution when there are measurement errors in the data of the input buses was also analyzed based on the published guideline for commercialized products.

The proposed approach is best suited for LV networks which have insufficient measurements. Furthermore, it is advantageous for not requiring measurement data for training of supervised learning algorithms unlike the previous works. Moreover, no additional process of error modeling, such as GMM, other than data generation, training, and estimation is required unlike [1]. In addition, the proposed approach is applicable even though one of input buses data is missing or delay occurs due to communication failure, measurement error, or other unknown reasons.

While the penetration of smart meters has been activated, it still needs more time to be installed at all points. The proposed approach uses information from small number of measurement devices to estimate the voltage information of consecutive unmeasured target buses. Therefore, it can also be used as a way to estimate the voltage information of the whole network using only the information acquired from some smart meters in the transition period before the smart meter is 100% deployed. However, the application of the proposed approach to terminal buses of the network has some limitation of poor accuracy. Moreover, the validation of the proposed approach has only been verified for networks consisting of consumers of the same load type without generators.

This paper is structured as follows: Section II describes how the data are generated and organized with the RLP for training supervised learning algorithms. How the generated data are used to estimate the voltage magnitude and angle,

and a brief explanation of the used supervised learning algorithms, are presented in Section III. The results are presented in Section IV, and Section V outlines the conclusions.

II. DATA CONFIGURATION

In order to obtain valid outputs for any inputs, sufficient data must be used to train for various input and output data. In this paper, the RLP and electricity charge of consumers were used to generate these data. Load modeling consists of two steps: selecting a load model structure and identifying the parameter of the load model structure. The types of load model structures include static load model, dynamic load model, ANN-based load model, etc., and there are two ways of identifying the parameters of load model structure which are component-based approach and measurement-based approach [21]. However, these approaches require the characteristics of individual load components or field measurements which are difficult to be obtain. Therefore, in this paper, the data are generated to train and estimate the voltage information based on the RLP and electricity charge of consumers which can be obtained without additional measurements.

A. REPRESENTATIVE LOAD PROFILE

The RLP can be defined as the demand pattern of a single or group of consumers over a period of time. It is mainly used to obtain pseudo-measurements, to select transformer ratings or to determine the charges of consumers, who do not have a measurement device [22]. In addition, if the monthly energy consumption of a particular consumer is known, then his daily load curve can be determined by multiplying the values in the RLP in p.u. of the range to which it belongs by its power base [23].

In other words, monthly energy consumption can be calculated through the electricity charge of certain consumer. Moreover, the peak load can be estimated through the monthly energy consumption using (1), as follow:

$$\text{monthly energy consumption} = 30 \times \int (\text{peak load}) \times RLP(t) dt \quad (1)$$

where $RLP(t)$ is the power load of RLP in percentage of peak load at t. Thus, the daily load curve of a particular consumer can be obtained if the RLP and the electricity charges of certain consumers are available.

B. DATA GENERATION

In this paper, we assumed that the electricity charges of consumers and the power curve supplied by the substation are known. If the power supplied by the substation can be monitored and load types (such as industrial, commercial or residential) of the customers who are supplied by the substation, are the same, then the RLP can be inferred through the power curve of the substation. Therefore, the load profiles of consumers can be obtained. However, the accuracy of the obtained load profiles is poor due to a high variation of load patterns of consumers. For this reason, this paper also analyzed the estimated voltage magnitude and angle according to the variation. This is called the "uncertainty" as shown in Fig. 1.

Fig. 1 presents the generation of M random load profiles within the range of the uncertainty based on the RLP for M
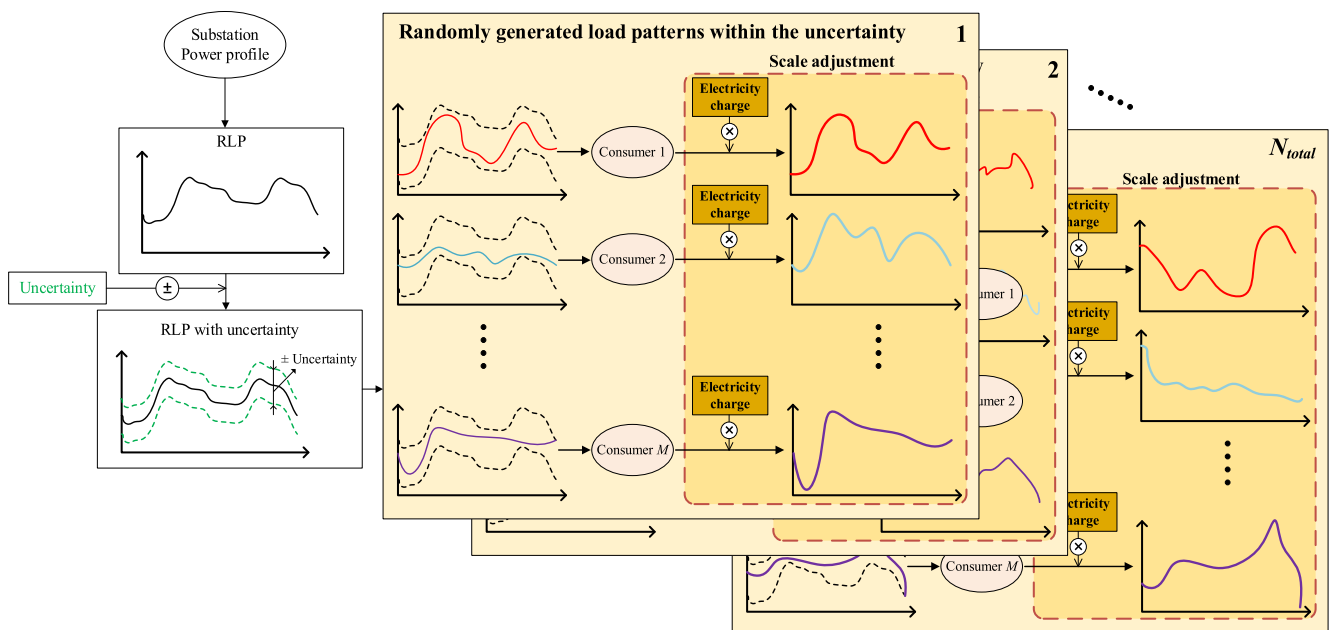

FIGURE 1. Overall process of data generation and configuration for *M* consumers.

TABLE 1. Variables for Generation load profiles.

Variable	Meaning
$P(n,t,m)$	Active power load of mth consumer of nth set at t (h)
$Q(n,t,m)$	Reactive power load of mth consumer of nth set at t
$RLP_P(t)$	Active power load of RLP in percentage of peak load at t
$RLP_Q(t)$	Reactive power load of RLP in percentage of peak load at t
$R_P(n,t,m)$	Random variable for active power load of mth consumer of nth set at t
$R_Q(n,t,m)$	Random variable for reactive power load of mth consumer of nth set at t
$S_P(m)$	Scale factor of mth consumer, which is calculated peak of active load from electricity charge in p.u.
$S_Q(m)$	Scale factor of mth consumer, which is calculated peak of reactive load from electricity charge in p.u.

consumers using (2) and (3), as follows:

$$P(n,t,m) = RLP_P(t) \times (1 + R_P(n,t,m)) \times S_P(m) \quad (2)$$

$$Q(n,t,m) = RLP_Q(t) \times (1 + R_Q(n,t,m)) \times S_Q(m) \quad (3)$$

where the variables are listed in Table 1. $R_P(n,t,m)$ and $R_Q(n,t,m)$ are introduced to consider the uncertainty, and these parameters have randomly generated values between $-uncertainty/(100\%)$ and $+uncertainty/(100\%)$. The unit of the uncertainty is %, and the uncertainty is a uniformly distributed random variable. In this paper, the uncertainty ranges from 10% to 50% in 10% intervals. For example, when the uncertainty is 20%, the range of values that $R_P(n,t,m)$ and $R_Q(n,t,m)$ can have is between –0.2 and +0.2. In addition, the values of these variables are different for n, m, and t. In other words, if one of n, m, and t is different, then the value of R_P is also different. The same is true for R_Q. The reason for adding 1 to the random variable in the second term in (2) and (3) is to vary the uncertainty from the original value.

