Event Identification Framework Based on Modal Analysis of Phasor Measurement Unit Data

Nima T.Bazargani, Gautam Dasarathy, Lalitha Sankar, and Oliver Kosut
School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ 85287

***Abstract*—This paper, propose a novel framework for event identification in power system based on modal analysis of PMU measurements using multi signal Matrix Pencil method. Modal analysis is conducted on each PMU measurement, including positive sequence voltage magnitude (VPm) and angle (VPa), positive sequence current magnitude (IPm), and angle (IPa), frequency (f) and Rate of Change of Frequency (DF) obtained from multiple locations in the grid. Then, each event is described as a set of features based on the modal analysis results. However, using various measurements obtained from multiple PMUs to extract features will result in the high dimensionality problem which may greatly degrade the performance of many machine learning algorithms. Hence, various feature selection techniques (Filter based and wrapper methods) are implemented in order to choose the best subset of features. Then, the performance of five well known classifiers (including Support Vector Machine (SVM) with polynomial kernel, SVM with rbf kernel, Decision tree, logistic regression and random forest) are compared based on the low-dimensional feature space obtained from each feature selection technique. Finally, we used a bootstrap re-sampling technique for a fair evaluation of each model accuracy.**

nomenclature

i	Index corresponding to the i^{th} PMU
$X \in \mathbb{R}^n$	Column vector of states variables
y_i	PMU measurement at bus i
λ_k	k^{th} system mode (pole in transfer function)
p	Order of the model
R_k	Residues corresponding to mode k
n , N	Sample number, total number of samples
m	Total number of PMUs
$Z_k^{(i)}$	Modes k of i^{th} PMU obtained from Modal analysis
$[Y_i]_p^{N-1}$	Column vector consists of i^{th} PMU measurements from sample p to $N-1$
$\hat{Y}_i = \mathbb{H}\{[Y_i]_0^{N-1}\}$	Hankel matrix consists of i^{th} PMU measurements from sample 0 to $N-1$
L	Pencil parameter in Matrix Pencil Method

I. Introduction

POWER systems are prone to various types of events (i.e. Line trips, Generation loss, etc) and identifying these events in real time can help operators in the control center to monitor, classify and analyze power system events and take suitable control actions, to ensure the system's reliability. However, due to system complexity, identification of these events is not an easy task. Increasing penetration of Phasor Measurement Units (PMU) in transmission systems, can help us to surmount the real time event identification issue by providing rich real-time data which can be used to monitor power system dynamics.

PMU provide time-synchronize voltage and current phasor measurements from various points of power system with high sampling rate. Although high sampling rate of PMUs can potentially help Independent System Operator (ISO) for Wide Area Monitoring Protection and Control (WAMPAC), it also presents challenges in the storage and analysis of massive scale PMU data [1]. On the other hand, power systems are inherently nonlinear and due to their complexity, in many cases, it is not possible to provide an accurate and sufficiently low order dynamical model [2]. To address these challenges, modal analysis has been widely used in power systems to extract modal information directly from the system response to a perturbation obtained from PMU measurements. There are various methods for modal analysis, each one having its own advantages and disadvantages. Considering robustness of matrix pencil method against noise compared with other modal analysis techniques (i.e. Prony's method), this method is used in this paper as the main tool for modal analysis (for more details on comparison between Matrix Pencil Method, Prony's analysis and Dynamic Mode Decomposition, we refer readers to [3], [4], and [5]).

Various studies have been conducted to provide a solution for real time event identifications problem based on extracted features obtained from modal analysis. Reference [1] proposed a feature based grid event classification approach where disturbance events are detected in the PMU data based on the extracted features from time-series data (i.e. oscillation frequency, participation factor, damping factor). They use extracted system modes from measured frequency signal for clustering grid events. Reference [6] proposed a robust and automated real-time monitoring system for detecting grid oscillations and analyzing their mode shapes obtained from single signal based Matrix Pencil method. However, in case of using single signal based Matrix Pencil method (i.e. [6]), one may need to apply Matrix Pencil on individual PMU signals and then classify PMU measurement channels with similar oscillation modes for further processing which often can be computationally expensive. Moreover, the classification result may not accurately represent the true underlying modes.

Another issue for identifying events in real time arises due to the high dimensionality of PMU data. Power system dynamical response can be observed in various PMU measurements (i.e. positive sequence voltage and current phasors, frequency and Rate of Change of Frequency) and thus, a key question is which of these signals can be used for modal analysis to improve the accuracy of event classification.

