A new method of decision tree based transient stability assessment using hybrid simulation for real-time PMU measurements

Tohid Behdadnia | Yusuf Yaslan | Istemihan Genc

1 Department of Electrical Engineering, Istanbul Technical University, Istanbul, Turkey
2 Computer Engineering Department, Istanbul Technical University, Istanbul, Turkey

Correspondence
Istemihan Genc, Department of Electrical Engineering, Istanbul Technical University, 34469 Istanbul, Turkey.
Email: gencis@itu.edu.tr

Abstract
In recent years, machine learning (ML) techniques have gained popularity in facilitating real-time transient stability assessment and prediction for early detection of blackouts in power systems. Conventionally, synchrophasor measurements, as the real-time indicators of system dynamics, are fed into the ML-based models. However, if the quality of the synchrophasors used in the process of developing and application of the ML models is not validated, these models could suffer from unreliability issues due to the unrealistic quantities obtained through simulations and due to the interference of erroneous measurements encountered during their application. In this paper, after investigating the properties of different simulation methods, a new hybrid-type simulator that generates a realistic dataset in a feasible time is proposed. Using this simulator, the distortion of the time-series data due to the dynamics of practical phasor measurement units (PMUs) following a disturbance is analysed and the intervals in which the PMU measurements are significantly erroneous are determined. Moreover, a new method of time-series data arrangement for the dataset to be used in ML models is proposed. With this method, the erroneous parts of the time-series measurements are effectively removed, while the remaining relevant information is retained to enhance the transient stability prediction accuracy.

1 INTRODUCTION

Because of the stressed operating conditions, which are aggravated by the growing load demand, and which are also unprecedented due to the high penetration of renewable energy sources, modern power systems have become more vulnerable to their credible contingencies. These contingencies could sometimes trigger a series of protection actions that might develop into cascading failures and might even lead to serious blackouts [1, 2]. In case of a disturbance driving the power system to transient instability, a fast prediction of its security status could be vital for allowing a sufficient time to take emergency control actions [3, 4].

In recent years, artificial intelligence methods, including machine learning techniques, have been broadly applied to real-time transient stability assessment (TSA) of power systems, mainly because of their non-linear modelling capabilities to learn the complex relationship between the stability status of a power system and its time-series measurable quantities of very short time interval following a disturbance [5–15]. Generally, the required time-series measurements are captured synchronously by PMUs located at different points in the power system network, and by arranging these time-series measurements, any instance of features can be built. As each instance conveys an information about the transient behaviour of the power system when it is subjected to a particular contingency, a large collection of instances associated with credible contingencies and operating conditions can be generated through simulation and can be arrayed in the form of a matrix, as a training matrix (TRM) or a testing matrix (TEM).

Despite the high accuracies in stability prediction and successes of ML models reported in literature, if these models are implemented on a practical power system, many of the current state-of-the-art models could suffer from unreliability issues due to an interference of misleading data in the process of model training and its application. In this paper, we explore the boundaries of accuracy and reliability of these models and emphasise the assumptions made in their reported accuracies and their
effects during a practical application. Moreover, in order to create a more reliable and accurate ML-based TSA models, a new methodology involving synthetic data generation and data quality improvement is proposed in this study as an alternative to the conventional approaches in data generation and data modification.

As in many other applications, to obtain a reliable ML model, any misleading features should be removed from the feature space. This requires first identifying the origins and the characteristics of the existing misleading data.

The first type of misleading data that we consider in our study is the unrealistic simulated data (URSD) [16, 17] that are originated from the phasor-based (PB) simulators, which are fast and suitable for generating large datasets, but on the other hand, possessing serious restrictions for the detailed representation of the dynamic aspects of key measuring elements, such as PMUs. These restrictions of PB simulators, especially, in modelling of the PMUs, which are the apparatuses that significantly affect the transients in the measured data, make the dataset of transient type unrealistic and error-free. In this paper, it is shown that the outputs of PB simulators cannot be considered as a reliable alternative to real-life measurements and may derail the ML-based models used for transient stability prediction. To overcome this problem, a new approach involving a hybrid-type simulation is proposed in this study for generating large realistic datasets required by the ML-based models in a feasible time.

The second type of misleading data that we consider in our study is the erroneous (low quality) data [18, 19]. Studies show that phasor calculation algorithm of practical PMUs may become significantly impaired in off-nominal frequency conditions associated with the fault occurrence and fault clearance. The effects of these impairments make the measurements highly erroneous and uninformative in specific periods of time [20–22]. Therefore, the erroneous part of time-series measurements cannot be considered as valid indicators of power system dynamics and should be removed from the feature space in a proper way.

Detection and removal of erroneous measurements requires the study of the transient behaviour of the PMUs located at different point of power system. In this paper, it is shown that highly erroneous measurements exist during a limited time interval (settling time of PMU measurements) following any disturbance. By determining the post-disturbance settling time of each PMU, the erroneous measurements can be effectively discovered. In a large power grid, settling time of a PMU varies widely depending on the size of abrupt changes in the magnitudes and phase angle of the three-phase input signal. In this paper, the settling time of phase-locked-loop (PLL) based PMUs are predicted by clustering and statistical analysis of the input transient signals that are received by the PMUs.

Based on the predicted settling times, erroneous measurements can be detected and isolated from the time-series data captured by PMUs. But it is shown that if the time-series data are arranged and transferred to the feature space by the conventional method [8, 13–15] of building a dataset to be used by machine learning based classifiers, a significant number of feature vectors within the feature space would include both erroneous and accurate measurements. This high interference between the erroneous and accurate data makes us incapable of removing the corrupted feature vectors (feature vectors that comprise highly erroneous measurements) without losing a significant amount of informative data during the process of feature cleansing. To overcome this problem, the time-series measurements are rearranged in the feature space based on a new approach. It is illustrated that the proposed method of data arrangement, without reducing the mutual information between the feature vectors and targets, prevents erroneous measurements from being highly interfered with accurate measurements within an individual feature vector. Thus, a large set of corrupted feature vectors can be easily removed from the dataset to guarantee the quality and relevance of the input feature space for TSA.

