PAC World Conference 2024 - Athens, Greece, June 17-20, 2024

Machine learning for distribution grid topology identification and state estimation

Haris Sapountzakis (el17174@mail.ntua.gr)
Kyriakos Andresakis (KyAnd@power.ece.ntua.gr)
Themistoklis Xygkis (txygkis@power.ece.ntua.gr)
Aris Dimeas (adimeas@power.ece.ntua.gr)
Georgios Korres (gkorres@cs.ntua.gr)
Nikos Hatziargyriou (nh@power.ece.ntua.gr)
National Technical University of Athens (NTUA)
Athens, Greece

Abstract

The ongoing integration of distributed energy resources and customer engagement in DNs render their operation condition increasingly dynamic. At the same time, fast updates of the current network operating state are essential for real-time monitoring and control functions. In this challenging environment, precise TI and DSSE, that is, processing network measurements to estimate the real-time grid configuration and state (e.g., nodal complex voltages), are imperative for extended DN visibility. Typically, the lack of observability owing to the limited availability of measurements, impedes the application of conventional, model-based DSSE methods. Lately, the introduction of SMDs in DNs, can reinforce the available measurement dataset. In this scope, ML is gradually being used to develop reliable TI and DSSE methodologies under limited SMD availability. The current paper aims to provide a complete framework for effective TI and DSSE, mainly focusing on the employment of dimensionality reduction and feature selection techniques. In particular, it investigates the use of RF and LASSO for the determination of minimum sets of SMDs which can be used to perform quality TI and DSSE based on DNNs. Numerical studies on the IEEE-33 bus benchmark test system, using the proposed ML techniques to train and evaluate the DNNs, are conducted. The efficacy and the computational speed of the proposed techniques, is compared to previous works yielding promising results.

List of abbreviations

Abbreviation	Description	Abbreviation	Description
DN	Distribution network	SE	State estimation
TI	Topology identification	WLS	Weighted least squares
DSSE	Distribution system state estimation	SCADA	Supervisory control and data acquisition
SMD	Synchronized measurement devices	MMSE	Minimum mean square error
ML	Machine learning	DL	Deep learning
RF	Random forest	DT	Decision tree
LASSO	Least absolute shrinkage selection operator	SFS	Sequential forward selection
DNN	Deep neural network	HC	Hierarchical clustering
NN	Neural network	MAPE	Mean absolute percentage error
AMI	Advanced metering infrastructure	MAE	Mean absolute error
RES	Renewable energy sources	MSE	Mean square error
IED	Intelligent electronic device	TVE	Total vector error
DG	Distributed generation	PCA	Principal component analysis

1. Introduction

Gradually, modern DNs are being transformed from traditionally passive conduits into smart grids that integrate digital technologies, such as AMI, synchrophasor measurement systems, control devices (e.g. IED), and DG including RES to optimize the production, distribution and consumption of electricity [1]. This transition to more active DNs necessitates real-time grid monitoring and analysis, to ensure efficient management of the multiple energy resources and to enhance their reliability and adaptability.

In this framework, DSSE is a crucial process for real-time monitoring [2], providing the operators with vital insights into system behavior by processing the measurement data gathered from the DN in order to: a) estimate the grid operating state mostly defined by the vector of nodal complex voltages, which are then used to compute the current and power flows at all network branches, and b) to filter out the measurement errors, thus, improving confidence in the available data [3]. DSSE also integrates TI capabilities, that is, processing the information about the statuses of network circuit breakers and tie-switches along with the available measurements to deliver the operating grid configuration [3]. Overall, the output of DSSE comprises a vital input for numerous distribution management applications related to real-time grid operation and control, such as volt-var optimization, fault detection, transmission-distribution interface coordination, cybersecurity etc., and planning tasks, such as proactive decision-making, contingency analysis, support of energy markets, forecasting etc. [4].

SE is, essentially, the numerical process of assigning values to unknown (missing) system parameters, i.e., state variables, based on measurement values [5]. To this end, estimators that seek to approximate the unknown state variables in an optimal way, considering the degree of measurement uncertainty, are constructed. Conventional DSSE techniques, such as the well-established WLS method, employ a deterministic approach, where maximum likelihood estimation and power flow equations to build the measurement model are utilized [6]. A serious impediment for the real-world application of this class of estimators in DNs is that, contrarily to transmission systems, the real-time measurements are limited given the huge number of the points of interest and heterogeneous, e.g., obtained from SCADA at primary substations, smart metering at customer premises etc. [7], [8]. Hence, unless pseudo-measurements. i.e., artificially generated data, are used (to the detriment of DSSE performance), the SE problem is unsolvable and the network is said to be non-observable [3].

The problem of low-observability can be circumvented by studying DSSE in a Bayesian framework [7]–[10]. Unlike the WLS method which minimizes the modelling error embedded through the measurement model, a Bayesian approach leads to the construction of MMSE estimators, which directly minimize the estimation error, i.e., the error between the actual and estimated state vector [7]. In this way, the need for using a measurement function is bypassed and, thus, the necessary conditions for observability, too. Given that a closed-form SE solution is unattainable, DNNs can be used to approximate a MMSE estimator [7]. Indeed, NNs have been proven to be a highly powerful tool to approximate any function given a sufficiently large amount of data for them to be trained on, especially non-linear ones. Hence, DNNs can be designed and trained via data driven techniques to construct a MMSE estimator for the purpose of DSSE [7]–[10]. Moreover, DL can be deployed for performing TI [10], [11], [12].

Importantly, DL approaches are strongly affected by the measurement data used as their input [7], [10], [13]. Therefore, the development of data driven, ML techniques for optimal measurement selection is highly stimulated in an effort to effectively train and implement DL methodologies for real-time TI and DSSE [10], [12]. In this respect, SMDs which can provide highly accurate, time-synchronized voltage and current phasor data in real time, are on the focus of ML based optimal measurement selection techniques for various applications including TI and DSSE [10], [14]. It is notable that, since the need for observability as conceptualized in conventional DSSE is eliminated, it is possible to achieve quality DSSE by means of a limited number of strategically placed SMDs.

In light of the above considerations, ML based optimal measurement selection is a rather promising and, currently, under-explored field compared to the related works regarding conventional DSSE such as in [15]. Hence, motivated by [10], this paper focuses on the development of ML techniques, derived from the fields of dimensionality reduction and feature selection, for optimal selection of SMDs in order to perform quality DNN based TI and DSSE. More specifically, following the implementation of the ML techniques proposed in [10], additional ML techniques, namely, the RF and LASSO algorithms, are designed and developed with a view to selecting the optimal locations of SMD installations in terms of TI and DSSE performance. In the sequel, the data from the optimal (minimum) SMD sets provided by the studied techniques, are leveraged to train 2 individual DNNs which operate in a sequential manner; following the TI-DNN which yields the current network topology, the grid state is subsequently estimated by the DSSE-DNN. To generate the training dataset, a scheme based on Monte Carlo trials is implemented. Moreover, transfer learning is applied to account for the changes in topology while performing DSSE, so that a single DNN is used and a need for multiple trainings and hyperparameter tunings do not arise.