To address aforementioned issues, for each set of measurements (positive sequence voltage magnitude and angles, positive sequence current magnitude and angle, frequency and Rate of Change of Frequency) obtained from PMUs installed in multiple locations in the grid, we apply Multi-signal Matrix Pencil method to find a single set of modes which best represent the underlying dynamical behavior of the system observed in that measurement signal. The order of the underlying dynamical model is approximated based on the Low rank property of Hankel matrix obtained from PMU measurements.

Then, using the available PMU data for historical events, each event (line trip and generation loss) can be described as a set of d features obtained from Matrix Pencil method. However, using various measurements from multiple PMUs to extract features will result in the high dimensionality problem. When dimension d is high (compared to the number of samples), we may assume that only small number of features contribute to the prediction variable [7]. In this paper, various feature selection techniques (Filter based and wrapper models) are implemented in order to choose the best subset of features.

Filter based methods select a subset of features from high dimensional data sets without using a learning algorithm. These methods rank the features based on some statistical measure and then select only $d' < d$ features with highest rank. In this paper five measures (one-way ANOVA [1] F-test [8], Sure Independence Screening [7], Mutual Information [9], [10], Pearson correlation [9], and Kendall correlation [11]) are used to quantify relevance of features to the target variable. Due to the limited number of samples in our data set, we applied a bootstrap strategy to select the most important features.

To compare performance of filter approach with wrapper approach, a sequential forward selection (SFS) technique is implemented in order to find the most important features. SFS is a greedy search based algorithm which uses a specific classification technique to find a desired number of important features by adding one feature at the time based on the classifier performance. The resulting optimal set of features obtained from each of these feature selection techniques are used to validate the performance of five classification methods: Support Vector Machine (SVM) with polynomial kernel (SVM_poly), SVM with rbf kernel (SVM_rbf), Decision tree (DT), logistic regression (LR) and random forest (RF). We used a bootstrap re-sampling technique for a fair evaluation of each model accuracy.

The remainder of the paper is organized as follows. Section II briefly explains the theory behind modal analysis and Matrix Pencil method. The proposed framework for feature extraction from time series PMU data is presented in Section IV. Section VI describes the validation methodology which is used to evaluate the performance of models based on different feature selection techniques. VII applies the proposed framework and illustrates the simulation results on real PMU data. Finally, Section VIII concludes the paper.

[1] Analysis Of Variance

II. Mathematical Formulation of Modal Analysis

A. Linear State Space Model in Power System Small Signal Stability

The power system is a high-order nonlinear system. However, for small perturbations, we may assume that the system is linear around its operation point (or the equilibrium point). For a Single output, the state space can be defined as [12]:

$$\begin{aligned} \dot{X} &= AX + bu \\ y &= cX \end{aligned} \tag{1}$$

Where $A \in \mathbb{R}^{n_s \times n_s}$, $X \in \mathbb{R}^{n_s}$ where n_s is the number of states, and $b, c, u,$ and $y \in \mathbb{R}$. After some calculation and for an impulse input, transfer function can obtained as:

$$G(s) = \frac{y(s)}{u(s)} = c(sI - A)^{-1}b \tag{2}$$

Then the transfer function which for an impulse input is equal to the output signal, can be written as a fraction of zeros and poles and further can be simplified as follow:

$$G(s) = k\frac{(s-Z_1)(s-Z_2)...(s-Z_z)}{(s-P_1)(s-P_2)...(s-P_p)} = \sum_{k=1}^{p} \frac{R_k}{s-\lambda_k} \tag{3}$$

In general, each output signal i can be written as the summation of a set of poles and the corresponding residues as follow:

$$y_i(t) = \sum_{k=1}^{p} R_k^{(i)} exp(\lambda_k^{(i)} t) \quad , i = 1, 2, ..., m \tag{4}$$

When $y_i(t)$ is sampled at constant sampling rate, T_s, in discrete form, we have:

$$y_i(n) = \sum_{k=1}^{p} R_k^{(i)} (Z_k^{(i)})^n \tag{5}$$

where n is the sample number and for each mode $k = 1, \ldots, p$:

$$\underbrace{Z_k^{(i)}}_{\text{Mode } k} = exp(\lambda_k^{(i)} T_s) \quad , \lambda_k^{(i)} = \underbrace{\sigma_k^{(i)}}_{\text{damping ratio}} \pm j \underbrace{\omega_k^{(i)}}_{\text{angular frequency}} \tag{6}$$

Note that, for each PMU signal $i = 1, \ldots, m$, residues or mode shapes $R_k^{(i)}$, are defined by it's magnitude $|R_k^{(i)}|$ and angle $\theta_k^{(i)}$.