In this paper, decision tree (DT)-based classifiers are trained and tested by using the datasets that are generated and arranged through the conventional approach and the proposed method. The results obtained in this study shows the importance of using hybrid simulator to generate a realistic synthetic dataset and the efficacy of the proposed data arrangement method in improving the classification accuracies by enabling the removal of erroneous measurements.

This paper is organised as follows: In Section 2, the process of generating realistic synchrophasor data, using hybrid-type simulation, is presented. Section 3 is devoted to scrutinising the time-series measurements captured by a practical model of PMUs (simulated in EMT domain) and to illustrating how synchrophasor data accuracy is impacted by filtering and estimation algorithms. Section 4 introduces an efficient and a reliable method for predicting the settling time of PMUs located at different points in the network. In Section 5, the approach for ML-based transient stability status prediction is explained, where the proposed method of data arrangement is described in subsection 5.2.3. In Section 6, the proposed method is applied on two test systems, IEEE 68 bus and a western systems coordinating council (WSCC) 127-bus system and, then, its performance is evaluated and compared. Finally, the conclusions are drawn in Section 7.

2 | SIMULATING REALISTIC SYNCHROPHASOR DATA

Offline trained ML-based prediction algorithms have been proposed for decades for online TSA, where synthetic synchrophasor measurements are commonly selected as the inputs to these algorithms. To generate synthetic synchrophasor data, different simulation methods can be used. The accuracy and reliability of the generated data strongly depends on the simulation method that is used to simulate the PMUs, which are the key measuring apparatus. Usage of inappropriate simulators to simulate PMUs may result in generating unrealistic simulated data (URSD). Clearly, URSD is not a good alternative to the practical data, because utilising the URSD either to train or to test a ML-based classifier may result in creating an unreliable and
pseudo-accurate model, which would not be suitable for an online TSA.

2.1 Properties and restrictions of different simulators to model the PMUs

2.1.1 Phasor-based (PB) simulator

In phasor-based simulators (also known as transient stability programs), the voltages and currents are computed as phasors. Although PB simulators can be used to effectively represent the quasi-steady-state angle dynamics for TSA purposes, these simulators present serious restrictions in representing the dynamic aspects of PMUs. The constraints in modelling the dynamic effect of PMUs and in simulating their transient responses, cause PB simulators to generate a set of unrealistic and error-free time-series data following a disturbance.

2.1.2 Electromagnetic transient (EMT) simulator

The EMT simulation method, which is based on three-phase waveforms of instantaneous voltages and currents, provides a high degree of accuracy in representing the dynamic phenomena of shorter time-scales. EMT methods use detailed models that enable us to evaluate fast system transients and their effects on the dynamic behaviour of nonlinear and key measuring components, such as PMUs. Although the synthetic data generated by EMT-type programs are quite realistic, the EMT simulation may not be feasible, and most of the times, would be impossible to apply for generating the large datasets required by the ML-based algorithms.

2.2 Application of hybrid-type simulation to generate realistic PMU data

In order to generate large datasets characterising the transient stability of a power system where the transients in PMU measurements are also accurately represented and included with less computation time, the usage of a hybrid simulator can be considered as an effective solution. The proposed hybrid phasor-based and electromagnetic transient (HPE) simulation method is based on both phasors and three-phase waveforms of currents and voltages. The main objective of HPE simulation is to split the original network into two parts, in such a way that, based on the required modelling accuracy, one part is simulated by the PB simulator while the other is carried out by the EMT simulator. EMT-type simulation is used for the smaller part in which more detailed and accurate results are needed. This part may comprise PMUs or any other elements whose dynamic behaviour are to be characterised more accurately by simulating their detailed models with smaller time-steps. The other part that embraces extensive portions of the network (including generating units, governors, exciters, stabilisers, loads etc.) is simulated by the PB simulator in phasor domain. In this part, whereas less detailed models of the components are sufficient, the capability of the simulator for fast computation is essential. Accordingly, interfacing PB and EMT simulators builds an HPE simulator that inherits the merits of both simulators. In an HPE simulator, EMT and PB simulators are run on two separate zones: (1) the detailed system (DS) and (2) the external system (ES). Thus, each simulator requires a true picture of the other zone which adequately reflects its characteristics. This needs a converter block which should be able to convert phasors of PB simulators to equivalent three-phase waveforms.

To clarify the idea, we illustrate the total scheme of the proposed simulation method in Figure 1(a). Additionally, Figure 1(b) shows the general structure and principal components of a typical phase locked loop based PMU used in Figure 1(a).

3 CHARACTERISING THE SYNCHROPHASORS MEASURED BY PRACTICAL MODEL OF PMUS

Studies on the transient behaviour of a detailed model of PMU show that the phasor calculation algorithms of these apparatuses become significantly impaired in the off-nominal frequency conditions mainly due to the presence of large and abrupt changes in the input three-phase signal. The effects of this unavoidable impairment of phasor estimation algorithm appear as transient ripples on measurements, persisting for a few cycles and then dying out over time. The data measured during the interval that these ripples exist, are highly affected by the dynamics of PMUs and would be erroneous and misleading in characterising the transient stability of the power system. Thus, these erroneous quantities either simulated or observed in practice should be detected and are not to be used in the transient stability prediction or any control algorithm.

To detect the erroneous measurements among the reported time-series data, post-disturbance behaviour of PMUs should be studied and the die-out time of transient ripples should be determined. Basically, the die out time of the transient ripple after a disturbance for any PMU is dependent on the changes in magnitude and frequency of three-phase input signals, where the electrical distance of the PMU to the fault is an important factor. It can be shown that the duration of post-disturbance transient ripples (known as post-disturbance settling time) is longer for PMUs with lower electrical distance from the fault location because of the relatively larger step changes in the input signals.