Concisely, the paper is structured as follows: in Section 2, an overview of ML applications to power systems is provided, while the design and development of the studied ML techniques for optimal SMD placement are detailed in Section 3. The results from numerical studies on the benchmark IEEE-33 bus test system [16] to investigate the performance of DNN-based TI and DSSE given the different SMD configurations provided by the individual ML techniques, are cited in Section 4. The conclusions and future implications of the work, are discussed in Section 5 and 6.

2. ML applications to power systems

ML is a sub-field of artificial intelligence that allows computational systems to learn from experience and, thus, improve on specific, predetermined tasks. This kind of systems perform training iterations over vast amounts of data and "learn" patterns hidden within these. This allows them to perform classification, regression or various others tasks on new "unseen" data, by gaining experience from the similar data they were trained on.

Power systems, like several other fields of study, have long benefited from the introduction of ML models into their applications and research. Typical examples include prediction of RES production [17], [18] and, lately, real-time data management and analytics in energy control centers [1]. Especially, the utilization of NNs and, particularly, DNNs can yield exceptional results.

2.1 Fundamentals of NNs and DNNs

A standard NN is composed of various types of layers, each containing its own parameters. Its typical structure is provided in Fig. 1. These parameters are updated during the training process for the purpose of minimizing the error between the target and the predicted data. Typically, each NN structure is composed of an input layer that receives the input data. The next layers are the hidden layers and can vary in number. In these layers, the computations that occur allow the NN to identify the problem-specific patterns and relationships between the input data. The output layer produces the final output, depending on the specific problem. Adjacent layers are linked between them through weights. In a simple NN structure all neurons (nodes) of the previous layer are linked with all the neurons of the next layer.

Neurons are the building blocks of a NN layer and they are responsible for processing information. Each one receives the weighted sum of all the neurons of the previous layer, adds it to its bias (additional parameter of the neuron) and utilizes it as input into an activation function. The output of the activation function is passed to the neurons of the next layer.

The NN learns the input data patterns for the means of predicting the output with a process called training. The input data is entered into the model and the signal of information is passed forward throughout the network from one layer to the next. The predicted output is calculated and compared to the target data. Once the error is estimated, the weights and biases of the neurons are updated. The signal travels back updating all neurons of layers in reverse order, through a process called backpropagation. The latter involves computing the gradients of the loss function with respect to each weight and bias in the network using the chain rule of calculus.

The difference between a simple NN and a DNN is the depth of the architecture. For DNNs, the number of hidden layers is multiple (typically 1 or more). This allows these structures to identify more complex patterns hidden within the input data. In practical terms, DNNs often have several hidden layers, sometimes ranging from 5 to hundreds or even thousands of layers, depending on the specific architecture and task. These deep architectures enable the network to learn complex and hierarchical representations of data, capturing intricate patterns and relationships.

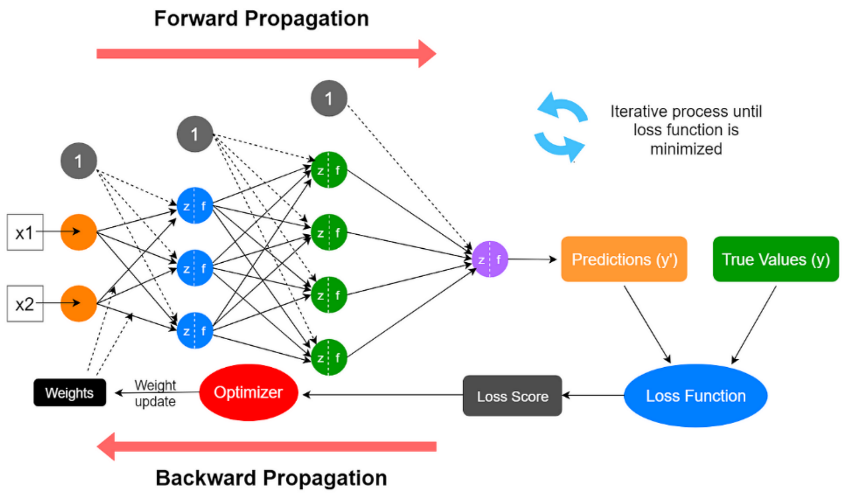

Fig. 1. Overview of learning process of a NN [19].

2.1.1 Use of DNNs for the TI and DSSE problem

As discussed in the Section 1, DL can comprise an effective approach to model and solve TI and DSSE tasks in a Bayesian framework. This means that the DNNs can be used to approximate a MMSE estimator, thus, targeting to minimize the error between the estimated and actual grid state which is quantified by the following metric [7]:

$$\min_{\hat{x}} E\|x - \hat{x}(z)\|^2 \tag{1}$$

where x ($\hat{x}$) is the actual (estimated) grid state vector, z comprises the available measurement set and $E(\cdot)$ is the mean value operator.

In this context, observability as perceived in conventional DSSE methods is not required; no measurement model is utilized, thus, the need for an over-determined system, i.e., the measurements to outnumber the state variables, is eliminated. As a result, it is feasible to solve the TI and DSSE problems by deploying a minimized set of measurements. Synchrophasors obtained by optimally placed SMDs appear to be a suitable solution for this purpose due to their high precision and time synchronization [10], [14]. In case of their exclusive use, the problem of measurement heterogeneity no longer exists.

In general, the TI task can be conceptualized as a classification problem where each topology is treated as a different class [10]–[12], whereas the DSSE task can be treated as a regression problem with the nodal complex voltages being considered as the grid state variables [7]–[10]. All the above works corroborate the accuracy of the DL based methodologies, which relies on a carefully designed and executed training process, as well as the fast online computation times.

2.2 ML techniques for optimal measurement selection

Aiming at selecting optimal measurement sets for DL based TI and DSSE, ML techniques have already been proposed in [10] and [13]. In [10], sequential forward selection and hierarchical clustering are employed to determine optimal sets of SMDs, whereas partial correlation and minimal redundancy maximum relevance are the techniques utilized in [13].

For the scope of this paper, two individual ML models are designed and developed focusing on optimal SMD placement. The former is based on the RF algorithm and the latter deploys the LASSO technique. RF is used in [20] in order to select the optimal set of features to be inputted into the TI-DNN and DSSE-DNN. In this work, the RF is used for identifying the importance of each SMD metric among a set of candidate ones in terms of TI and DSSE accuracy, acting as a preprocessing step for feature selection and dimensionality reduction. Moreover, LASSO, whose use is demonstrated in [21], is also utilized in this paper as an alternative method to the RF.