In a large power system as a multi-output system, multiple signals will oscillate at the same modes. Because of noise and non-linearity effects, individual analysis of measured signal often provides varying modal estimate. Hence, one goal is to analyze multiple measured signals simultaneously and obtain one optimum set of mode estimates which can accurately model the dynamic behavior of the system. To this end, we use Multi-signal Matrix pencil to obtain one optimum set of modes which best represent the underlying dynamical behavior of a set of measurement. So, instead of finding $Z_k^{(i)}$ for each signal i, we would have a single set of modes Z_k as follow:

$$y_i(n) = \sum_{k=1}^{p} R_k^{(i)} (Z_k^{(i)})^n \Rightarrow y_i(n) \approx \sum_{k=1}^{p} R_k^{(i)} (Z_k)^n \tag{7}$$

After finding a single set of Modes, i.e. $Z_1, \dots, Z_p$, then we can reconstruct each signal by finding the residues $R_k^{(i)}$ for each signal $i = 1, \dots, m$ and $k = 1, \dots, p$ using Eq. 8.

$$\underbrace{\begin{bmatrix} y_i(0) \\ y_i(1) \\ \vdots \\ y_i(N-1) \end{bmatrix}}_{[Y_i]_0^{N-1}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ Z_1 & Z_2 & \cdots & Z_p \\ \vdots & \vdots & \ddots & \vdots \\ Z_1^{N-1} & Z_2^{N-1} & \cdots & Z_p^{N-1} \end{bmatrix} \begin{bmatrix} R_1^{(i)} \\ R_2^{(i)} \\ \vdots \\ R_p^{(i)} \end{bmatrix} \tag{8}$$

Details on multi-signal Matrix Pencil and order approximation of the model can be found in Section III.

III. Matrix Pencil Method

For signal i, we construct the Hankel matrix as follow:

$$\hat{Y}_i = \mathbb{H}\{[Y_i]_0^{N-1}\} = \underbrace{\begin{bmatrix} y_i(0) & y_i(1) & \cdots & y_i(L) \\ y_i(1) & y_i(2) & \cdots & y_i(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ y_i(N-L-1) & y_i(N-L) & \cdots & y_i(N-1) \end{bmatrix}}_{(N-L)\times(L+1)} \tag{9}$$

where $[Y_i]_0^{N-1} = \begin{bmatrix} y_i(0), & y_i(1), & \cdots, & y_i(N-1) \end{bmatrix}^T$. We define $\hat{Y}_i^{(1)}$ as a matrix consists of first L columns of $\hat{Y}_i$ and $\hat{Y}_i^{(2)}$ as a matrix consists of last L columns of $\hat{Y}_i$. Then, one can write $\hat{Y}_i^{(1)}$ and $\hat{Y}_i^{(2)}$ as:

$$\hat{Y}_i^{(1)} = \mathbf{Z_L R Z_R} \quad , \quad \hat{Y}_i^{(2)} = \mathbf{Z_L R Z_0 Z_R} \tag{10}$$

where

$$\mathbf{Z_L} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ Z_1 & Z_2 & \cdots & Z_p \\ \vdots & \vdots & \ddots & \vdots \\ Z_1^{N-L-1} & Z_2^{N-L-1} & \cdots & Z_p^{N-L-1} \end{bmatrix}}_{(N-L)\times p}$$

$$\mathbf{Z_R} = \underbrace{\begin{bmatrix} 1 & Z_1 & \cdots & Z_1^{L-1} \\ 1 & Z_2 & \cdots & Z_2^{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Z_p & \cdots & Z_p^{L-1} \end{bmatrix}}_{p\times L}$$

$$\mathbf{Z_0} = diag(Z_1, Z_2, ..., Z_p)$$

$$\mathbf{R} = diag(R_1, R_2, ..., R_p)$$

Then, the matrix pencil is defined as :

$$\hat{Y}_i^{(2)} - \lambda \hat{Y}_i^{(1)} = \mathbf{Z_L R}(\mathbf{Z_0} - \lambda \mathbf{I})\mathbf{Z_R} \tag{11}$$

Now, when $p \le L \le N-p$, rank of $\hat{Y}_i^{(2)} - \lambda' \hat{Y}_i^{(1)}$ is p if $\lambda' \neq Z_k$. However, for any $\lambda' = Z_k$, the k^{th} row of $\mathbf{Z_0} - \lambda' \mathbf{I}$ becomes zeros and the rank of $\hat{Y}_i^{(2)} - \lambda' \hat{Y}_i^{(1)}$ is reduced by 1. Therefore, one can prove that the parameter Z_k's are the generalized eigenvalues of $(\hat{Y}_i^{(1)})^\dagger \hat{Y}_i^{(2)}$.