The randomly generated load profiles, which are based on the RLP and the uncertainty, are scaled based on Sm calculated through each consumer's electricity charge. These generated load profiles are used for power flow calculations to obtain the voltage magnitudes, voltage angles, and generated daily data including the voltage magnitudes, angles, active power, and reactive power. These are defined as a "set," as shown in Figs. 1 and 2.
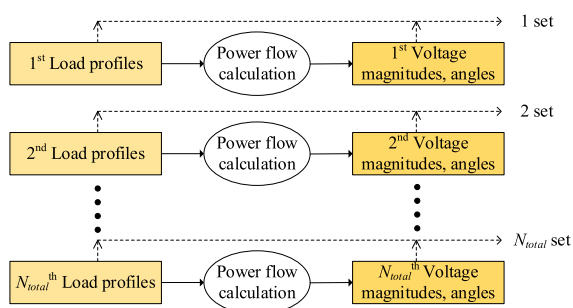

FIGURE 2. Generation data "set" through power flow calculation.

In addition, each data set includes different randomly generated load profiles and the obtained voltage magnitudes and angles calculated using these load profiles. The power

flow calculation was conducted using OpenDSS, which is a power system simulation tool developed by the Electric Power Research Institute (EPRI). Therefore, the additional data, which are the voltage magnitudes, angles, and the power of the slack bus, can be obtained from the generated load profiles by power flow calculations for training.

The data were generated based on IEEE 13-, 34-, and 37-node test feeders. A total of N_{total} sets were generated for each IEEE network. N_{train} sets were used for training and N_{test} sets were used for testing of the training results. A graph of changes in the rate of peak load usage during the day in [24] was used as the RLP, and the load value of each bus of the IEEE network was assumed as the peak load to generate the data for training and testing. In other words, the load profiles for all buses of each IEEE network were generated by multiplying the load usage pattern graph in [24] by the load usage of each bus based on (2) and (3).

III. SIMULATION ENVIRONMENT
A. SUPERVISED LEARNING ALGORITHMS
In this paper, three supervised learning algorithms–FFNN, LR, and SVM–were used. The artificial neural network was motivated by the structure of a real brain. Each neuron, which together comprise the network, is connected to least one neuron, and each connection is evaluated by a weight coefficient [25]. The network learns the relationship between inputs and outputs using the activation function and the cost function through the weight coefficient. The role of the activation function in the neural network is to represent a nonlinear relationship between inputs and outputs. This feature makes the neural network suitable for nonlinear relationships. The FFNN is a representative supervised learning algorithm based on the ANN [26].

Simple linear regression is an algorithm that estimates the response variable by obtaining an equation of a straight line that represents best the given data points in the coordinate distribution of the dependent variable for an independent variable. In addition, it can be applied to multi independent variables [27].

The SVM is a technique that was proposed by Vapnik and coworkers in 1992. The objective of the SVM is to find a hyperplane in an N-dimensional space that distinctly classifies the data points by maximizing the margin. In addition, it can be used to estimate an unknown target value as a regression algorithm [28].

The relation between the power and the voltage is as follows:

$$P_i = \sum_{k=1}^{N} |V_i||V_k| (G_{ik}\cos\theta_{ik} + B_{ik}\sin\theta_{ik}), \quad i=1,2,\cdots,N \quad (4)$$

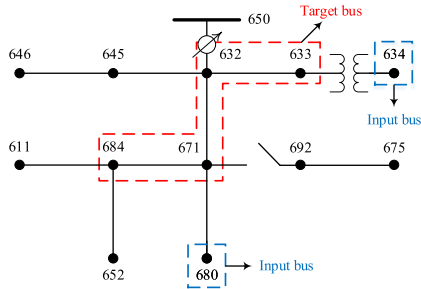$$Q_i = \sum_{k=1}^{N} |V_i||V_k| (G_{ik}\sin\theta_{ik} - B_{ik}\cos\theta_{ik}), \quad i=1,2,\cdots,N \quad (5)$$
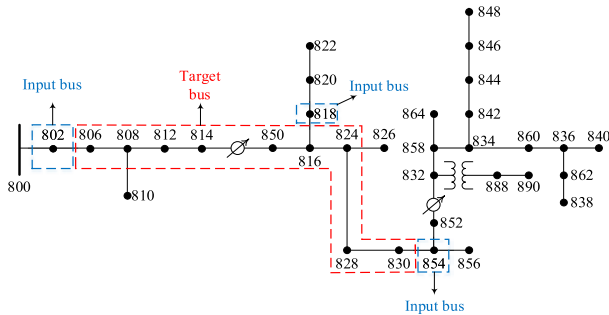
FIGURE 3. IEEE 13-node Test Feeder.
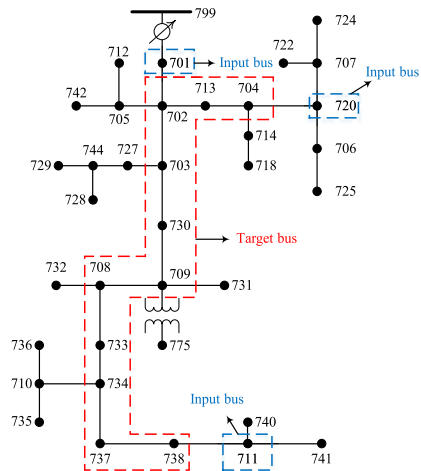

FIGURE 4. IEEE 34-node Test Feeder.

FIGURE 5. IEEE 37-node Test Feeder.

TABLE 2. Study cases.

Network type	Target bus	Input bus
13-node Test Feeder	632, 671, 633, 684	680, 634
34-node Test Feeder	806, 808, 812, 814, 850, 816, 824, 828, 830	802, 818, 854
37-node Test Feeder	702, 703, 730, 709, 708, 733, 734, 737, 738, 713, 704	701, 711, 720

TABLE 3. Matrices for training and test.

Matrix Name	Meaning	Matrix Size
P_s	Active power of slack bus	$N_{total} \times 24$
Q_s	Reactive power of slack bus	$N_{total} \times 24$
V_{in}	Voltage magnitude of adjacent buses	$N_{total} \times 24 \times M_{in}$
θ_{in}	Voltage angle of adjacent buses	$N_{total} \times 24 \times M_{in}$
t	Hour	$N_{total} \times 24$
V_{tar}	Voltage magnitude of target buses	$N_{train} \times 24 \times M_{tar}$
V_{true}	True voltage magnitude of target buses for calculating the error	$N_{test} \times 24 \times M_{tar}$
V_{est}	Estimated voltage magnitude of target buses	$N_{test} \times 24 \times M_{tar}$
θ_{tar}	Voltage angle of target buses	$N_{train} \times 24 \times M_{in}$
θ_{true}	True voltage angle of target buses for calculating the error	$N_{test} \times 24 \times M_{in}$
θ_{est}	Estimated voltage angle of target buses	$N_{test} \times 24 \times M_{in}$

FIGURE 6. FFNN training concept.

where P_i, Q_i, G_{ik}, B_{ik}, V_i and V_k are the active and reactive power of the ith bus of the network, the elements of the conductance matrix and susceptance matrix, and the voltages of the ith bus and k^{th} bus, respectively. As denoted in (4) and (5), the relationship between the power and voltage is nonlinear. The FFNN was used to learn these nonlinear features and to estimate state variables. Moreover, the LR was used to learn regarding the unexpected linear characteristics between the power and the voltages. However, the linearity of the relationship cannot be predicted easily. Hence, the SVM was used, since it is advantageous when the relationships between data are not well understood [29].

B. SIMULATION CASES

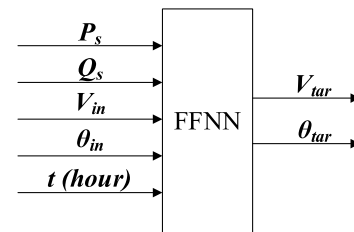IEEE 13-, 34-, and 37-node Test Feeder networks are shown in Figs. 3–5, respectively. The proposed approach was tested assuming unmeasured consecutive buses for each network, as described in Table 2, and a single phase was considered for estimating the voltage magnitudes and angles of the target buses.