2.2.1 Fundamentals of RF

RF is a powerful ensemble learning method used for classification and regression tasks. It operates by constructing a multitude of DTs during training and outputs the mode (classification) or mean prediction (regression) of the individual trees, as depicted in Fig. 2. Each DT in the forest is built using a random subset of the training data and a random subset of features. It is often used as a standalone model for classification and regression tasks due to its robustness and ability to handle large datasets with high dimensionality. However, it can also be used as a preprocessing step mainly because of the dimensionality reduction and feature importance. RF can be used to assess the importance of different features in the dataset. Features with higher importance scores are considered more informative and may be selected for further analysis, while less important features can be discarded as part of feature selection. Also, by ranking features based on their importance scores, RF can be used to perform dimensionality reduction by selecting only the most informative features for subsequent modelling. This can help to reduce the computational complexity and improve the performance of other algorithms.

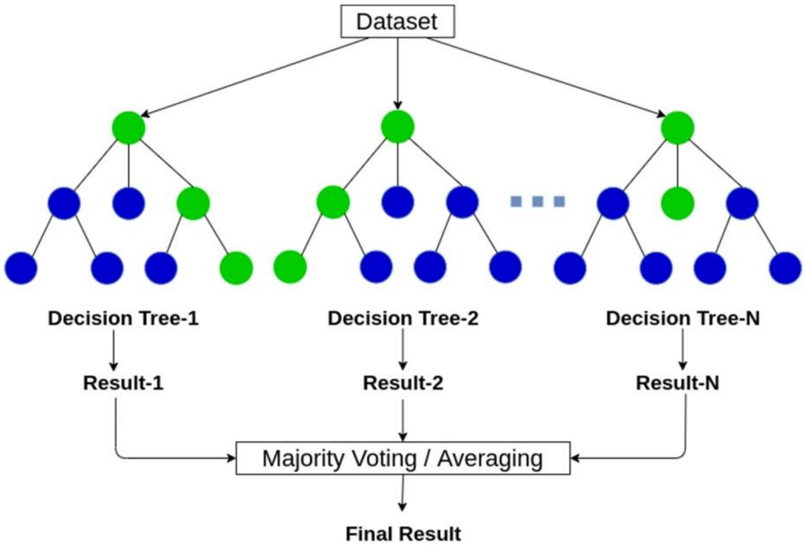

Fig. 2. Operation diagram of the RF algorithm [22].

2.2.2 Fundamentals of LASSO

LASSO is a regularization technique used in linear regression to reduce overfitting and perform feature selection. However, LASSO can also be used as a preprocessing step in ML pipelines to select the most relevant features before training a model. The technique imposes a penalty on the absolute magnitude of the coefficients of the regression model. This penalty encourages sparse solutions, meaning that many coefficients are driven to zero, effectively performing feature selection. Features with non-zero coefficients after applying LASSO regularization are considered important or relevant to the prediction task. By adjusting the regularization parameter, the degree of sparsity in the resulting model can increase or decrease. A higher regularization parameter leads to more coefficients being pushed to zero, resulting in a sparser model with fewer features. Thus, above property allows for smaller input sizes and choice of more relevant input features.

3. Model design and development

To develop the RF and LASSO techniques and the DNNs for TI and DSSE, supervised learning is performed with the dataset consisting of input data samples and their corresponding output data samples. For proper training of the DNNs, a large amount of data is generated based on a thoroughly designed process. Once the dataset is available, the feature selection process pertaining to the SMD placement task, is implemented. The training of the TI and DSSE DNNs and hyperparameter optimization is then conducted. Finally, practices from transfer learning are also applied.

3.1. Dataset generation

The generation of the dataset of measurements for the purpose of SMD placement is carried out using a Monte Carlo trial scheme, that can be executed considering multiple operating topologies for the grid under study.

For the consideration of different load scenarios, the following algorithm is devised:

1. The active power loads of the buses obtained from the network database (nominal or average values) are set as the base values for calculations.
2. Every bus is assigned a random number between 0 and 1 (using a uniform distribution).

3. In case the number assigned to the bus is greater than 0.5, its active power load is obtained by assuming a normal distribution with mean value equal to the corresponding base value (step 1) and standard deviation of 20% of the mean value.
4. The above steps are repeated for the reactive power loads of the buses.

The above steps are repeated 1000 times for each topology. Each time, after the random updates of the loading conditions are performed, power flow analysis is carried out to acquire the nodal voltage and branch current phasors. In the sequel, random noise following normal distribution is added to the above phasor data, in order to simulate the actual measurement data provided by SMDs.

Overall, the procedure described guarantees the generation of a large enough dataset including a multitude of loading scenarios. A key feature is the inclusion of step 2, which contributes to extra variability. Adding errors to model actual measurements provides the ML algorithms with training data that better describe real-world operating conditions. Through this step, the DNNs to be trained also become robust to real measurement errors. It is noted that the error size utilized might be smaller when compared to real-life scenarios, yet, it is deemed acceptable with a view to demonstrating the effectiveness of the proposed ML techniques.

3.2. Feature selection process

In this step of the process, the SMD locations for TI and DSSE purposes need to be identified. The methodology used in [10] is implemented for reasons of comparison with the proposed methodology. In general, the SMD placement amounts to a feature selection process based on an iterative ML algorithm. A single feature, that is, an SMD measurement pertaining to a voltage/current phasor, is chosen from the feature set per iteration; the selected feature is the one maximizing the evaluation metric of the corresponding DNN using it as input. The algorithm is terminated if the desired level of accuracy (TI or DSSE) is reached by means of the SMD units placed. It is noted that no budget limit is considered for the SMD installation.

3.2.1 TI feature selection – SMD placement

In order to perform the SMD placement to achieve the desired TI accuracy, SFS in cooperation with DNN is utilized in [10]. The SFS is a time-consuming process, since at each step of the above algorithm all SMD measurements (input features) require testing. At every step, in order to conclude on the optimal input feature, all remaining features require examination. Another DNN is trained from scratch, while hyperparameter optimization is repeated.

The above process can be improved by replacing the SFS algorithm with the RF algorithm. An RF model can be utilized for the means of classifying the topology using the SMD measurements as input. Instead of the estimation results, the importances of the different input features, as calculated by the RF algorithm, are extracted. By sorting the input features in descending importance order, instead of examining all of them per iteration step, a single one is chosen in advance. At each iteration the next input feature, in descending importance metric order, is used until the above termination criteria are met.

The importance metric is determined by the multiple DTs that contribute to the estimation. Each DT calculates the importance of a single feature by calculating its contribution at reducing the impurity of the estimation and, thus, providing more information for the output. All DT importances of the RF are averaged and a single list with the final importance metrics is calculated. Because of the above, the descending importance metric choice of features per iteration step guarantees that the feature minimizing the estimation error is chosen.

As in [10], the feature selection process leads to dimensionality reduction, because the optimal subset of input features is chosen, instead of the whole set.