A. Multi Signal Matrix Pencil and Low rank property of Hankel Matrix

Matrix Pencil Method may be extended to find a set of modes which best represent the underlying dynamical behavior of a set of measurement. So, for each PMU measurement $i = 1, 2, \dots, m$, one can vertically concatenate each Hankel matrix $\mathbb{H}\{[Y_i]_0^{N-1}\}$ as follow:

$$\mathbb{H}\{[\mathbf{Y}]_0^{N-1}\} = \underbrace{\begin{bmatrix} \mathbb{H}\{[Y_1]_0^{N-1}\} \\ \vdots \\ \mathbb{H}\{[Y_i]_0^{N-1}\} \\ \vdots \\ \mathbb{H}\{[Y_m]_0^{N-1}\} \end{bmatrix}}_{m(N-L)\times(L+1)} \tag{12}$$

Now, The same method as the original Matrix Pencil for single signal can be applied to the large hankel matrix ($\mathbb{H}\{[\mathbf{Y}]_0^{N-1}\}$) to identify a set of poles Z_i, $i = 1, 2, \dots p$. Then for each signal i we can find the residues corresponding to each mode by solving Eq. (13). These residues later can be used to reconstruct each signal.

$$\underbrace{\begin{bmatrix} [Y_1]_0^{N-1} & \cdots [Y_m]_0^{N-1} \end{bmatrix}}_{N\times m} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ Z_1 & Z_2 & \cdots & Z_p \\ \vdots & \vdots & \ddots & \vdots \\ Z_1^{N-1} & Z_2^{N-1} & \cdots & Z_p^{N-1} \end{bmatrix}}_{N\times p} \underbrace{\begin{bmatrix} R_1^{(1)} & \cdots & R_1^{(m)} \\ R_2^{(1)} & \cdots & R_2^{(m)} \\ \vdots & \ddots & \vdots \\ R_p^{(1)} & \cdots & R_p^{(m)} \end{bmatrix}}_{p\times m} \tag{13}$$

B. Model Order Approximation and Low rank property of Hankel matrix

Assuming a linear dynamical system without an input, the state space model can be written as follow:

$$\begin{aligned} X(n+1) &= AX(n) \\ Y(n) &= CX(n) \end{aligned} \tag{14}$$

where $X(n) \in \mathbb{R}^{n_s}$ and $Y(n) \in \mathbb{R}^m$. Now, let $\bar{\mathbf{Y}} = [Y(0), Y(1), \dots, Y(N-1)] \in \mathbb{R}^{m\times N}$ contain all the measurements from sample 0 to $N-1$. Now, one can construct Hankel matrix from $\bar{\mathbf{Y}}$ as follow:

$$\mathbb{H}\{[\bar{\mathbf{Y}}]_0^{N-1}\} = \underbrace{\begin{bmatrix} Y(0) & Y(1) & \cdots & Y(L) \\ Y(1) & Y(2) & \cdots & Y(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ Y(N-L-1) & Y(N-L) & \cdots & Y(N-1) \end{bmatrix}}_{m(N-L)\times(L+1)} \tag{15}$$

Note that, since row operations do not change the row space, thus the row space and rank of this Hankel matrix would be the same as the $\mathbb{H}\{[\mathbf{Y}]_0^{N-1}\}$ obtained from Eq. (12).

Now, if A could be diagonalized as $A = \Phi\Lambda\Psi$ where $\Phi = [\phi_1, \dots, \phi_{n_s}]$ and $\Psi = [\psi_1, \dots, \psi_{n_s}]$ are the left and right eigenvectors and Λ is a diagonal matrix contains eigenvalues of matrix A, one can formulate measurement vector $Y(n)$ as:

$$Y(n) = CA^nX(0) = C\Phi\Lambda^n\Psi X(0) = \sum_{k=1}^{n_s} \lambda_k^n C\phi_k\psi_k^T X(0) \tag{16}$$

Assuming that only r ($r << n_s$) modes are enough to represent the underlying dynamical behavior and the remaining modes are associated with the noise in the signal, Eq. (16) can be approximated by

$$Y(n) = CA^nX(0) \approx \sum_{k=1}^{r} \lambda_k^n C\phi_k\psi_k^T X(0) \tag{17}$$

Then one can prove that both $\mathbb{H}\{[\mathbf{Y}]_0^{N-1}\}$ and $\bar{\mathbf{Y}}$ are rank r matrices [13].