The FFNN is organized with one hidden layer of 12 neurons between the input layer and the output layer. The data matrices used for training the FFNN, LR, and SVM are represented in Table 3. The number of target buses and input buses are defined as M_{tar} and M_{in}, respectively. $V_{tar}(n,t,m)$ is the element of matrix V_{tar} of the mth target consumer at t in the nth set, and this expression is also applied to other matrices including V_{true}, V_{est}, θ_{tar}, θ_{true}, and θ_{est}. The matrices used as input or output are shown in Figs. 6 and 7 for each algorithm. In the case of the LR and SVM, the training was conducted for every hour and for each target bus repeatedly, unlike the FFNN.

In other words, since the voltage magnitude and angle of the target bus at the time determined by the user was
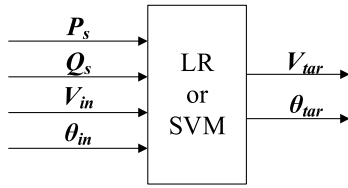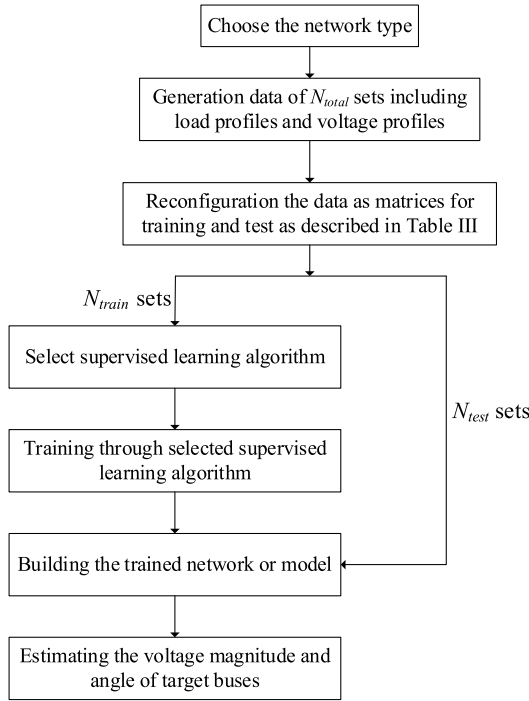
FIGURE 7. LR or SVM training concept.

FIGURE 8. Flowchart of training and testing.

estimated, no time data was used as the input. Fig. 8 shows a flowchart of the overall learning process using the generated data described in the previous section. The total N_{total} sets of load profiles were generated, and the same amounts of the voltage magnitudes and angles were obtained through these load profiles. The load profiles, voltage magnitudes, and angles were rearranged into matrices for use in training and testing, as shown in Table 3. N_{train} sets of the N_{total} sets were used to train the three supervised learning algorithms, and the rest were used for testing.

The configuration, training, and testing of the FFNN, LR, and SVM were performed using MATLAB [30], [31].

IV. SIMULATION RESULTS

In this paper, the maximum absolute magnitude error was calculated to analyze the error. Each equation is as follows:

$$\varepsilon_{V,p.u.}(m) = \max \begin{pmatrix} a_{11} & \cdots & a_{1t} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nt} \end{pmatrix},$$
$$m = 1, 2, \cdots, M_{tar} \qquad (6)$$

$$\varepsilon_{\theta,deg}(m) = \max \begin{pmatrix} b_{11} & \cdots & b_{1t} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nt} \end{pmatrix},$$
$$m = 1, 2, \cdots, M_{tar} \qquad (7)$$

where

$$a_{nt} = |V_{true}(n, t, m) - V_{est}(n, t, m)| \qquad (8)$$
$$b_{nt} = |\theta_{true}(n, t, m) - \theta_{est}(n, t, m)| \qquad (9)$$

for $n = 1, 2, \cdots, N_{test}$ and $t = 1, 2, \cdots, 24$. Equations (6) and (7) are the maximum absolute magnitude error of the estimated voltage magnitude and the angle for the mth of M_{tar} target consumers, respectively. The elements of the matrices in (6) and (7) are described in (8) and (9). In other words, $\varepsilon_{V,p.u.}(m)$ and $\varepsilon_{\theta,deg}(m)$ are the magnitude of the maximum absolute errors of the mth target consumer for all day and, all test sets.

A. DETERMINATION OF NUMBER OF SETS

In order to determine the values of N_{total} and N_{train}, the maximum value of the absolute percentage error of the estimated voltage magnitude for a 13-node network was analyzed when the uncertainty equals 50%. As shown in Fig. 9, the maximum absolute percentage error of the estimated voltage magnitude was obtained as N_{total} increased. N_{train} was set to 80% of N_{total}, and the FFNN was used to obtain the error tendency. The minimum error occurred when the value of N_{total} was 300, and the error increased drastically when N_{total} was 500 or higher. For this reason, the value of N_{total} was set to 300. Fig. 10 shows the maximum error of the estimated voltage magnitude according to the ratio of N_{train} to N_{total} when N_{total} equals 300. The error was decreased when the ratio was 60% or higher, and the minimum error occurred when the ratio was 80%. Hence, N_{train} was set to 240, and N_{test} was set to 60. These tendencies were equally obtained for most uncertainties in 13-, 34-, and 37-node test feeder networks, and the cases with the most prominent changes were shown in Figs. 9 and 10.
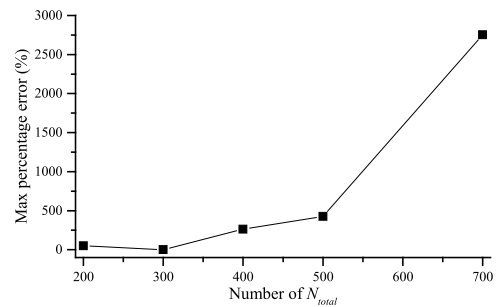

FIGURE 9. Change of maximum error of estimated voltage magnitude through FFNN according to number of N_{total}.

B. VOLTAGE MAGNITUDE

The maximum values of $\varepsilon_{V,p.u.}$ in 10^{-2} p.u. of the estimated voltage magnitude with respect to the uncertainty are listed in Table 4. The values in Table 4 were obtained for all target

TABLE 4. Error of estimated voltage magnitude.

Algorithm type	Network type	Maximum values of $\varepsilon_{V,p.u.}(m)$ in 10^{-2} p.u. for all m according to the uncertainty, $m=1,2,\cdots,M_{tar}$				
		10% uncertainty	20% uncertainty	30% uncertainty	40% uncertainty	50% uncertainty
FFNN	13-node	0.42	0.42	0.77	1.00	0.86
	34-node	0.03	0.06	0.07	0.09	0.11
	37-node	0.04	0.06	0.22	0.52	0.48
LR	13-node	0.04	0.08	0.15	0.22	0.34
	34-node	0.02	0.05	0.07	0.08	0.08
	37-node	0.02	0.05	0.07	0.07	0.14
SVM	13-node	0.12	0.12	0.13	0.17	0.29
	34-node	0.33	0.37	0.46	0.70	0.67
	37-node	0.19	0.12	0.15	0.24	0.22
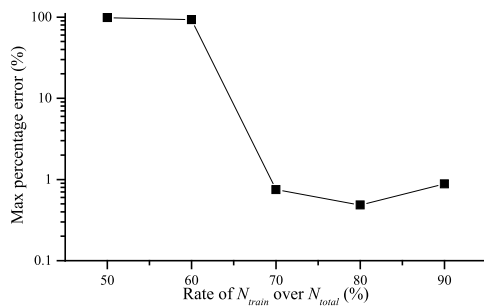
FIGURE 10. Change of maximum error of estimated voltage magnitude through FFNN according to ratio of N_{train} to N_{total}.

FIGURE 12. Max magnitude error result through LR of 13-node Test Feeder when uncertainty equals 50%.