In order to demonstrate the evolution of the measurements of PMUs located at different points in the grid, the western systems coordinating council (WSCC) 127-bus system is simulated using the proposed hybrid-type simulation. Figure 2 illustrates the actual and measured voltage magnitude (VM) obtained from PMUs located at three different distances from a three-phase fault. Based on the standard IEEE Std C37.118.1-2011, a measurement error ($E$) less than 1% is normal and can be neglected. But synchrophasors with $E > 1\%$ are considered as significantly erroneous and should be detected and not used.
As can be seen in Figure 2, the lengths of settling times after a fault occurrence ($T_{DS}$) or after a fault clearance ($T_{PS}$) depend considerably on the size of abrupt changes in the input signal’s magnitude. Therefore, when a post-contingency transient stability assessment is to be done based on PMU measurements, the post-disturbance settling times of the PMUs should be predicted considering the size of abrupt changes in magnitude after disturbances to detect and avoid the interference of erroneous measurements. In this study, a new method for predicting the post-disturbance settling times of PMUs located at different points in a power system is proposed and presented in Section 4.

### 4 PREDICTING THE SETTLING TIMES OF PMUS

According to Figure 2 and the explanations given in Section 3, due to the differences in the settling time of the PMUs located at different points of a power grid, a fixed $T_{DS}^P$ and $T_{PS}^P$ cannot be assumed for all PMUs. On the other hand, in order to quickly detect the erroneous part of time-series measurements, we need to have a reliable knowledge in advance about the settling times for all PMUs in case of a given contingency [18, 19]. While the settling times, $T_{DS}^P$ and $T_{PS}^P$, of any PMU are highly dependent on the characteristics of its input three-phase signal (specially the size of abrupt changes), PMUs receiving input signals with similar changes both in magnitude and form would possess similar settling times [20–22].

A clustering method can be used to determine the settling time of any PMU following a contingency, where the settling time is strongly dependent on the location of PMU with the respect to the place of contingency. Consider a power system equipped with $P$ number of PMUs. After a contingency, $P$ different transient signals are received by the installed PMUs. Given an instance, consider the set of input transient signals $\{x_1, \ldots, x_p, \ldots, x_P\}$, where $x_p$ is the time-series data collected from $p$th PMU. Using k-means algorithm, each of these signals is assigned to one of the clusters, $Y_r$, $r = 1, \ldots, R$, such that

$$
\arg\min_{\{Y_1, \ldots, Y_R\}} \sum_{r=1}^R \sum_{x \in Y_r} \| x - \mu_r \|^2,
$$

(1)
where $\mu_r$ is the centre of cluster $Y_r$. The clustering process can be started with an arbitrarily chosen instance forming a set of centres. Then, it can be carried out by updating the centres as each new member is added to the cluster. In this paper, PMUs that receive similar transient signals (signals assigned to the same cluster) in different instances are considered as PMUs with the same rank. It can be deduced that as the post-disturbance settling times of equal-rank PMUs are significantly similar due to the similarity of their input signal, a fixed $T_D$ and $T_P$ can be assumed for the equal-rank PMUs for any credible future scenarios. Therefore, in practice and online applications, it would be sufficient to determine the cluster of the input signal based on (1), and then, according to the settling time that is calculated off-line for the centre signal of each cluster, the approximate values of $T_D$ and $T_P$ of the PMUs in that cluster are simply specified.

5 | ML-BASED TRANSIENT STABILITY PREDICTION

5.1 | Off-line supervised dataset generation

Synchrophasor measurements are usually generated through off-line time-domain simulations. In these simulations, the power system is subjected to a large number of credible contingencies [23] involving disturbances that can make the system unstable. These contingencies can include a sudden three-phase short circuit, an outage of a transmission line or a generator followed by a subsequent tripping. The clearing time ($t_d = t_C - t_r$) can vary within $t_{d_{\text{min}}}$ and $t_{d_{\text{max}}}$ ($t_{d_{\text{min}}} \leq t_d \leq t_{d_{\text{max}}}$). For a reliable TSA, the training set should cover a sufficient number of operating points so that the models developed can be a good representation of the practical power system and can be tolerant to the uncertainties in operating conditions. For each operating point, time-domain simulations are executed to identify the transient stability index with respect to any credible contingency. The system's transient stability can be assessed and classified, as either stable or unstable, with a label $L$

$$L = \begin{cases} 1 \text{ (stable)} & \text{for } \eta > 0 \\ -1 \text{ (unstable)} & \text{for } \eta \leq 0 \end{cases} \quad (2)$$

where $\eta$ is a power angle-based stability index, which can be obtained by observing the maximum angle separation between any two generators at the same time in post-fault response [24], to determine if any generator in the system is out of synchronism.

5.2 | Construction of feature spaces with different feature vector configurations

The input feature space of an ML-based TSA model can be represented by a $S \times F$ matrix, where $S$ is the total number of instances (rows) and $F$ is the total number of features (columns). In order to construct such a matrix from the generated dataset, synchrophasor measurements are mapped to the fixed-dimensional sub-vectors and, then, arranged in a meaningful manner, thus, the instances of features are created [25]. In this subsection, at first, we explain how the variable-length
time-series measurements can be mapped to the constant-size
sub-vectors, and, then, arranged in different ways to construct
the feature space matrix required for the learning process.

### 5.2.1 Mapping variable-length time-series measurements to fixed-dimensional sub-vectors

Since \( t_d \) is varied in different scenarios \( (t_{\text{cl}} \leq t_d \leq t_{\text{cl max}}) \), in the cases where fault-on measurements are used as predictor features, the dimensions of different instances would not be the same, and consequently a fixed-size feature space matrix cannot be built. In order to overcome this problem and for better description of the proposed method we define three initially empty sub-vectors of data for mapping the captured synchrophasors without disrupting their time-series format. These sub-vectors are represented by (3)–(5):

\[
\alpha_{(s,p)} = \left[ \alpha^1_{(s,p)}, ..., \alpha^i_{(s,p)}, ..., \alpha^{n}_{(s,p)} \right]_{1 \times f},
\]

\[
\beta_{(s,p)} = \left[ \beta^1_{(s,p)}, ..., \beta^i_{(s,p)}, ..., \beta^{m_{\text{cl}}}_{(s,p)} \right]_{1 \times m_{\text{max}}},
\]