Based on the above representation of $\mathbb{H}\{[\bar{\mathbf{Y}}]_0^{N-1}\}$ and our assumption on the order of the model, one can simply show that the order of the big Hankel matrix would be equal to the number of considered modes. In this setting, parameter (p) in the Matrix Pencil Method which represents the order of the model can be approximated as $p = r$. Note that the best way to approximate r is to take the SVD of $\mathbb{H}\{[\bar{\mathbf{Y}}]_0^{N-1}\}$ and then only consider the r largest singular values based on a predefined threshold.

Figure 1 demonstrates the sensitivity analysis of Hankel matrix approximation based on various values for parameters L and r in Matrix Pencil method.

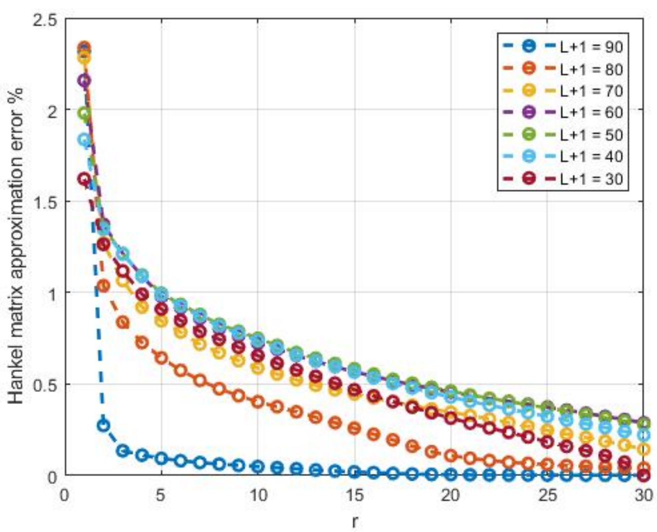

Fig. 1. Approximation error of Hankel matrix for different L and r parameters.

IV. Feature Extraction from PMU time series data

In order to capture dynamical response of the system after a disturbance, various PMU measurements can be used for modal analysis. One goal of this study is to decide which of these signals can potentially help us to classify events more accurately. To this end, two different approaches can be used to extract features from time series PMU data.

A. Method 1:

In this approach, modal analysis is conducted on each PMU measurement, including positive sequence voltage magnitude (VPm) and angle (VPa), positive sequence current magnitude (IPm), and angle (IPa), frequency (f) and Rate of Change of Frequency (DF) obtained from multiple locations in the grid. Then each event j can be described as a set of features obtained from Matrix Pencil method as follow:

$$\mathbf{x}_j = [x_{\text{VPm}}, x_{\text{VPa}}, x_{\text{IPm}}, x_{\text{IPa}}, x_{\text{f}}, x_{\text{DF}}]^T \tag{18}$$

where each x_s, $s \in$ {VPm, VPa, IPm, IPa, f, DF} consists of modal analysis results corresponding to the selected PMU measurement. For instance,

$$x_{\text{VPm}} = [\{\sigma_k\}_{k=1}^p, \{\omega_k\}_{k=1}^p, \{\{|R_k^{(i)}|\}_{i=1}^m\}_{k=1}^p] \tag{19}$$

consists of p angular frequencies, p damping factors and corresponding residues magnitudes for each PMU $i = 1, \ldots, m$ and mode $k = 1, \ldots, p$ which are obtained from multi signal Matrix Pencil method applied on voltage positive sequence measurements. Note that, residues magnitudes are the sorted values of $|R_k^{(i)}|$ in descending order. Based on our simulations, among all the installed PMUs in the grid (in our case $m \approx 500$), only small number of PMUs ($m' = 20 < m$) capture the dynamic behavior of the system after disturbances and therefore, just residues magnitude of those PMUs are included in the feature space.

B. Method 2:

In second approach, we apply multi signal Matrix Pencil on all types of PMU measurements (from multiple PMUs). For each historical line trip and generation loss event, all the recorded measurement (including VPm, VPa, IPm, IPa, f, DF) from multiple PMUs and for 100 samples (after the exact start of the event) are collected into a matrix. This matrix can be denoted as follow:

$$M = \begin{bmatrix} M_{\text{VPm}} \\ M_{\text{VPa}} \\ M_{\text{IPm}} \\ M_{\text{IPa}} \\ M_{\text{f}} \\ M_{\text{DF}} \end{bmatrix} \tag{20}$$

Then multi-signal Matrix Pencil is applied on matrix M to obtain a single set of dynamic modes which can best represent the dynamical behavior of the system observed in all the measurements.