FIGURE 11. Max magnitude error result through FFNN of 13-node Test Feeder when uncertainty equals 40%.
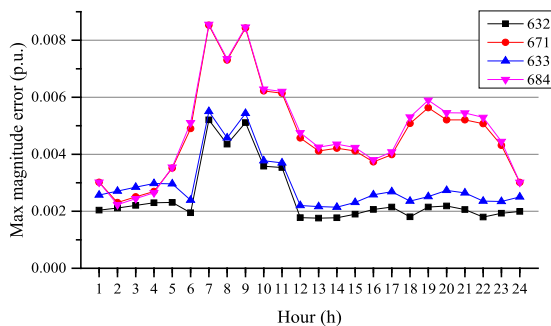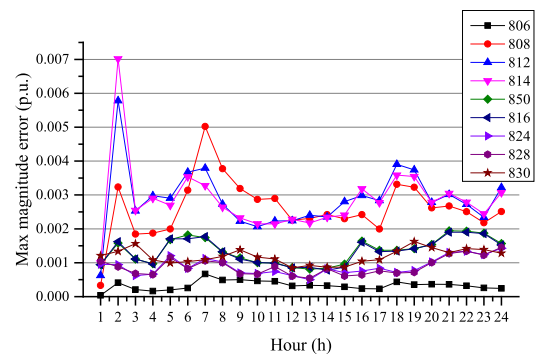

FIGURE 13. Max magnitude error result through SVM of 34-node Test Feeder when uncertainty equals 40%.

buses, all test sets, and for a day. All values of $\varepsilon_{V,p.u.}$ through the FFNN, LR and SVM are below 1.00×10^{-2} p.u., 0.34×10^{-2} p.u., and 0.70×10^{-2} p.u., respectively. Therefore, it can be said that the voltage magnitudes of two or more unmeasured buses at a time were successfully estimated by the generated data and, not based on the measurements. In addition, cases, that have the maximum values of $\varepsilon_{V,p.u.}$ in Table 4 are described in Figs. 11–13, and the maximum magnitude errors of each target bus with respect to time (hours) are shown. In general, as the uncertainty increases, the maximum error of the estimated voltage magnitude also increases as represented in Fig. 14. As shown in Figs. 11 and 12, the errors of the estimated voltage magnitude of target buses for 13-node network were exceptionally large between 6 and 11 o'clock
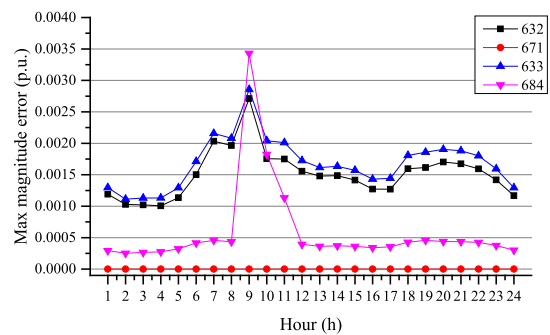
compared to other networks, and its reason will be described in Appendix.

In general, the error in the voltage magnitude estimated by the FFNN is relatively greater than that of the LR and the SVM. This is because the FFNN is optimized for learning the nonlinear features. The FFNN is relatively disadvantageous to learning linear relationships between the input and output because it is the algorithm for learning nonlinear charac-teristics through the activation functions. In addition, the relationship between the voltage and power is nonlinear as represented in Equations (4) and (5); however, the voltage difference becomes smaller when the voltage is expressed

TABLE 5. Error of estimated voltage angle.

Algorithm type	Network type	Maximum values of $\varepsilon_{\theta,deg}(m)$ in degree for all m according to uncertainty, $m = 1, 2, \cdots, M_{tar}$				
		10% uncertainty	20% uncertainty	30% uncertainty	40% uncertainty	50% uncertainty
FFNN	13-node	0.031	0.080	0.176	0.189	0.157
	34-node	0.002	0.003	0.003	0.006	0.008
	37-node	0.012	0.019	0.025	0.034	0.051
LR	13-node	0.030	0.082	0.106	0.146	0.181
	34-node	0.002	0.003	0.003	0.006	0.007
	37-node	0.012	0.017	0.028	0.033	0.053
SVM	13-node	0.032	0.061	0.058	0.109	0.115
	34-node	0.032	0.061	0.058	0.109	0.115
	37-node	0.020	0.029	0.047	0.057	0.076
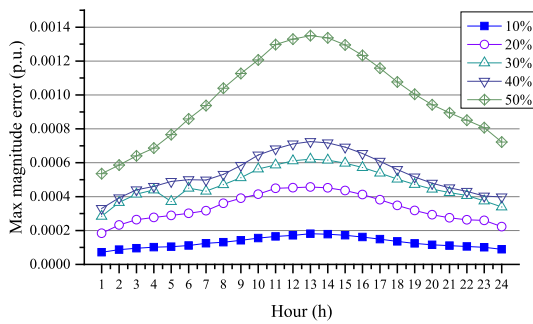
FIGURE 14. Max magnitude error tendency through LR of 733 bus of 37-node Test Feeder.

FIGURE 15. Max magnitude error result through FFNN of 34-node Test Feeder when uncertainty equals 50%.

FIGURE 16. Max magnitude error result through FFNN of 13-node Test Feeder when uncertainty equals 50%.

in p.u. This makes the linear feature stronger, and the resulting error of the estimated voltage magnitude by the FFNN is large relatively compared to that of other algorithms. This argument is supported by the relatively small error in the voltage magnitude estimated by the LR compared to other algorithms.

The methodology proposed in [1] was used to estimate the voltage magnitudes and angles of only two buses out of 95 buses network of the U.K. Generic Distribution System. On the other hand, the number of unmeasured target buses estimated by the proposed approach in this paper is 11 consecutive buses out of 37-node network, which is far higher than that of methodology in [1]. Moreover, the maximum value of percentage error of the estimated voltage magnitude from the methodology in [1] was 0.883% assuming 10% uncertainty in pseudo-measurement, whereas that of the proposed method in this paper was 0.51% and the minimum value was 0.10% for all target buses, where the uncertainty is assumed to be 50%. This implies that not only the number of estimated buses in the proposed approach is much higher than that in [1], but also the accuracy of the estimation is far higher.

C. VOLTAGE ANGLE

Table 5 lists the maximum values of $\varepsilon_{\theta,deg}$ in degrees of the estimated voltage angles for each algorithm, and for each network according to the uncertainty. The maximum value
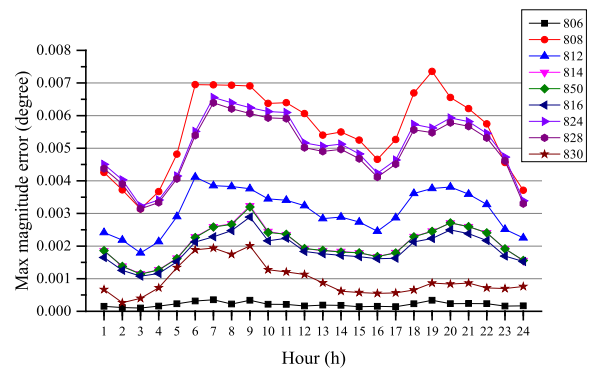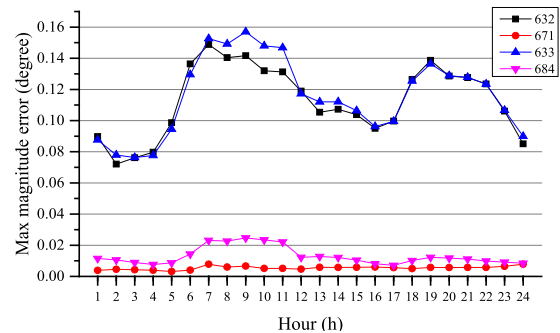
of $\varepsilon_{\theta,deg}$ of the 13-node network is higher than that of the others. For the 34-node network, the error differences in degrees are lower than 0.010° through the FFNN and the LR. The SVM is the best algorithm for the 13-node network. In addition, the greater the uncertainty, the larger the error in common with the result of the estimated voltage magnitude. Figs. 15–20 show the best and the worst cases of the results for each algorithm when the uncertainty equals 50%. As shown in Figs. 16 and 18, the errors of the estimated voltage angle of target buses for 13-node networks were exceptionally large between 6 and 11 and between 18 and 22, and the reason will be described in Appendix.
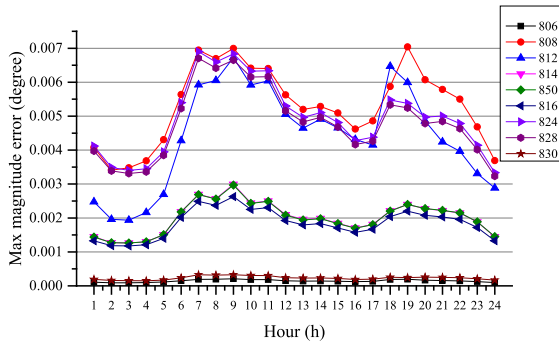
FIGURE 17. Max magnitude error result through LR of 34-node Test Feeder when uncertainty equals 50%.
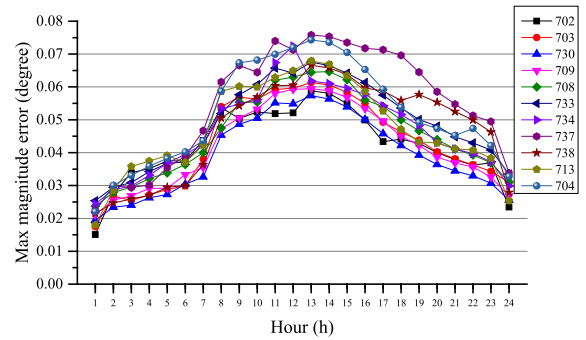