\[
\gamma_{(s,p)} = \left[ \gamma^1_{(s,p)}, ..., \gamma^i_{(s,p)}, ..., \gamma^{m_{\text{cl}}}_{(s,p)} \right]_{1 \times a},
\]

where \( \alpha_{(s,p)}, \beta_{(s,p)} \) and \( \gamma_{(s,p)} \) are pre-fault, during-fault and post-fault time-series sub-vectors, respectively. Here, the subscripts \( p \) is the identification (ID) number of PMU, \( p \in \{1, ..., P\} \), and \( s \) is the identification number of the simulated scenario, \( s \in \{1, ..., S\} \). The entries \( \alpha^i_{(s,p)}, \beta^i_{(s,p)}, \gamma^i_{(s,p)} \) are filled by the \( i \)th sample of synchrophasors reported during the respective intervals from different stages (pre-fault, during-fault and post-fault). Depending on the maximum allowed prediction time, and the maximum clearing time, the size of the sub-vectors can vary. In this paper, \( \alpha_{(s,p)} \) contains only one pre-contingency, thus \( f = 1 \), whereas the post-fault vector covers all the reported data before the maximum prediction time allowed, thus \( n \) can be selected as \( n = t_{\text{cl max}} / R_t \), where \( t_{\text{cl max}} \) is the maximum prediction time allowed, and \( 1 / R_t \) is the reporting frequency. Since the fault-on sub-vectors must cover all the during-fault data associated with any scenario with a varying clearing time, the size of the vector must be equal with \( m_{\text{max}} = t_{\text{cl max}} / R_t \), which is the number of samples that can be captured during the maximum clearing time. Thus, for a given scenario, with \( m_{d} \) number of during-fault samples \( m_{d} = \left| t_C - t_f \right| / R_t \), the entries of the \( \beta_{(s,p)} \) can be filled as follow:

\[
\beta^i_{(s,p)} = \begin{cases} 
 f(t) \in [t_f + (i - 1) \times R_t, t_f + i \times R_t] & \text{if } 1 \leq i \leq m_d \\
 \text{NaN} & \text{if } m_d < i \leq m_{\text{max}}
\end{cases}
\]

There are different methods to handle non-existent values which are represented by NaNs in the fault-on sub-vectors. In our case NaNs can simply be replaced by zeros.

### 5.2.2 Conventional method (static arrangement of time-series data)

Conventionally, in order to construct a feature space, sequential PMU data of a certain type of electrical quantities, for example, voltage magnitude, are arranged in a static manner, for example, in the form of a matrix purely based on the ID numbers of PMUs. Equations (7) and (8) illustrate how a feature space matrix (FSM) is constructed based on a conventional method. In (7), the vector \( \xi \) contains all spatio-temporal features \([12]\), representing dynamic characteristics of the power system when \( t_b \)th scenario occurs, and \( \xi_{(s,p)} \) contains the portion of spatio-temporal features (time-synchronised measurements) that are reported from the PMU with ID number \( p \). In (8), the subscript \( CV \) indicates that FSMs constructed based on conventional method.

\[
\xi_s = [\xi^1_{(s,1)}, ..., \xi^i_{(s,p)}, ..., \xi^{m_{\text{cl}}}_{(s,p)}]_{1 \times F},
\]

\[
\text{FSM}_{CV} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_S \end{bmatrix}_{S \times F},
\]

### 5.2.3 Arranging time-series data based on rank of PMUs (proposed method)

As mentioned in Section 3, some parts of synchrophasor measurements are significantly erroneous in the presence of disturbances and transients. When the time-series measurements are arranged and transferred to the feature space based on the ID number of PMUs (conventional method), erroneous data become highly interfered with accurate data and erroneous measurements, that is, erroneous data can lie anywhere in the \( F \)-dimensional feature space. In this regard, significant number of feature vectors (columns of \( \text{FSM} \)) are corrupted by erroneous data. Feeding corrupted feature vectors into the statistical learning models can significantly derail the classification and decision-making process. To prevent such an integration of erroneous data with accurate data, a new method of data arrangement is proposed in this section. In the proposed method, time-series measurements are arranged consecutively according to the rank of PMUs. With this arrangement, without reducing mutual information between the feature vectors and targets, the erroneous data are placed in some specific feature vectors. This property makes us able to remove erroneous data by eliminating a limited number of feature vectors containing...
non-informative or unreliable information about the power system dynamics. The details are presented in the Appendix. Equations (9) and (10) illustrate the details of the proposed arrangement procedure:

$$\vartheta_j = [\xi_{(1)}, \ldots, \xi_{(n)}, \ldots, \xi_{(n)}]_{1 \times F}, \quad (9)$$

$$FSM_{New} = \begin{bmatrix} \vartheta_1 \\ \vdots \\ \vartheta_j \\ \vdots \\ \vartheta_{j \in F} \end{bmatrix}. \quad (10)$$

In (9), the vector $\vartheta_j$ contains all spatio-temporal features, representing dynamic characteristics of the power grid when $j$th scenario occurs, and $\xi_{(n)}$ contains the portion of spatio-temporal features (time-synchronised measurements) that are reported from PMU with rank $r$. In (10), the subscript $New$ indicates that $FSM_{New}$ is constructed based on the new method.

5.3 | Reliability-based feature cleansing and its effects on the informativity of dataset

Feature cleansing or feature cleaning (FC) is the process of detecting and removing a subset of features that are corrupted, inaccurate, invalid or distorted. Feature cleansing is a part of the data pre-processing process by which the size of feature space can be reduced for a faster learning speed, better accuracy. Additionally, by FC, the quality of the input feature space can be increased for obtaining more reliable ML models.