3.2.2 DSSE feature selection – SMD placement

The next step in [10] is the identification of additional SMD installations to attain the desired DSSE accuracy. For the purpose of this, a spearman correlation matrix from the voltage measurements is generated and HC in cooperation with DNN is performed on its data, in order to conclude on new SMD locations. However, since the spearman correlation coefficient requires calculation between every bus of the network, increased time is likely to be consumed for the purpose of performing the HC, similar to the TI process described above.

To improve this procedure, the proposed methodology is to replace the HC on the spearman correlation matrix with an RF model and, alternatively, a LASSO algorithm. RF importances are calculated as above to extract the descending importance metric list. At each iteration, a subsequent feature is extracted from the list to improve each new corresponding DNN performance until the evaluation metric threshold is met. Alternatively, LASSO is utilized to generate a descending importance list, based on the coefficients it generates during training. The value of each coefficient determines the contribution of each input feature to the final output, effectively providing a sorted importance list, as it is the case with the RF importances. Both of the above algorithms perform dimensionality reduction again, since they reduce the input feature space.

3.3 Training process

For the training process, multiple models are developed. DNNs are implemented for TI and DSSE tasks, respectively. Hyperparameter optimization is performed for both. At the core of each SMD placement algorithm, additional DNNs are trained and evaluated. Overall, multiple DNN trainings are carried out in order to implement the SFS, HC, RF and LASSO techniques. For all training cases, the train-validation-test ratio is 80-10-10.

3.3.1. TI DNN

A classification DNN is built, trained and evaluated for the purpose of TI. The measurements from the optimally placed SMDs are utilized as input to track the switch statuses. As far as its architecture is concerned, the number of neurons in the output layer is equal to the number of feasible topologies of the network. The number of hidden layers and neurons per layer are hyperparameters that require tuning. The ReLU activation function, is used in the input and hidden layers. For the output layer, the softmax[10] activation function is used [10]. While training the model, the categorical crossentropy[10] is set as the loss function and the optimizer utilized is adam [10], [20]. Its learning rate is experimented upon. Lastly, dropout layers are also introduced between the hidden layers to avoid overfitting. The model is evaluated based on its estimations on the test set, using the following metric:

$$\mathrm{TI}_{\mathrm{acc}} = \frac{\text{number of correctly estimated topologies}}{\text{total number of topologies}} 100\% \tag{2}$$

For hyperparameter optimization, the following hyperparameters are investigated: the number of hidden layers ranged from 1 to 5. The number of neurons in the hidden layers also range from 32 to 2048. The number of epochs during training also is experimented upon, starting from 10 to 200. Four batch sizes of 16, 32, 64 and 128 are used and learning rates from 0.0001 to 0.001 are searched. In addition, dropout rate of dropout layers ranges from 0.2 to 0.4. Lastly, increasing or decreasing the number of neurons per hidden layer is also examined.

There are various techniques for tuning the hyperparameters, namely grid search, random search and more complex methods such as Bayesian optimization. In this paper, a combination of random and grid searches is implemented. First random combinations are examined and then, a number of grid searches are executed. The latter guarantees optimal results within the specified search grid, since it systematically explores all possible combinations.

3.3.2. DSSE DNN

The DNN developed at this stage receives 3 different inputs. Initially it is trained using the current phasor data selected previously from the SMD placement for TI. After this initial training, the same DNN is retrained using additional voltage phasor data from the SMD placement for DSSE. The input per different topology changes, since new branches are active and their corresponding SMD measurements need to be taken into account.

The above regression model could be swapped with 2 separate DNN models, the first estimating the voltage phase angles and the latter one the voltage magnitudes. However, this approach would cause more computational burden, as it would require 2 DNNs to be implemented and, then, appropriately tuned. Thus, a model which splits into two symmetrical branches, is developed.

The number of neurons in the output layer of each branch is equal to the number of buses on the system, in order to effectively estimate the state variables of each bus. The linear function is chosen as the activation function per output branch. The activation functions for the hidden layers are the same as the ones also used by the TI DNNs. The mean square error, given in (1), is set as the loss function.

The metrics considered to evaluate the performance of the DSSE DNN, are the mean absolute percentage error ($MAPE_{\mathrm{v}}$) for the bus voltage magnitudes:

$$MAPE_{\mathrm{v}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{V_{m,i} - \hat{V}_{m,i}}{V_{m,i}}\right|100\% \tag{3}$$

and the mean absolute error (MAE_{a}) for the corresponding phase angles:

$$MAE_{\mathrm{a}} = \frac{1}{n}\sum_{i=1}^{n}\left|V_{a,i} - \hat{V}_{a,i}\right| \tag{4}$$

where $V_{m,i}$ ($V_{a,i}$) and $\hat{V}_{m,i}$ ($\hat{V}_{a,i}$) are the actual and estimated values of the estimated voltage magnitudes (phase angles), and n is the number of states variables for the whole dataset.

As hyperparameter optimization for the DSSE DNNs, the following hyperparameters are investigated: the number of neurons of each hidden layer ranged from 16 to 2048. The number of the hidden shared layers range from 1 to 4, while the number of the branch layers ranges from 1 to 3. Moreover, the number of epochs and the batch size for the training are examined, with the first ranging from 100 to 1000 while the latter from 16 to 128. The learning rate of the adam optimizer is also experimented upon, with possible values set at 0.01, 0.001 and 0.0001. Lastly, similar to the TI model, increasing or decreasing the number of neurons successively is considered.

3.3.3. Transfer learning implementation

By 2020 [23], there are 4 main learning types in inductive transfer learning so far: learning on instances, learning on features, learning on parameters, and learning on relations. Parameter transfer is used in this paper, as it is assumed that an alteration in the topology of the network doesn’t constitute a significant domain shift. This method focuses specifically on transferring the parameters of a pre-trained model to a new untrained one, utilizing them for initialization purposes. For the scope of this paper, in order to speed up the DNN convergence, parameter sharing is utilized between all DNNs used in the same placement algorithms correspondingly: SFS, HC, LASSO, and RF algorithms.

3.3.5. RF implementation

In the RF algorithm, implemented for feature selection and dimensionality reduction purposes, different hyperparameters are used. The number of DTs to be generated, is explored with values ranging from 20 to 200. Increasing the number of this hyperparameter typically leads to more accurate results, since the output is averaged over more DTs, which cover more potential scenarios. However, a large value could lead to a slower training process, due to more DTs being trained at the same time. The Gini impurity is used as the error function for the TI problem:

$$G = 1 - \sum_{i=1}^{k} p_i^2 \tag{4}$$

where p_i is the probability of randomly selecting a data point from class i and k is number of classes.

The MSE is the corresponding metric for the DSSE problem. The RF training stops, when the training of all the DTs is terminated. The criteria used for convergence of the DTs is the convergence of their corresponding error functions. Finally, the sorted importance list is obtained after the training completion.