FIGURE 19. Max magnitude error result through SVM of 37-node Test Feeder when uncertainty equals 50%.
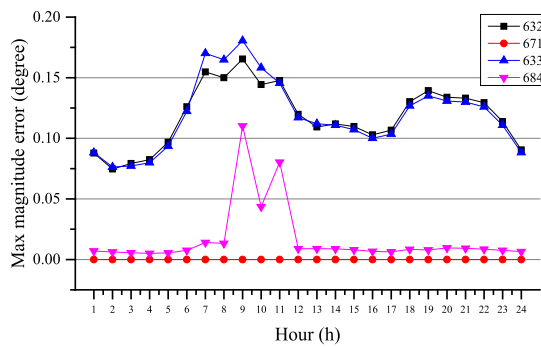

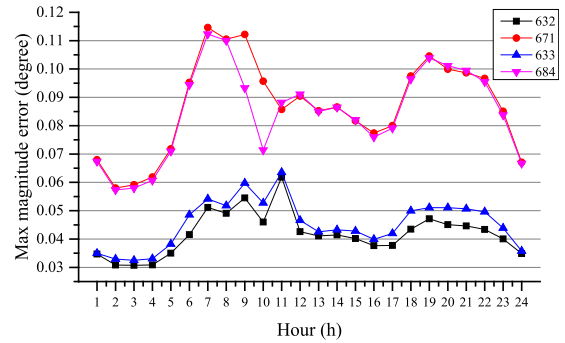**FIGURE 18.** Max magnitude error result through LR of 13-node Test Feeder when uncertainty equals 50%.

FIGURE 20. Max magnitude error result through SVM of 13-node Test Feeder when uncertainty equals 50%.

and $\varepsilon_{\theta,deg}(m)$ for each target bus in the error distribution results shown in Figs. 21–28.

1) 13- AND 34-NODE TEST FEEDER

Figs. 21 and 22 show error distributions of the estimated voltage magnitudes and angles of 13-node network for three supervised learning algorithms according to input buses used among the adjacent buses, respectively. In Figs. 21 and 22, the first row with red letters in used input buses is the result of the study case described before, and the second and third rows show the results of missing one input bus from the original study case. In general, the error was small in the original study case of the first row when the voltage magnitudes and angles were estimated through the FFNN and LR. On the other hand, the error had the largest value
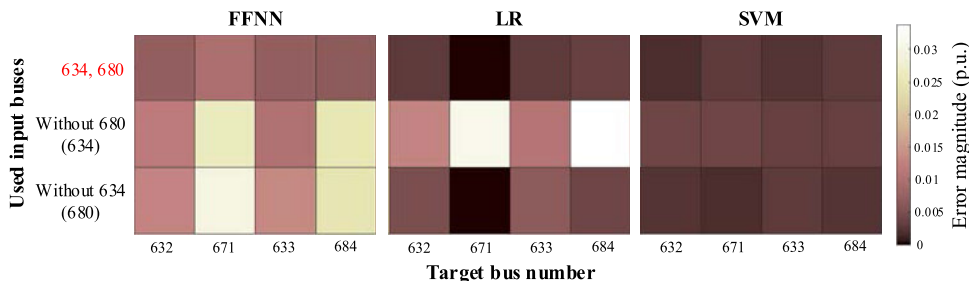
D. IMPACT OF INPUT BUS

The estimated voltage magnitudes and angles of target buses were analyzed when one of the input buses shown in Table 2 was omitted, as there are no additional available adjacent buses of consecutive target buses to use as the input data on selected phase for 13- and 34-node networks. However, the results were further analyzed when additional more adjacent buses as input buses were used to estimate voltage information, as 37-node network had more additional available adjacent buses of consecutive target buses. All of the results described in this section are those with the uncertainty of 50%, and the error magnitudes are all values of $\varepsilon_{V,p.u.}(m)$

FIGURE 21. Error distribution of estimated voltage magnitudes for 13-node Test Feeder over target buses according to input buses.
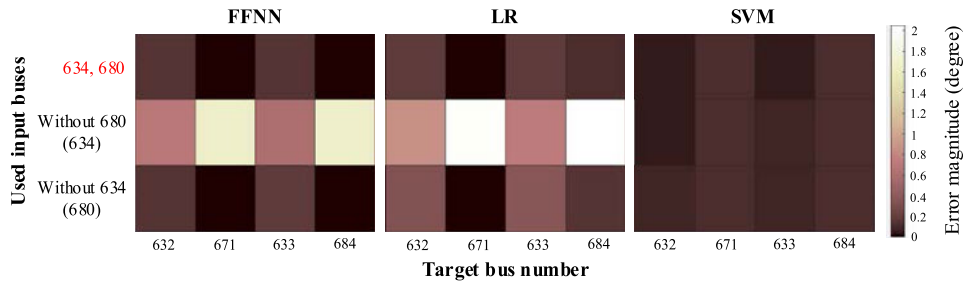
FIGURE 22. Error distribution of estimated voltage angles for 13-node Test Feeder over target buses according to input buses.
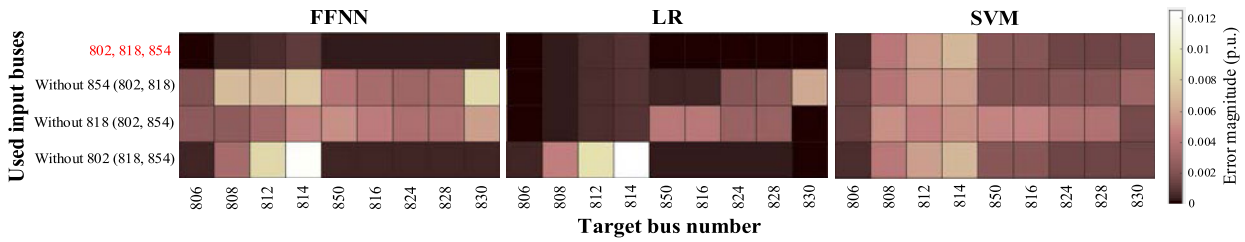

FIGURE 23. Error distribution of estimated voltage magnitudes for 34-node Test Feeder over target buses according to input buses.
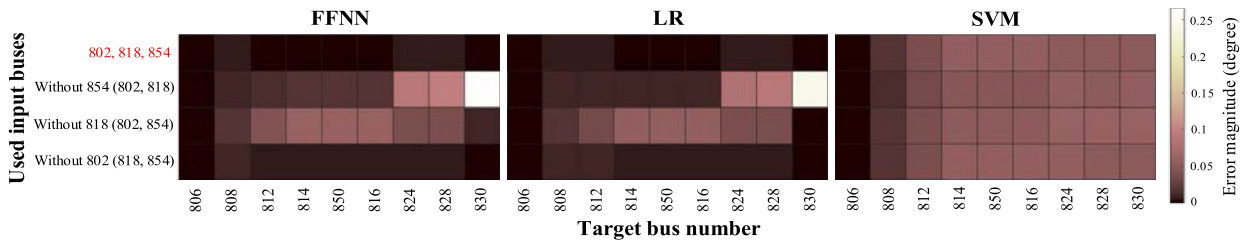

FIGURE 24. Error distribution of estimated voltage angles for 34-node Test Feeder over target buses according to input buses.

when the data of 680 bus was missing. However, the results of estimated voltage magnitudes and angles through the SVM were not significantly affected by the lack of one input bus.