An efficient feature cleansing algorithm should effectively remove the erroneous and misleading data, while keeping the informativity of the dataset. To ensure the efficiency of the used FC algorithm, information theoretic quantities, such as average normalised mutual information (ANMI), can be computed and compared before and after the feature cleansing. Let $A_j$ and $C$ be the $j$th feature vector of the feature space matrix and the vector of class label, respectively. In essence, after FC, the average normalised mutual information can be calculated as follows:

$$ANMI = \frac{\sum_{j=1}^{j=m} NMI_j(A_j; C) - \sum_{j=n}^{n=N} NMI_j(A_j; C)}{F - k_{CFV}} \quad (11)$$

where $NMI_j(A_j; C)$ is the mutual information between the feature vector $A_j$ and the vector of targets, while $\lambda_{CFV}$ is the set of corrupted feature vectors, which contain significant amount of erroneous data. A significant reduction in $ANMI$ after applying FC, reduces informativity of the input feature space, which could negatively affect the classification performance.

5.4 | Decision tree based classifiers

A decision tree (DT) is constructed using a set of training instances, and is, then, applied to classify a set of unseen instances. DTs are built using a top-down search performance for the data classification. Beginning from a root node, a feature, called the splitting feature, is selected to divide the feature space into two subsets. The result is two child nodes, either terminal or internal. An internal child node becomes a parent node, at which another splitting feature and corresponding splitting value are chosen. This sequential splitting procedure finally leads to terminal nodes called leaf nodes. In a two-class classification problem, each leaf node is classified and labelled with one of the classes, as either secure or insecure, as studied in this paper. A path from the root node to a leaf node is characterised by a series of questions (yes/no questions) about the features associated with the data. As illustrated in Figure 3, a complete DT consists of a series of splitting rules ($\vartheta_1, \vartheta_2, \ldots$), a root node, internal nodes, and leaf nodes.

Basically, to answer each yes/no question, a set of statistical criteria is used for data classification. The two commonly used measures are Gini and the Entropy indexes. For a typical two-class dataset (negative and positive), the entropy of the data as the purity degree of the dataset is defined as follows:

$$\Delta I(S^+, S^-) = -D_{S^+} \log_2(D_{S^+}) - D_{S^-} \log_2(D_{S^-}), \quad (12)$$

where

$$D_{S^+} = \frac{S^+}{S^+ + S^-}, \quad (13)$$

$$D_{S^-} = \frac{S^-}{S^+ + S^-}, \quad (14)$$

where $S^-$ is the number of input instances with negative target class (unstable) and $S^+$ is the number of input instances with positive target class (stable). The entropy is highest when the number of instances with positive class equals to the number of data with negative class (i.e. $S^+ = S^-$), whereas it is minimum when all the instances have the same target classes (i.e. $D_{S^+} = 1$ or $D_{S^+} = 1$). The Gini index $G(\varphi)$ at a node $i$ is defined as:

$$G(\varphi) = 1 - \sum_{i} P(i | \varphi)^2, \quad (15)$$

where $P(i | \varphi)$ is conditional probability of category $i$ in node $\varphi$ and it can be defined via (16) to (18):

$$P(i | \varphi) = \frac{P(i, \varphi)}{P(\varphi)}, \quad (16)$$

$$P(i, \varphi) = \frac{\pi_{i} N_{i}(\varphi)}{N_{i}}, \quad (17)$$
where $\pi_i$ is the prior probability value for category $i$, $N_i(o)$ is the number of records in category $i$ of node $o$, and $N_i$ is the number of records of category in the root node.

5.5 Indices for model performance evaluation

Since TSA has the defect of ignoring unstable instances, it is unreasonable to use only the accuracy of all instances as the evaluation index for the model. In order to effectively evaluate the performance of the ML-based TSA models this paper uses the confusion matrix [10] shown in Table 1.

The indices to evaluate the performance of the ML-based TSA model are defined as follows:

\[
ACC = \frac{w_{11} + w_{22}}{w_{11} + w_{12} + w_{21} + w_{22}} \times 100\%,
\]

\[
TUR = \frac{w_{22}}{w_{12} + w_{22}} \times 100\%.
\]
where \( ACC \) represents the overall accuracy, \( TUR \) represents the proportion of correct results predicted to be unstable for all unstable instances, \( TSR \) represents the proportion of correct results predicted to be stable from all stable instances.

6 | CASE STUDIES ON THE TEST SYSTEMS

In this section, analysis is implemented on an IEEE 68 bus and a WSCC 127-bus system in order to test the effectiveness of the proposed methods. Both are well-known test systems for similar TSA studies [12, 26]. There are 16 generators and 86 transmission lines in the 68 bus test system, and 37 generators and 211 transmission lines in the 127 bus system. For online monitoring, all of the buses are equipped by PMUs in the 68 bus test system, while in the 127-bus system synchrophasors are captured only from a limited number of PMUs that are optimally located for state estimation based on [27]. The one-line diagram of 127-bus system and the location of PMUs are shown in Figure 4.

In this paper TSAT software [24], MATLAB/SIMULINK, and PYTHON are used for phasor-type simulation, EMT-type simulation, and implementing machine learning algorithms, respectively.

6.1 | Database generation using hybrid-type simulation

The datasets are generated via time-domain hybrid-type simulation. Based on the proposed method, power system components, including generating units, exciters, governors, stabilisers, loads etc. are simulated as external system in phasor domain using TSAT software, and the phase locked loop (PLL)-based PMUs (conforming with IEEE Std C37.118.1-2011) are simulated as detailed systems in EMT domain using MATLAB software.

For the test systems we studied, the loading level of each system is varied from 80\% to 120\%. The contingencies considered are the three-phase-to-ground fault that occur on all buses and on all lines, where the fault is located at 25\%, 50\%, and 75\% of their length. The line faults are cleared by removing the transmission line. The start time of the fault is uniformly set to \( t_{F} = 0.02s \), the operation time of the proximal circuit breaker is set to 4–8 cycles (4 \( \leq \) \( m_{cl} \leq 8 \)), and the time-domain simulation duration is set to 10 s. A class label (“−1” or “1”) is assigned to each simulated scenario based on descriptions given in Section 5.1.