3.3.6 LASSO implementation

Lasso is utilized only for the DSSE DNN preprocessing, since it can only be employed for regression tasks. It introduces the hyperparameter alpha which controls the strength of regularization applied to the coefficients in a linear regression model. This hyperparameter has a huge impact on the ranking of the features provided by LASSO, so the choice of a proper value is essential. The range examined for alpha spans from 0.000001 to 10. Greater contamination values result in feature importances being reduced to smaller values, thus, performing stronger dimensionality reduction. Once the training is over, the sorted coefficients list is extracted.

Fig. 3 and 4 depict the implemented algorithms for SMD placement regarding TI and DSSE, respectively. The desired accuracy values per task, are the $\mathrm{TI}_{\mathrm{acc}}$, $MAPE_{\mathrm{v}}$, and MAE_{a} thresholds.

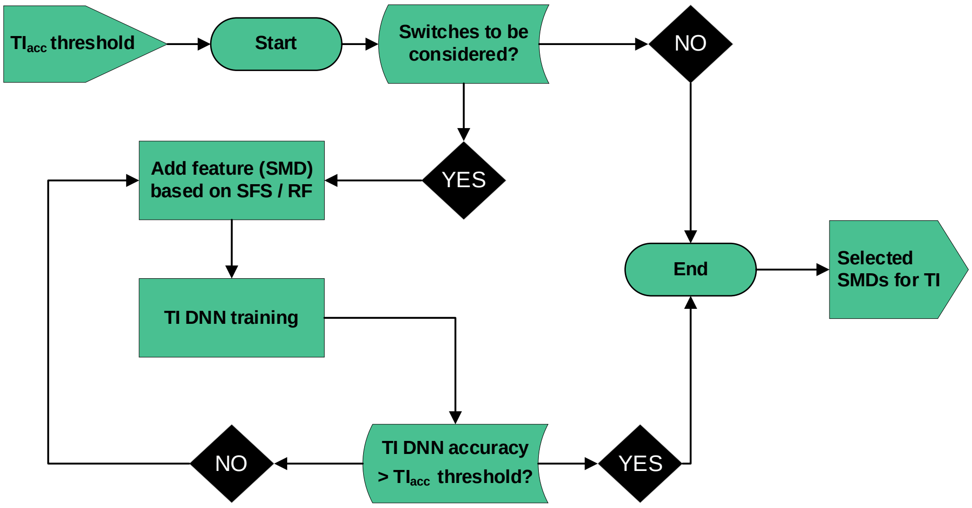

Fig. 3. Algorithm for SMD placement considering TI accuracy.

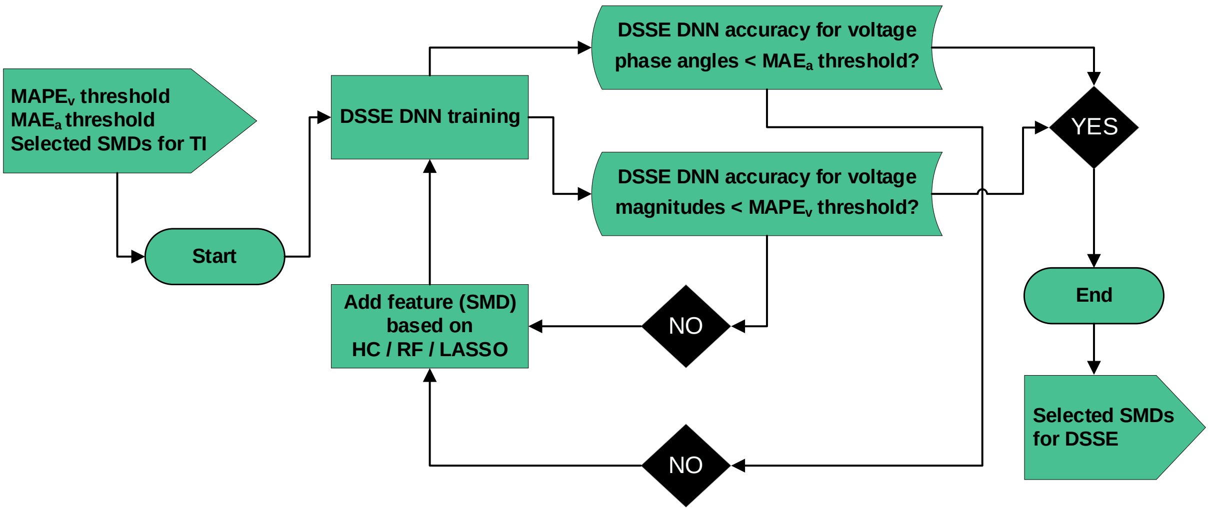

Fig. 4. Algorithm for SMD placement considering DSSE accuracy.

4. Results and showcase

To showcase the use and assess the effectiveness of the proposed methodologies, numerical studies are performed using the IEEE 33-bus benchmark test system as detailed in [16]. All nodes are considered as candidate locations for SMD installation. Each SMD is assumed to measure one nodal voltage and one branch current phasor. For the purpose of TI, the branch current phasors comprise the candidate features to be selected by the SMD placement techniques. For DSSE, the nodal voltage phasors are used as the input features under selection.

As regards the training dataset, the procedure detailed in Section 3.1 is implemented. A set of 16 radial and meshed topologies are taken into account for constructing the full dataset. The MATPOWER toolbox is utilized to carry out the necessary power flow calculations [24]. To account for the measurement errors, a total vector error (TVE) of SMD data below 1% is considered, as reported in the IEEE standard [25]. As commonly assumed, the synchrophasor data errors follow a Gaussian distribution, even though in reality this may be a simplification [26]. Under these circumstances, an error of ±0.1% is assumed for voltage and current magnitude measurements. Also, an error ±0.5% (in degrees) is considered for the voltage and current phase angles.

The target values for the accuracy metrices used for the placement are $\mathrm{TI}_{\mathrm{acc}} = 95\%$, $MAPE_{\mathrm{v}} = 0.30\%$, $MAE_{\mathrm{a}} = 0.15^{\mathrm{o}}$. The selected SMD locations using a) the combination of SFS (for TI) and HC (for DSSE) reported in [10], b) exclusively the proposed RF (for both TI and DSSE) and c) the proposed combination of RF (for TI) and LASSO (for DSSE), are depicted in Fig. 5, 6, and 7, respectively. It is noted that the dashed lines denote the switchable branches of the DN [16], and that the arrows indicate the branch current flows measured by each SMD.