The error distributions of the estimated voltage magnitudes and angles for 34-node network through three supervised learning algorithms are shown in Figs. 23 and 24, respectively, according to the input buses used among the adjacent buses. As shown in Figs. 23 and 24, the results of original study case with red letters in used input buses through the FFNN and LR had small errors relatively as with the 13-node network. In addition, the error, when 818 bus or 854 bus was missing, had generally large value in both estimated voltage magnitudes and angles. However, as with 13-node network, the error of the estimated voltage magnitudes and angles through the SVM were not significantly changed according to the input buses.

In both networks, the error of results through the FFNN and LR was large when the input bus, which is far from the slack bus, was missing. In original study cases in Table 2, the error of result through the SVM was generally larger than the result through other two algorithms. However, it is

advantageous when one of the input buses is omitted as it was not significantly changed by lack of one input bus.

2) 37-NODE TEST FEEDER

Figs. 25 and 26 show error distributions of the estimated voltage magnitudes and angles through three supervised learning algorithms in 37-node network according to used input buses, respectively. Unlike 13- and 34-node network, results when input buses were additionally used in the original study case were also analyzed. In Figs. 25 and 26, red letters are the used input buses in original study case, and first and second row are the results when additional 6 input buses and 3 input buses among adjacent buses were used, respectively. When one of the input buses was missing, it can be seen that the error through the FFNN and LR was larger than the original study case as in 13- and 34-node network, and the error was also prominently large when the input bus, which is far from the slack bus, was missing. When the voltage information was estimated using the additional input buses as shown in first and second row through the FFNN and LR, it was obtained that the more input buses used, the less error in the estimated voltage magnitudes and angles.
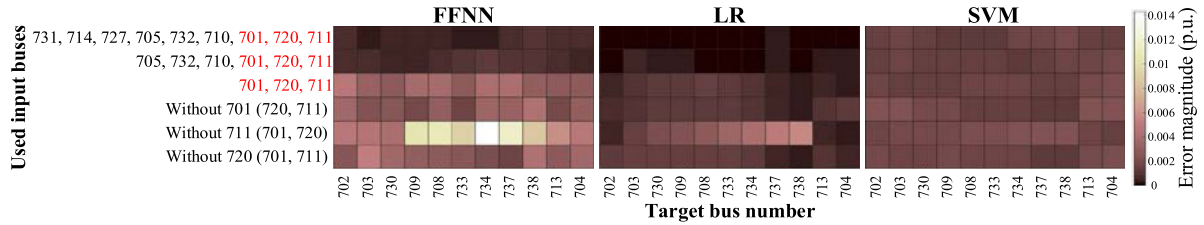
FIGURE 25. Error distribution of estimated voltage magnitudes for 37-node Test Feeder over target buses according to input buses.
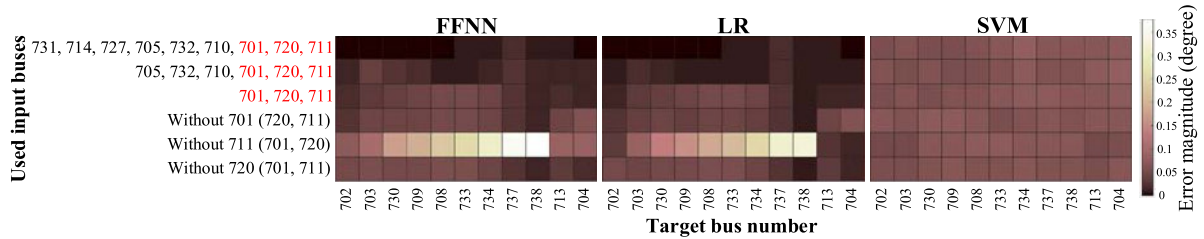

FIGURE 26. Error distribution of estimated voltage angles for 37-node Test Feeder over target buses according to input buses.
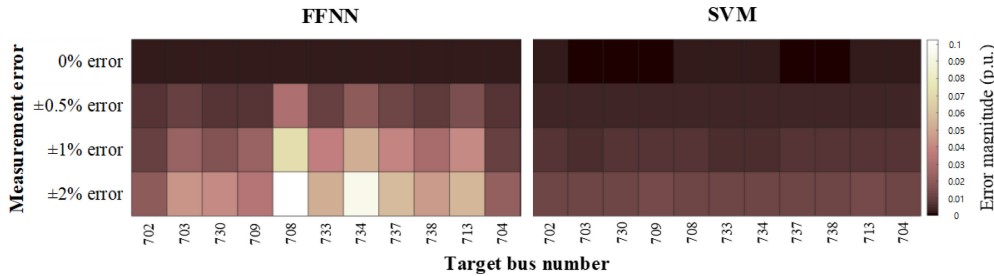

FIGURE 27. Error distribution of estimated voltage magnitudes for 37-node Test Feeder over target buses according to measurement errors.
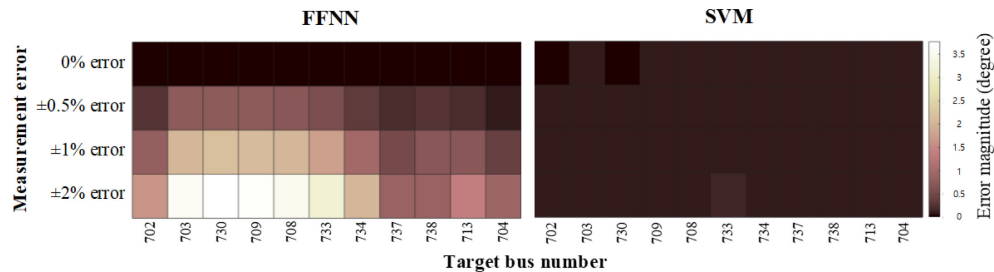

FIGURE 28. Error distribution of estimated voltage angles for 37-node Test Feeder over target buses according to measurement errors.

However, when the voltage magnitudes and angles were estimated through the SVM, there was little change in the results, and there was no significant difference even when more input buses were used. These are the results of well emphasized characteristics of the SVM, which is advantageous when the relationship between the data is not understood.

E. IMPACT OF MEASUREMENTS ACCURACY
There are several guidelines for the accuracy of commercialized measurement devices, however, the criteria are not very different. Thus, in this paper, error of the devices was considered based on the measuring instruments

directive (MID) published in official journal of the EU [32]. In MID, the accuracy of the device is classified into three classes according to maximum percentage error (MPE), with each set MPE of 0.5, 1, 2%. Hence, the error distributions of the estimated voltage magnitudes and angles were analyzed by setting the input data used in the testing step of the supervised learning algorithms to have random measurement errors within the range of ±0.5, ±1 and ±2%.