Finally, 30,000 samples are generated, for each test system. The ratio of unstable instances to stable instances is almost 1 : 2. To prevent the model from overfitting and effectively evaluating models performance, datasets are randomly divided into 1 : 3 as testing set and training set, respectively.
6.2 Comparing outputs of simulators

In order to show the difference between time-series PMU data obtained from the proposed HPE simulator, and synchrophasor data obtained from PB simulator (transient stability simulator), Figure 5 is given. In this figure three main electrical quantities, voltage magnitude (VM), voltage angle (VA), and frequency (F), are presented to illustrate the short time effects of dynamically fast events and transients on the PMU measurements. It is demonstrated how these destructive non-linear effects are overlooked in the commonly used PB simulations.

It can be seen in Figure 5(a) that after fault occurrence, $t = t_F$, and fault clearance, $t = t_C$, there are some delays and ripples in the outputs of the HPE simulator that could not be modelled by PB simulators. These ripples persist for a few cycles and then die out over time. According to Figure 5(b), it can be seen that there is a phase difference between the outputs of PB simulator and outputs of HPE simulator which can be due to the filtering mechanisms being applied during the discrete Fourier transform (DFT)-based phasor estimation process [20–22]. In Figure 5(c), there are some significant overshoots in the outputs of HPE, which are not seen in the outputs of PB simulator. In general, it can be concluded that, in the presence of dynamically fast events and transients, measurements of the electrical quantities are highly affected by the inherent dynamics of PMUs. However, error-free outputs of PB simulator are devoid of actual PMU data attributes and dynamic influence the phasor estimation algorithm. Obviously, error-free measurements do not bear true representations of PMU datasets and could be inadequate for developing a reliable ML-based TSA model.
6.3 | Impacts of dynamic response of PMUs on classification performance

Generally, the classifiers use the post-disturbance measurements of the first few cycles to predict stability status of power system. Thus, these measurements should truthfully represent power system dynamics. Nevertheless, as it was shown in the previous section, the accuracy of measurements may highly be influenced by the presence of transient phenomena due to inherent dynamics of PMUs. To demonstrate the impacts of dynamic response of PMUs on the classification performance, at first, we train several well-known DT-based classifiers with the datasets obtained from PB simulator, and then we retrain the same models with the datasets obtained from the HPE simulator. In this section, we choose the voltage magnitude as the input features and the time window used for all models is 0.1s.

The parameters of all classifiers are tuned using a 5-fold cross validation.

Tables 2–5 show the performance of different classifiers for the testing datasets of the IEEE 68-bus and WSCC 127-bus systems. By comparing the results reported in Tables 2 and 3, it can be seen that all models show a poor performance when they are tested with realistic synthesised data, containing erroneous measurements of PMUs. The deficiency in performance is more evident in predicting unstable cases ($TUR\% < 87\%$).

6.4 | Settling time of equal-rank PMUs

In order to detect erroneous parts of time-series measurements, obtained from the PLL-based PMUs, the mean and standard deviation of both $T_{DS}$ and $T_{PS}$ are calculated over a large number of instances for equal-rank PMUs (see Section 4). The obtained results are shown in Figures 6 and 7 for the 68-bus and 127-bus systems, respectively. A very small standard deviation $\sigma_r (\sigma_r \leq 1 \text{ cycle } | r = 1, 2, \ldots, R)$, indicates that the settling times of equal-rank PMUs tend to be very close to the mean. Thus, the calculated mean value (mean of $T_{DS}$ and $T_{PS}$) can be assumed as a good approximation for the settling time of equal-rank PMUs when the power system is subjected to the disturbances.

6.5 | Effects of feature cleansing on classification performance

In order to guarantee the quality of the input feature vectors and minimising misleading effects of erroneous measurements, a reliability-based FC is applied on both $FS_{MCV}$ and $FS_{MNew}$. Thereby, corrupted feature vectors (feature vectors that contain significant amounts of erroneous measurements) are removed from $FS_{MCV}$ and $FS_{MNew}$. The ANMI before and after FC is shown in Table 6. It can be seen that, after applying FC on the conventionally arranged dataset, the ANMI is decreased to 53% and 66% in the 68-bus and 127-bus systems, respectively. This could be objectionable and may have a considerably negative impact on the classification performance. Nevertheless, due to the special properties of the proposed method of

---

**TABLE 2**  TSA results for 68-bus test system (68 PMUs)

| Model               | ACC (%) | TUR (%) | TSR (%) | ACC (%) | TUR (%) | TSR (%) |
|---------------------|---------|---------|---------|---------|---------|---------|
| XGBoost             | 98.7    | 96.4    | 99.2    | 92.1    | 81.5    | 96.4    |
| Random forest       | 98.8    | 97.1    | 99.4    | 94.1    | 86.5    | 97.3    |
| Decision tree       | 98.0    | 95.8    | 98.8    | 90.6    | 81.0    | 94.3    |
| Bagged tree ensemble| 98.3    | 96.3    | 99.0    | 91.8    | 81.4    | 96.2    |

**TABLE 3**  TSA results for 127-bus test system (39 PMUs)

| Model               | ACC (%) | TUR (%) | TSR (%) | ACC (%) | TUR (%) | TSR (%) |
|---------------------|---------|---------|---------|---------|---------|---------|
| XGBoost             | 97.6    | 95.4    | 98.5    | 90.8    | 80.1    | 95.2    |
| Random forest       | 97.9    | 96.0    | 98.6    | 92.2    | 84.3    | 95.6    |
| Decision tree       | 96.8    | 94.4    | 97.9    | 89.5    | 79.8    | 93.7    |
| Bagged tree ensemble| 97.1    | 95.0    | 98.1    | 90.6    | 80.0    | 95.1    |
TABLE 4  TSA results for 68-bus test system (68 PMUs)

| Model                  | ACC (%) | TUR (%) | TSR (%) | ACC (%) | TUR (%) | TSR (%) |
|------------------------|---------|---------|---------|---------|---------|---------|
| XGBoost                | 87.0    | 84.8    | 87.9    | 91.9    | 80.3    | 96.2    |
| Random forest          | 88.8    | 86.9    | 89.5    | 93.9    | 86.4    | 97.1    |
| Decision tree          | 86.8    | 84.1    | 87.5    | 90.2    | 81.3    | 94.0    |
| Bagged tree ensemble   | 87.2    | 85.1    | 88.2    | 91.7    | 81.5    | 96.1    |