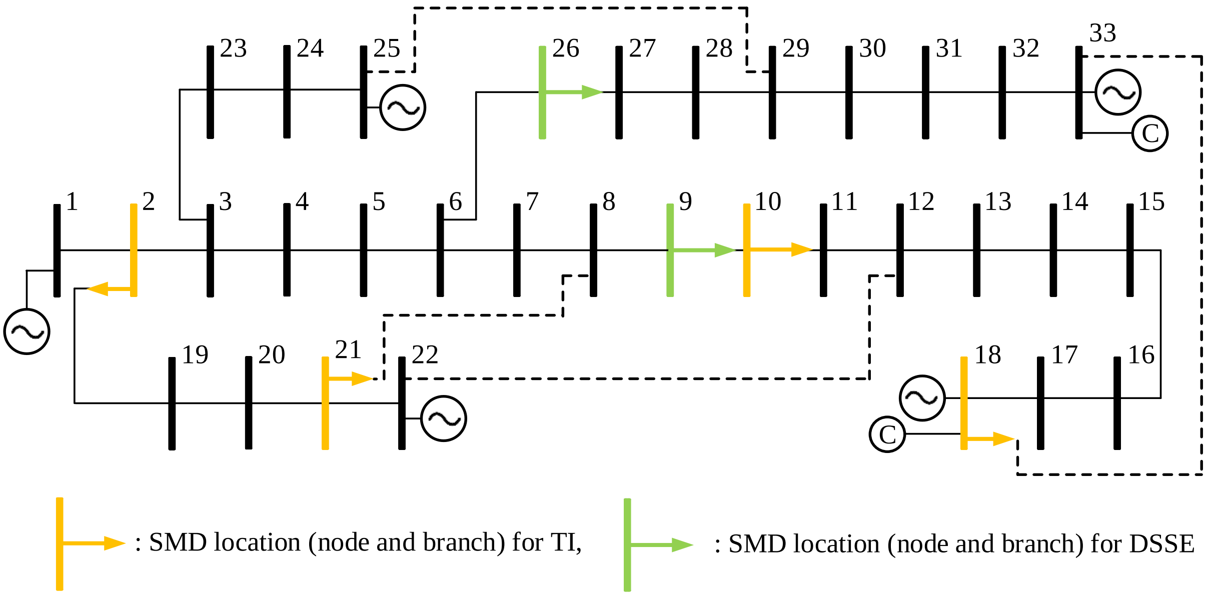

Fig. 5. Selected SMD locations using the methodology in [10] (combination of SFS and HC).

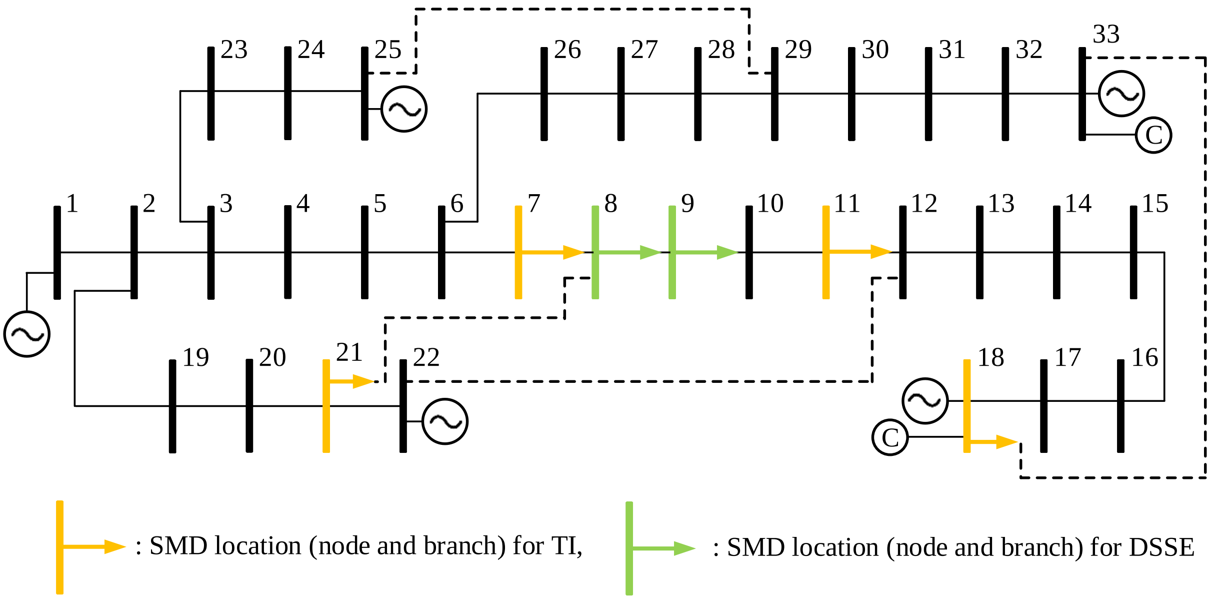

Fig. 6. Selected SMD locations using the proposed RF methodology.

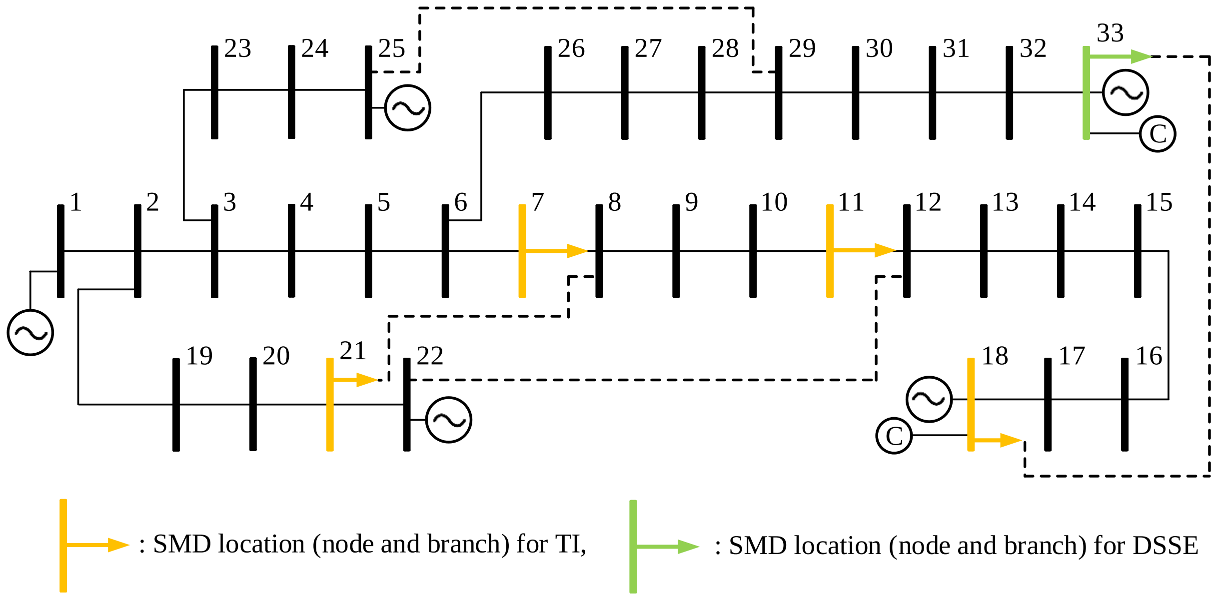

Fig. 7. Selected SMD locations using the proposed combination of RF and LASSO.