Figs. 27 and 28 show error distributions of the estimated voltage magnitudes and angles through the FFNN and SVM when there were measurement errors previously assumed for 37-node network. The results through the FFNN and SVM were analyzed because the results through the FFNN
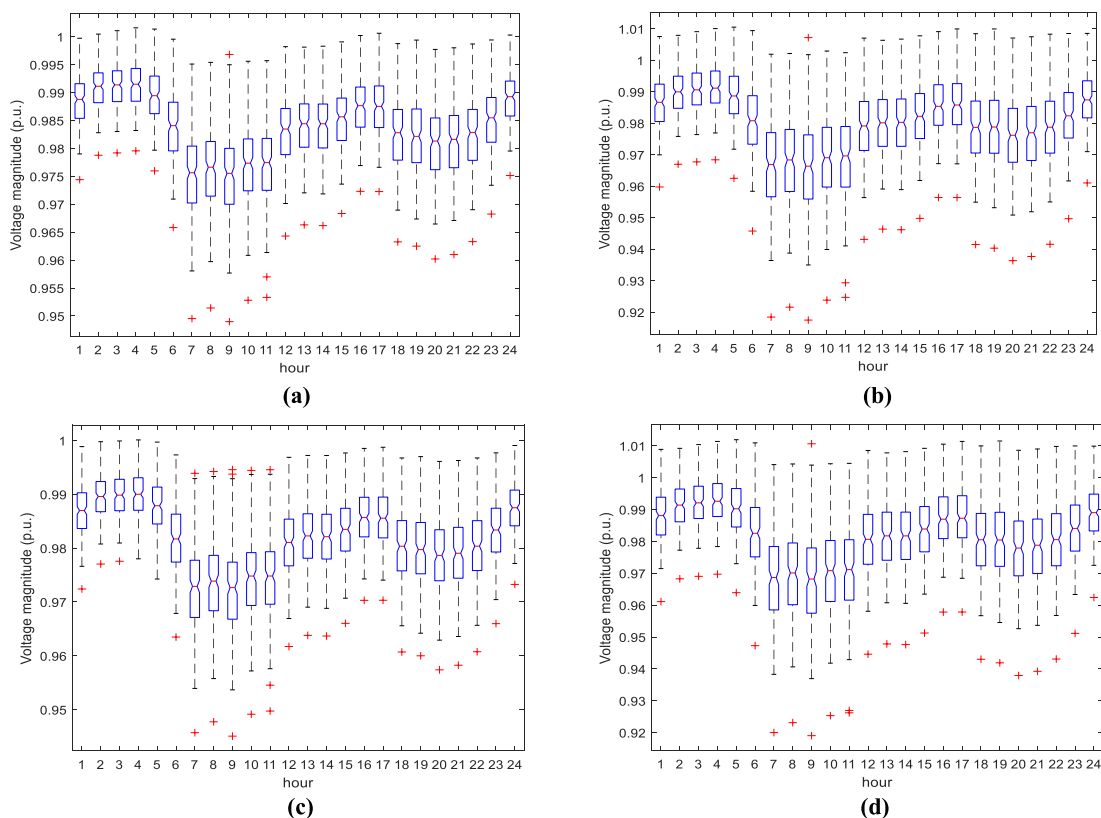
FIGURE 29. ANOVA test results of voltage magnitudes of each target bus for 13-node Test Feeder: (a) 632 bus; (b) 671 bus; (c) 633 bus; (d) 684 bus.

and LR were similar and the results through the SVM were significantly different with the previous results. In addition, since the results of 37-node network were most clearly distinguished, the measurement errors were assumed for 37-node network.

First row of Figs. 27 and 28 is result of error distribution without measurement errors, which is the original study case. Second, third, and fourth rows in Figs. 27 and 28 are error distributions of the results when measurement errors of ± 0.5, ± 1, and $\pm 2\%$ were assumed, respectively. When using the FFNN, error magnitude of both the estimated voltage magnitudes and angles significantly increased as the measurement error range increases. However, with the SVM, error magnitude of the estimated voltage magnitudes and angles was significantly smaller than the results using the FFNN.

V. CONCLUSIONS
Three supervised learning algorithms were used to estimate the voltage magnitudes and angles of the consecutive unmeasured buses, and the representative load profile and the electricity charges of consumers were needed to generate data for training and testing. The data were generated based on an IEEE 13-, 34-, and 37-node test feeders considering the uncertainty. The proposed approach is best suited for LV networks when measurements are not sufficient and the size of the network is not very large.

This approach requires minimal metering devices for measuring voltage information of adjacent buses of the unmeasured buses when applied in practice; however, the measurement data are not required for the training.

In addition, the proposed approach can estimate the voltage magnitudes and angles of several consecutive buses all at once. The maximum absolute magnitude error of the estimated voltage magnitude was 0.01 p.u., and the maximum absolute magnitude error of the estimated voltage angle was 0.189°. These results can be improved by using more input buses among the adjacent buses. The best result of the estimated voltage magnitude was obtained by the LR since the linearity of the voltage magnitude was enhanced by converting to p.u. and a small voltage magnitude difference.

In order to analyze the impact of the adjacent buses used as input buses on the results, the results when one of the input buses in the proposed study cases is missing, and additional input buses are further used were obtained through the FFNN, LR, and SVM. In case of using the FFNN and LR, the error magnitude increased when one input bus was missing. Especially, the error magnitude in the case of missing the input bus, which is far from the slack bus, was larger than in other cases. That is, the proposed approach in this paper is greatly affected by the input bus which is distant from the slack bus. Therefore, this approach cannot be applied to the terminal buses of networks due to lack of adjacent bus data. Although
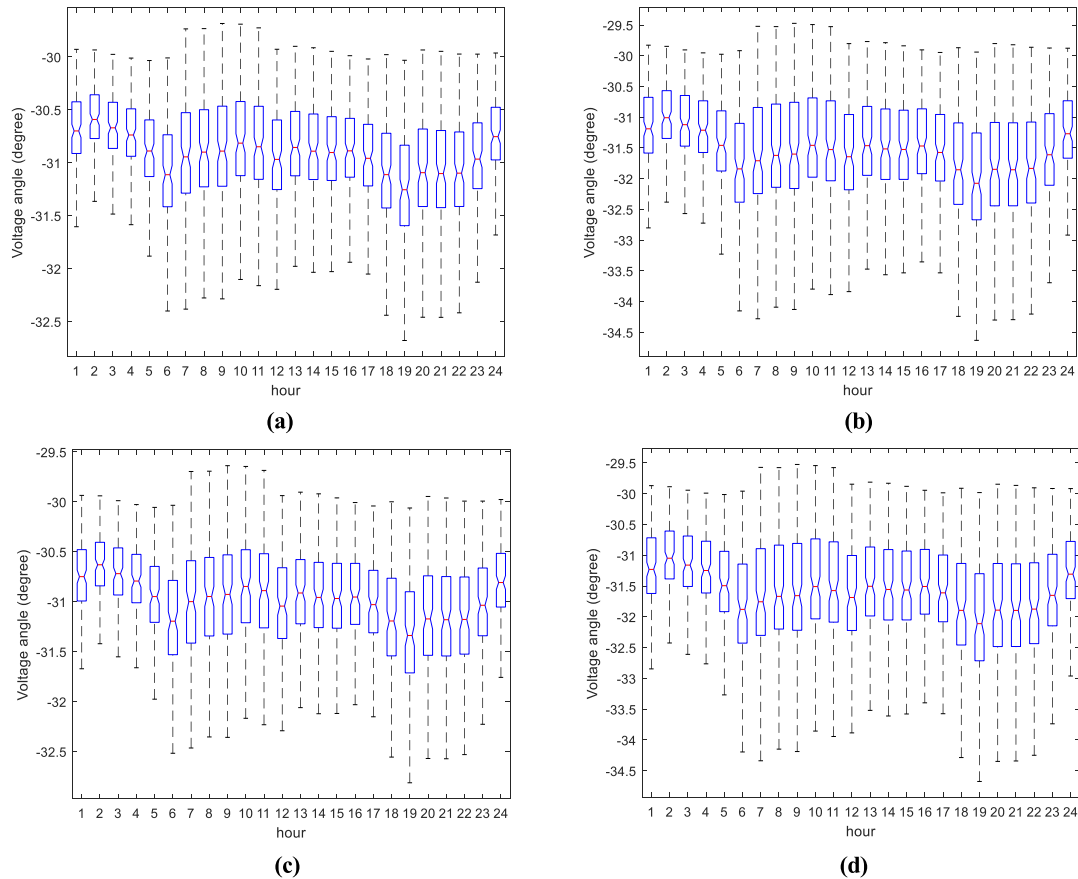
FIGURE 30. ANOVA test results of voltage angles of each target bus for 13-node Test Feeder: (a) 632 bus; (b) 671 bus; (c) 633 bus; (d) 684 bus.

results can be obtained through the proposed approach in case of the terminal bus, the accuracy may be worse than expected. When using the FFNN and LR, it can be seen that the error magnitude decreases as number of input buses used increases more when using the SVM.

On the other hand, results obtained using the SVM presented no significant difference when one of input bus is missing or additional input buses are used. This feature was also shown in results of estimating the voltage magnitudes and angles with assumed measurement errors, and illustrates advantageous characteristics of the SVM when the relationship between inputs and outputs is not well understood. In conclusion, it is advantageous to use the SVM if there are measurement error or if there are few adjacent buses to use as input buses, but there is no way to improve the accuracy because the results using the SVM are not affected by addition of input buses significantly, unlike when using the FFNN or LR.