TABLE 5  TSA results for 127-bus test system (39 PMUs)

| Model                  | ACC (%) | TUR (%) | TSR (%) | ACC (%) | TUR (%) | TSR (%) |
|------------------------|---------|---------|---------|---------|---------|---------|
| XGBoost                | 83.1    | 80.9    | 84.0    | 90.5    | 80.0    | 94.9    |
| Random forest          | 86.0    | 84.1    | 86.7    | 92.0    | 84.2    | 95.3    |
| Decision tree          | 80.9    | 78.5    | 82.0    | 88.9    | 79.0    | 93.2    |
| Bagged tree ensemble   | 82.9    | 80.8    | 83.9    | 90.4    | 79.9    | 94.8    |

data arrangement, ANMI of $\beta_{M_{N_{\text{ep}}}}$ does not decrease much with FC, and the informativity of dataset is retained. Table 7 illustrates the performance of four different DT-based classifiers. It can be seen that the performance of all classifiers are considerably improved after applying FC. This performance improvement is more evident in predicting unstable cases, which shows the importance and effectiveness of the proposed method.

FIGURE 6  Mean and standard deviation of the post-disturbance settling times for PMUs with equal rank (68-bus system)
In this section in order to evaluate the performance of the proposed method under various disturbances, the simulated contingency scenarios are divided into three groups:

- Group 1: Three phase fault at a bus: the fault is cleared without tripping any line.
- Group 2: Three phase fault on a transmission line: The fault is cleared by tripping the faulted line.
- Group 3: Three phase fault near a generator bus: The generator is tripped in order to protect it.

The total number of instances is 3000, 6000, and 2500 in Group 1, Group 2, and Group 3, respectively. The ratio of unstable instances to stable instances is almost 1:2 in each group. The datasets are randomly divided into 1:3 testing set and learning set, respectively. The TSA results are provided in Table 8. It can be seen that all of the developed classifiers perform well when they are trained by the instances of the first and second group. However, the classifiers show a poor performance when they are trained and tested with the instances of the third group. This is basically due to the significant topology changes after generator tripping.

7 | CONCLUSION

In this paper, to develop a reliable ML-based TSA model, a new approach of hybrid-type simulation is proposed for generating large datasets of realistic synchrophasors in a feasible time. It is shown that the proposed simulation method provides an adequate implementation environment for simulating the nonlinear effects of PMUs on the wide-area measurements. The outputs of the hybrid-type simulator are compared with the outputs of the commonly used phasor-based simulators; then the influence of dynamic response of PMUs on the stability classification performance is illustrated. It is shown that some portions of the realistically simulated time-series data could not be reliable for representing the transient stability characteristic of a power system, basically because of the errors originating from the dynamic behaviour of PMUs. To make the input dataset more reliable, the approach of detection and elimination of erroneous measurements from the dataset is proposed. For effectively removing the erroneous measurements, a new and effective method of data arrangement is developed. It is shown that the proposed method of data arrangement enables us to remove the erroneous measurements without reducing the informativity of the dataset and the proposed method provides a significant improvement in prediction accuracies for online TSA, as it is compared with the alternative approaches. We also investigated how the method performs in various transient disturbances. In our future work, we will study the impacts of non-fundamental frequency components on the PMU.
measurements, and then study their effects on the classification performance of ML-based models. In addition, in order to enhance the prediction performance, deep learning techniques will be used to mitigate erroneous measurements instead of removing them from the feature space.

**ACKNOWLEDGEMENT**

This work is supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under project no. 118E184.

**ORCID**

Istemihan Genc [https://orcid.org/0000-0001-7077-8895](https://orcid.org/0000-0001-7077-8895)

**REFERENCES**

1. Liscouski, B., Elliot, W.: Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations. Rep. US Dept. Energy 40, 4 (2004)
2. Makarov, Y., et al.: Blackout prevention in the United States, Europe, and Russia. Proc. IEEE 93(11), 1942–1955 (2005)
3. Misra, S., et al.: Fast and robust determination of power system emergency control actions, [https://arxiv.org/abs/1707.07105](https://arxiv.org/abs/1707.07105) (2019)
4. Begovic, M., et al.: Wide-area protection and emergency control. Proc. IEEE 93(5), 876–891 (2005)
5. Zhang, Y., et al.: Online power system dynamic security assessment with incomplete PMU measurements: A robust white-box model. IET Gener., Transm., Distrib. 13(5), 662–668 (2019)
6. Tan, B., et al.: A deep imbalanced learning framework for transient stability assessment of power system. IEEE Access 7, 81759–81769 (2019)
7. Rashidi, M., Farijah, E.: LES based framework for transient instability prediction and mitigation using PMU data. IET Gener. Transm. Distrib. 10(14), 3431–3440 (2016)
8. Dan, H., et al.: Wide-area measurement system-based model-free approach of post-fault rotor angle trajectory prediction for on-line transient instability detection. IET Gener. Transm. Distrib. 12(10), 2425–2435 (2018)
9. Yan, R., et al.: Fast transient stability batch assessment using cascaded convolutional neural networks. IEEE Trans. Power Syst. 34, 2802–2813 (2019)
10. Wang, B., et al.: Power system transient stability assessment based on big data and the core vector machine. IEEE Trans. Smart Grid 7(5), 2561–2570 (2016)
11. Tang, Y., et al.: Hybrid method for power system transient stability prediction based on two-stage computing resources. IET Gener. Transm. Dis. 12(8), 1697–1703 (2018)
12. Li, N., et al.: Transient stability assessment of power system based on sg smoother and factorization algorithm. IEEE Access 8, 28403–28414 (2020)
13. Yu, J.J.Q., et al.: Intelligent time-adaptive transient stability assessment system. IEEE Trans. Power Syst. 33(1), 1049–1058 (2016)
14. Shi, Z., et al.: Convolutional neural network-based power system transient stability assessment and instability mode prediction. Appl. Energy 263, 114586 (2020)
15. Gomez, F.R., et al.: Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. IEEE Trans. Power Syst. 26(3), 1474–1483 (2011)
16. Idehen, I., et al.: Large-scale generation and validation of synthetic PMU data. IEEE Trans. Smart Grid 11(5), 4290–4298 (2020)
17. Jones, L., Brown, D.: PMU data feature considerations for realistic, synthetic data generation. In: 2019 North American Power Symposium (NAPS), pp. 1–6. IEEE, Piscataway (2019)
18. Wu, M., Xie, L.: Online identification of bad synchrophasor measurements via spatio-temporal correlations. In: Proceedings of Power Systems Computation Conference, pp. 1–7. IEEE, Piscataway (2016)
19. Wu, M., Xie, L.: Online detection of low-quality synchrophasor measurements: A data driven approach. IEEE Trans. Power Syst. 99, 1–1 (2016)
20. Wang, Y., et al.: An adaptive filters based PMU algorithm for both steady-state and dynamic conditions in distribution networks. Int. J. Electr. Power Energy Syst. 117, 105714 (2020)
21. Ren, J., et al.: Characterizing dynamic behavior of PMUs using step signals. Eur. Trans. Elect. Power 21(4), 1496–1508 (2011)
22. Ren, J., Kezunovic, M.: An adaptive phasor estimator for power system waveforms containing transients. IEEE Trans. Power Deliv. 27, 735–745 (2012)
23. Dhar, R.N.: Computer Aided Power System Operation and Analysis. Tata McGraw-Hill Publishing Co., New Delhi (1982)
24. DSATools Dynamic Security Assessment Software, http://www.dsatools.com/
25. Zhang, R., et al.: Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system. IET Gener. Transm. Distrib. 9(3), 296–305 (2015)
26. Mahdi, M., Genc, V.M.I.: Artificial neural network based algorithm for early prediction of transient stability using wide area measurements. In: Proceedings of 5th International Istanbul Smart Grid and Cities Congress and Fair, pp. 17–21. IEEE, Piscataway (2017)
27. Gou, B.: Generalized integer linear programming formulation for optimal PMU placement. IEEE Trans. Power Syst. 23(3), 1099–1104 (2008)