As observed, 6 SMD installations are required to attain the desired TI and DSSE accuracy using either the methodology in [10] or exclusively the proposed RF. In case the proposed methodology of RF and LASSO is used, 5 SMD installations are needed. Nodes 18 and 21 are selected by all SMD placement methodologies, a fact which highlights their strategic position.

The computation time per methodology also is a parameter of increased interest. Table 1 includes the computation times and attained TI accuracy using a) the SFS as described in [10], and b) the proposed RF algorithm. As showed, although the accuracy is approximately the same, the computation time is significantly smaller for the latter. Hence, the RF algorithm is less time-consuming while providing quality solutions, thus, it performs much more efficiently.

Table 1. Attained accuracy and computation time per SMD placement methodology for TI.

SMD placement methodology	TI accuracy (%)	Computation time (s)
SFS	95.5	987.1
RF	95.9	50.3

The corresponding time values for DSSE are cited in Table 2. As seen, all the values are comparable, with the ones pertaining to the proposed RF and LASSO being slightly smaller. This difference may have been even larger in case the test system was more compex.

Table 2. Computation time per SMD placement methodology for DSSE.

SMD placement methodology	Computation time (s)
HC	76.6
RF	73.4
LASSO	63.5

As regards the attained DSSE accuracy, Fig. 8 and 9 depict the bar diagrams of the attained $MAPE_{\mathrm{v}}$ and MAE_{a} values for the bus voltage magnitudes and phase angles, respectively, which are estimated by the DSSE DNNs using the SMDs selected per individual methodology.

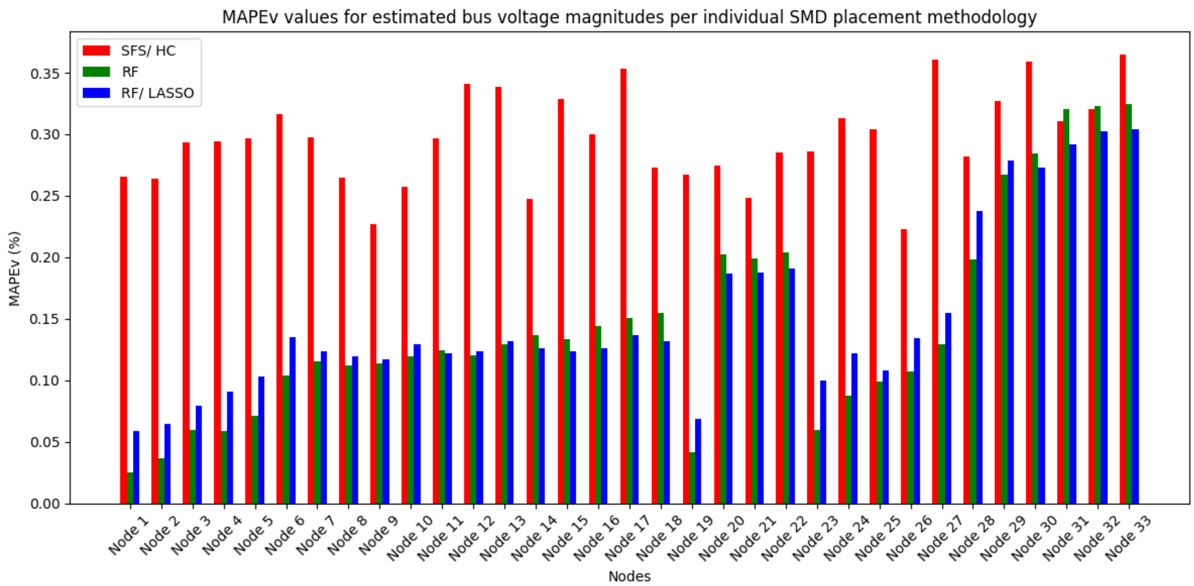

Fig. 8. Bar diagrams of attained $MAPE_{\mathrm{v}}$ values for bus voltage magnitudes per SMD placement methodology.

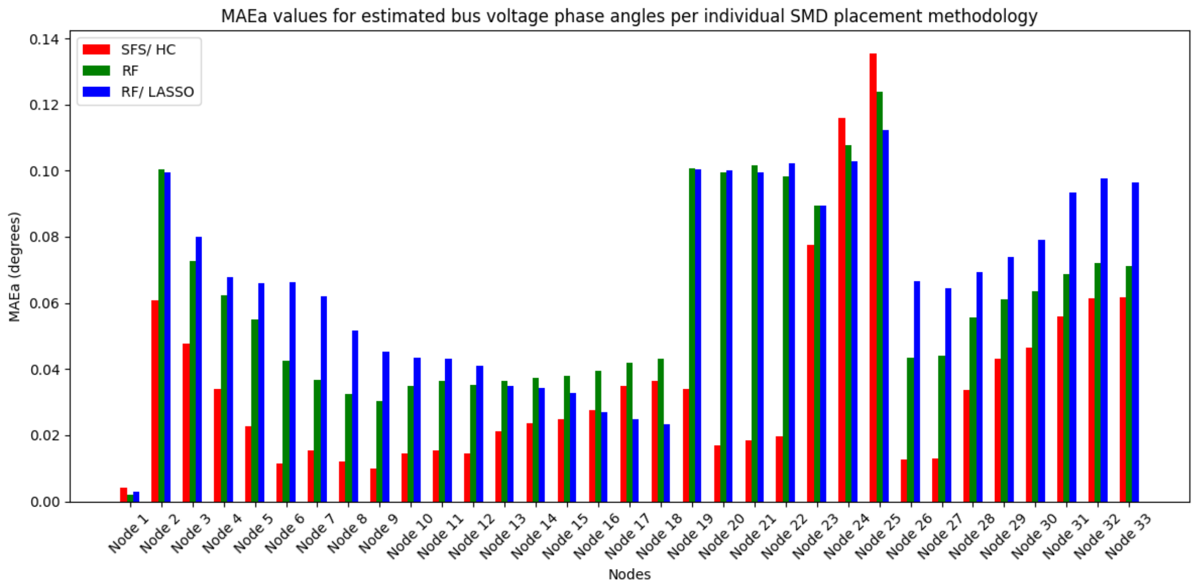

Fig. 9. Bar diagrams of attained MAE_{a} values for bus voltage phase angles per SMD placement methodology.