The proposed approach is not applicable to terminal buses and has the disadvantage that estimating voltage information using the FFNN and LR will be largely affected by measurement errors. In addition, the validity was verified only when the load types of consumers, which make up the network, are the same, and optimization of measurement placement is not carried out in this paper. However, a small number of input

buses were used to estimate the voltage information of the target buses with high accuracy, and this feature can reduce technical costs. Although there are still data missing occurs from the smart meter, valid results can be obtained even if one of input buses is missing or there are measurement errors. Thus, these results can be advantageous over conventional methods.

The validity of the proposed approach was verified only for PQ buses with no generators and only loads. Future works will be carried out on networks that include PV buses with generators, with load forecasting and generation forecasting.

APPENDIX
ANOVA (analysis of variance) test was conducted to analyze the cause of errors in the estimated voltage magnitudes and angles using MATLAB. Figs. 29(a–d) show ANOVA test results of V_{tar} of 13-node network, such as 632, 671, 633, 684 buses according to time when the uncertainty equals 50%, respectively. On each blue box, the central mark is the median and the edges of the box are the 25th and 75th percentiles. The black broken lines extend to the most extreme data points that are not considered outliers. The outliers are plotted individually using the red '+' symbol.

As shown in Figs. 29(a–d), for all target buses, the distributions of voltage magnitudes, which are the length of

broken lines, between 6 and 11 were larger than other times. In addition, the outliers had relatively large differences from the medians compared to other times. In other words, the error of the estimated voltage magnitude at a time when the variance of the voltage magnitude was large tended to be large. This feature was shown not only in the results of the voltage magnitudes of target buses but also in the results of the voltage angles of target buses, which are the elements of θ_{tar}, for 13-node network when the uncertainty equals 50% in Figs. 30(a–d).

As shown in Figs. 30(a–d), the distribution of the voltage angles of target buses between 6 to 11 and between 18 to 22, which had large errors in the estimated voltage angles, can be seen to be wider compared to other times. These results support the rationale that the larger the distribution of voltage magnitude and angle of the target bus, the larger errors of the estimated voltage magnitude and angle.

REFERENCES

[1] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1888–1896, Nov. 2012.

[2] A. Angioni, T. Schlosser, F. Ponci, and A. Monti, "Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 70–77, Jan. 2016.

[3] S. Bhela, V. Kekatos, and S. Veeramachaneni, "Enhancing observability in distribution grids using smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5953–5961, Nov. 2018.

[4] B. Brinkmann and M. Negnevitsky, "A probabilistic approach to observ-ability of distribution networks," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1169–1178, Mar. 2017.

[5] R. Singh, B. C. Pal, and R. B. Vinter, "Measurement placement in distri-bution system state estimation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 668–675, May 2009.

[6] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2919–2928, Nov. 2015.

[7] S. Ryu, M. Kim, and H. Kim, "Denoising autoencoder-based missing value imputation for smart meters," *IEEE Access*, vol. 8, pp. 40656–40666, 2020.

[8] N. Etherden, A. K. Johansson, U. Ysberg, K. Kvamme, D. Pampliega, and C. Dryden, "Enhanced LV supervision by combining data from meter, secondary substation measurements and medium voltage supervisory con-trol and data acquisition," in *Proc. 24th Int. Conf. Exhib. Electr. Distrib. (CIRED)*, Jan. 2017, pp. 1089–1093.

[9] R. Bush, "Predicting distribution transformer failures," *T&D World*, vol. 70, no. 7, pp. 24–28, Jul. 2018.

[10] H. Wang, W. Zhang, and Y. Liu, "A robust measurement placement method for active distribution system state estimation considering network reconfiguration," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2108–2117, May 2018.

[11] A. Bernieri, G. Betta, C. Liguori, and A. Losi, "Neural networks and pseudo-measurements for real-time monitoring of distribution systems," *IEEE Trans. Instrum. Meas.*, vol. 45, no. 2, pp. 645–650, Apr. 1996.

[12] H. Wang and N. N. Schulz, "A revised branch current-based distribution system state estimation algorithm and meter placement impact," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 207–213, Feb. 2004.

[13] Y. R. Gahrooei, A. Khodabakhshian, and R.-A. Hooshmand, "A new pseudo load profile determination approach in low voltage distribution networks," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 463–472, Jan. 2018.

[14] A. Deihimi and A. Rahmani, "Application of echo state network for har-monic detection in distribution networks," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 5, pp. 1094–1101, Mar. 2017.

[15] A. Deihimi and A. Rahmani, "Application of echo state networks for estimating voltage harmonic waveforms in power systems considering a photovoltaic system," *IET Renew. Power Gener.*, vol. 11, no. 13, pp. 1688–1694, Nov. 2017.

[16] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4796–4805, Nov. 2019.

[17] K. Dehghanpour, Y. Yuan, Z. Wang, and F. Bu, "A game-theoretic data-driven approach for pseudo-measurement generation in distribution system state estimation," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 5942–5951, Nov. 2019.

[18] *IEEE 13 Node Test Feeder*. Accessed: Feb. 2020. [Online]. Available: https://site.ieee.org/pes-testfeeders/resources/

[19] *IEEE 34 Node Test Feeder*. Accessed: Feb. 2020. [Online]. Available: https://site.ieee.org/pes-testfeeders/resources/

[20] *IEEE 37 Node Test Feeder*. Accessed: Feb. 2020. [Online]. Available: https://site.ieee.org/pes-testfeeders/resources/

[21] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load modeling—A review," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5986–5999, Nov. 2018.

[22] J. A. Jardini, C. M. V. Tahan, M. R. Gouvea, S. U. Ahn, and F. M. Figueiredo, "Daily load profiles for residential, commercial and industrial low voltage consumers," *IEEE Trans. Power Del.*, vol. 15, no. 1, pp. 375–380, Jan. 2000.

[23] I. P. Panapakidis, M. C. Alexiadis, and G. K. Papagiannis, "Electricity customer characterization based on different representative load curves," in *Proc. 9th Int. Conf. Eur. Energy Market*, Florence, Italy, 2012. pp. 1–8.

[24] D. M. Tagare, *Reactive Power Management*. New Delhi, India: McGraw-Hill, 2007.

[25] G. Bebis and M. Georgiopoulos, "Feed-forward neural networks," *IEEE Potentials*, vol. 13, no. 4, pp. 27–31, Nov. 1994.

[26] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometric Intell. Lab. Syst.*, vol. 39, no. 1, pp. 43–62, Nov. 1997.

[27] X. Yan and X. G. Su, *Linear Regression Analysis: Theory and Computing*. Singapore: World Scientific, 2009.

[28] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Netw.*, vol. 12, no. 6, pp. 783–789, Jul. 1999.

[29] L. Auria and R. A. Moro, "Support vector machines (SVM) as a technique for solvency analysis," Deutsches Inst. Wirtschaftsforschung, Berlin, Germany, DIW Berlin Discussion Paper 811, 2009. [Online]. Available: https://ssrn.com/abstract=1424949

[30] *MATLAB Deep Learning Toolbox 12, User's Guide*. Accessed: Mar. 2020. [Online]. Available: http://www.mathworks.com

[31] *MATLAB Statistics and Machine Learning Toolbox 11, User's Guide*. Accessed: Mar. 2020. [Online]. Available: http://www.mathworks.com

[32] "Directive 2004/22/EC of the European parliament and of the council of 31 March 2004 on measuring instruments," *Off. J. Eur. Union*, pp. 1–80, Apr. 2004.

GWANGPYO HONG was born in Seoul, South Korea, in 1994. He received the B.S. degree in electrical engineering and computer science from the Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, where he is cur-rently pursuing the M.S. degree with the School of Integrated Technology.

YUN-SU KIM (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2010 and 2016, respectively.

From 2015 to 2017, he worked as a Senior Researcher with the Korea Electrotechnology Research Institute (KERI). He joined the Faculty of the Gwangju Institute of Science and Technol-ogy (GIST), in 2018, where he is currently an Assistant Professor with the School of Integrated Technology. He is also the Director of the Korean Society for New and Renewable Energy and the Korean Institute of Electrical Engineers. His research interests include distribution networks, distributed energy resources, microgrid, artificial intelligence, and wireless power transfer.

● ● ●