**How to cite this article:** Behdadnia T, Yaslan Y, Genc V. M. I. A new method of decision tree based transient stability assessment using hybrid simulation for real-time PMU measurements. *IET Gener Transm Distrib.* 2021;15:678–693. https://doi.org/10.1049/gtd2.12051

**APPENDIX**

To clarify our claim about the difference between the proposed method and the conventional method of time-series data arrangements and to illustrate their characteristics, a simple example is given in this section. Consider two PMUs (PMU₁ and PMU₂) that are located at two different points with a large electrical distance in between. The power system is subjected to two different contingencies. In the first scenario, the disturbance occurs nearby PMU₁. It is clear that the size of voltage drop (and rise) and subsequently the length of post-fault settling time of PMU₁ is expected to be larger due to a very small electrical distance from the fault location (see Section 3). However, because of the large electrical distance, the size of voltage drop (and rise) and the length of settling time is expected to be shorter for PMU₂. In essence, \( T^p_{\min} = \tau_{\max} \) can be attributed to PMU₁ and \( T^p_{\max} = \tau_{\min} \) can be attributed to PMU₂, where \( \tau_{\max} \gg \tau_{\min} \).

In the second scenario, the same type of fault occurs nearby PMU₂. In this case, with the similar arguments, \( T^p_{\max} = \tau_{\max} \) can be attributed to PMU₂ and \( T^p_{\min} = \tau_{\min} \) can be attributed to PMU₁ (in contrast to the previous scenario). By assuming \( \tau_{\max} = 5 \text{s} \) and \( \tau_{\min} = 1 \), the feature space matrix can be built and arranged as \( FS_{MCV} \) and \( FS_{MCV}^{new} \) based on (8) and (10), respectively. Figure A.1 shows the constructed matrices. It can be seen that, by removing erroneous measurements (marked in red) some of the informative and accurate data (marked in blue) are removed unintentionally from the dataset, which causes to reduce the informativity of the feature space.
Figure A.1: An example of matrices of datasets arranged with conventional and new methods.

\[ FSM_{CV} = \begin{bmatrix}
Y_{(1,1)}^1 & Y_{(1,1)}^2 & Y_{(1,1)}^3 & Y_{(1,1)}^4 & Y_{(1,1)}^5 & Y_{(1,1)}^6 & Y_{(1,1)}^7 \\
Y_{(2,1)}^1 & Y_{(2,1)}^2 & Y_{(2,1)}^3 & Y_{(2,1)}^4 & Y_{(2,1)}^5 & Y_{(2,1)}^6 & Y_{(2,1)}^7 \\
Y_{(2,2)}^1 & Y_{(2,2)}^2 & Y_{(2,2)}^3 & Y_{(2,2)}^4 & Y_{(2,2)}^5 & Y_{(2,2)}^6 & Y_{(2,2)}^7
\end{bmatrix} \]

Time-series Measurements of PMUs with the Same ID Number (ID No.1)

\[ FSM_{New} = \begin{bmatrix}
Y_{(1,1)}^1 & Y_{(1,1)}^2 & Y_{(1,1)}^3 & Y_{(1,1)}^4 & Y_{(1,1)}^5 & Y_{(1,1)}^6 & Y_{(1,1)}^7 \\
Y_{(2,1)}^1 & Y_{(2,1)}^2 & Y_{(2,1)}^3 & Y_{(2,1)}^4 & Y_{(2,1)}^5 & Y_{(2,1)}^6 & Y_{(2,1)}^7 \\
Y_{(2,2)}^1 & Y_{(2,2)}^2 & Y_{(2,2)}^3 & Y_{(2,2)}^4 & Y_{(2,2)}^5 & Y_{(2,2)}^6 & Y_{(2,2)}^7
\end{bmatrix} \]

Time-series Measurements of Equal-Rank PMUs (Rank No.1)

Time-series Measurements of PMUs with the Same ID Number (ID No.2)

Time-series Measurements of Equal-Rank PMUs (Rank No.2)