As perceived in Fig. 8, the DSSE DNNs estimate the bus voltage magnitudes more precisely using the SMDs selected by the proposed RF and LASSO models, since the corresponding $MAPE_{\mathrm{v}}$ values are considerably lower compared to the ones obtained using the SMDs placed by the methodology in [10]. Contrarily, the bus phase angles are estimated more accurately using the SMDs selected by the methodology in [10], as the corresponding MAE_{a} values are lower compared to the ones obtained from the proposed models. Overall, since monitoring voltage magnitudes is on the focus in energy control centers, the outcome of RF and LASSO deployment for SMD placement in terms of DSSE accuracy is favorable.

5. Conclusions

The main scope of this paper is to investigate SMD placement methodologies for conducting quality TI and DSSE using DNNs. The SMD placement problem is conceptualized as a feature selection problem. To model and solve this problem, RF and LASSO algorithms are designed and developed. For reasons of comparison, the related methods described in [10] are also implemented and used as a reference. At first stage, the SMD locations are selected so that the desired accuracy for the TI DNN is met. At next stage, keeping the already selected SMDs, additional ones are placed aiming at the achievement of the desired accuracy by the DSSE DNN. The proposed methodologies pertain to a) the use of RF technique for both stages of SMD placement, and b) the use of RF technique for the former stage and of LASSO for the latter one. The above sets of selected SMDs, along with the one provided by the methodology in [10], are utilized to train and evaluate 2 DNNs for TI and DSSE, respectively.

Numerical studies are carried out using the IEEE-33 bus test system in order to assess the proposed techniques. The conclusions drawn are that the inclusion of the proposed ML models in the feature selection steps can provide lower training times, especially for the TI cases. Also, for both TI and DSSE cases, it is possible to derive DNNs with higher performance, by utilizing RF and LASSO. In particular, the provided SMD selections lead to the reduction of the estimated voltage magnitude estimation errors.

6. Future steps

The RF and LASSO techniques, exploited in the current work, are only 2 out of the multiple ML algorithms currently being implemented both in power system applications. More of them, like principal component analysis (PCA), can be utilized for feature selection tasks aiming at lower training times and higher performance. Also, instead of a simple DNN architecture, more complex models such as recurrent and convolutional NNs, can be implemented for evaluating their performance in TI and DSSE tasks. Finally, the proposed ML techniques need to be tested on larger benchmark test systems, so that more conclusions regarding their applicability and advantages will be drawn.

7. Acknowledgments

This research is supported by the Horizon Europe project OPENTUNITY (Grant agreement Nº 101096333). This paper reflects only the author's views and does not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] K. H. Mohd Azmi, N. A. Mohamed Radzi, N. A. Azhar, F. S. Samidi, I. Thaqifah Zulkifli, and A. M. Zainal, "Active electric distribution network: Applications, challenges, and opportunities," *IEEE Access*, vol. 10, pp. 134655–134689, 2022.

[2] Conseil international des grands réseaux électriques, Ed., *Control and automation systems for electricity distribution networks (EDN) of the future*. Paris: Cigré, 2017.

[3] "State estimation concepts and terminology," Power and Energy Society (PES-TR20), Tech. Rep., 2016.

[4] https://www.platone-h2020.eu/data/deliverables/864300_M15_D4.21.pdf

[5] "IEEE recommended practice for conducting load-flow studies and analysis of industrial and commercial power systems," IEEE Std 3002.2-2018, pp. 1–73, 2018.

[6] A. Abur and A. Gómez Expósito, *Power system state estimation: theory and implementation*. New York, NY: Marcel Dekker, 2004.

[7] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4910–4920, Nov. 2019.

[8] L. Zhang, G. Wang and G. B. Giannakis, "Distribution system state estimation via data-driven and physics-aware deep neural networks," *2019 IEEE Data Science Workshop (DSW)*, Minneapolis, MN, USA, 2019, pp. 258–262.

[9] S. Dahale and B. Natarajan, "Bayesian framework for multi-timescale state estimation in low-observable distribution systems," *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4340–4351, Nov. 2022.

[10] B. Azimian, R. S. Biswas, S. Moshtagh, A. Pal, L. Tong, and G. Dasarathy, "State and topology estimation for unobservable distribution systems using deep neural networks," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, Apr. 2022.

[11] M. Xiao, S. Wang, Z. Ullah, Y. Li, and R. Arghandeh, "Topology detection in power distribution system using kernel-node-map deep networks," *IET Gener. Transm. Distrib.*, vol. 14, pp. 4033–4041, 2020.

[12] D. Gotti, H. Amaris, and P. L. Larrea, "A deep neural network approach for online topology identification in state estimation," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5824-5833, Nov. 2021.

[13] M. Ferdowsi, A. Benigni, A. Monti, and F. Ponci, "Measurement selection for data-driven monitoring of distribution systems," *IEEE Syst. Journal*, vol. 13, no. 4, pp. 4260–4268, Dec. 2019.

[14] P. M. Joshi and H. K. Verma, "Synchrophasor measurement applications and optimal PMU placement: A review," *Elect. Power Syst. Res.*, vol. 199, Jun. 2021.

[15] T. C. Xygkis and G. N. Korres, "Optimized measurement allocation for power distribution systems using mixed integer SDP," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 11, pp. 2967-2976, Nov. 2017.

[16] S. H. Dolatabadi, M. Ghorbanian, P. Siano, and N. D. Hatziargyriou, "An enhanced IEEE 33 bus benchmark test system for distribution system studies," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2565-2572, May 2021.

[17] C. Li, G. Tang, X. Xue, A. Saeed, and X. Hu, "Short-term wind speed interval prediction based on ensemble GRU model," *IEEE Trans. Sust. Energy*, vol. 11, no. 3, pp. 1370-1380, Jul. 2020.

[18] Z. Si, Y. Yu, M. Yang, and P. Li, "Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks," *IEEE Trans. Ind. Appl.*, vol. 57, no. 1, pp. 5–16, Jan.-Feb. 2020.
[19] https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa
[20] K. P. Andresakis, A. Dimeas, and N. Hatziargyriou, "Block order bids prediction for day-ahead market," *2023 IEEE PES Innovative Smart Grid Technol. Europe (ISGT EUROPE)*, Grenoble, France, 2023, pp. 1–6.
[21] S. Zhang, Q. Ding, G. Hu, and C. Nan, "Power distribution network topology reconstruction based on adaptive Lasso," *2018 2nd IEEE Conf. Energy Internet Energy Syst. Integration (EI2)*, Beijing, China, 2018, pp. 1–6.
[22] https://anasbrital98.github.io/blog/2021/Random-Forest/
[23] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Trans. Artif. Intell.*, vol. 1, no. 2, pp. 151-166, Oct. 2020.
[24] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education, " *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
[25] “IEEE/IEC International Standard - Measuring relays and protection equipment - Part 118 -1: Synchrophasor for power systems - Measurements," IEC/IEEE 60255 -118 -1:2018, pp.1 -78, Dec. 2018.
[26] T. Ahmad and N. Senroy, "Statistical characterization of PMU error for robust WAMS based analytics," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 920-928, Mar. 2020.