In this setting, each event j can be described as a set of features obtained from Matrix Pencil method applied on matrix M as:

$$\begin{aligned}\mathbf{x}_j =&[\{\sigma_k\}_{k=1}^p, \{\omega_k\}_{k=1}^p, \{\{|R_{\text{VPm}}^{(i,k)}|\}_{i=1}^m\}_{k=1}^p, \\ &\{\{|R_{\text{VPa}}^{(i,k)}|\}_{i=1}^m\}_{k=1}^p, \{\{|R_{\text{IPm}}^{(i,k)}|\}_{i=1}^m\}_{k=1}^p, \\ &\{\{|R_{\text{IPa}}^{(i,k)}|\}_{i=1}^m\}_{k=1}^p, \{\{|R_{\text{f}}^{(i,k)}|\}_{i=1}^m\}_{k=1}^p, \\ &\{\{|R_{\text{DF}}^{(i,k)}|\}_{i=1}^m\}_{k=1}^p]^T\end{aligned} \tag{21}$$

where $\{\omega_k\}_{k=1}^p$ and $\{\sigma_k\}_{k=1}^p$ are angular frequency and damping ratio of $k'th$ mode ($k = 1, 2, \ldots, p$), respectively.

$\{\{|R_s^{(i,k)}|\}_{i=1}^{m}\}_{k=1}^{p}$ are residues magnitudes for each PMU $i = 1, \ldots, m$ and mode $k = 1, \ldots, p$ corresponding to various types of measurement $s \in \{$VPm, VPa, IPm, IPa, f, DF$\}$. [2]

Using either of above methods, each event j can be described as a set of features $\mathbf{x}_j \in \mathbb{R}^d$ and a label y_j which describes the type of event (line trips and generation loss events are labeled as -1 and 1, respectively). Note that number of features, d, obtained using each feature extraction method would be different. However, to avoid complexity in the notation, we use notation d to specify number of features obtained from each feature extraction method. We define our data set as $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N]^T \in \mathbb{R}^{N\times d}$ and $\mathbf{y} \in \{-1, 1\}^N$ where N here is the total number of historical events. The goal is to classify two types of events from the constructed data set $D = \{\mathbf{x}_j, y_j\}_{j=1}^N$.

V. Feature Selection

Based on the available historical events, total number $N = 70$ events (23 Generation loss events and 47 Line trip events) are used for feature extraction. For each event j, considering that only $p = 3$ modes are enough to represent the underlying dynamical behavior[3] observed in 6 different types of measurements, and only $m' = 20$ PMUs can capture the dynamic behavior of the system, then we would obtain $d = 6 \times (p + p + m' \times p) = 396$ features using the first feature selection method and $d = p + p + 6 \times m' \times p = 366$ using the second Method. High dimensionality of feature space compared to the number of samples, also know as curse of high-dimensionality, may greatly degrade the performance of many machine learning algorithms. This is because of the fact that many of these features can contain irrelevant or redundant information [9]. So, a necessary pre-processing step before using any classification algorithm is to select relevant and most informative features. Common statistical approaches for selecting features include filter methods, wrapper methods and embedded methods [14]. The focus of this paper is to compare performance of filter methods and wrapper method. Note that, for feature selection, we split our data into a train set D_{train} with $N_{\text{tr}} = 59$ samples and test data set D_{test} with $N_{\text{te}} = 11$ samples, and all the feature selection techniques are applied on the train data. The test data is kept for evaluation and comparison of the final model.

A. Filter Method

In order to evaluate the goodness of features for classification, five different approaches (including one-way ANOVA F-value [8], Sure Independence Screening [7], Mutual Information [9], [10], Pearson correlation [9], and Kendall correlation [11]) are implemented using off-the-shelf sklearn packages in Python to quantify the correlation between features and target class. Then, based on this definition, a feature is good if it is highly correlated to the target class.

Based on the correlation of each feature f_i and target variable $\mathbf{y}$, we sort the features and then remove all the features except d' the highest correlation with target variable. However, due to the small number of samples, we use bootstrap technique to generate B distinct data set by repeatedly re-sampling observations from the original data set with replacement, as shown in Fig. 2.

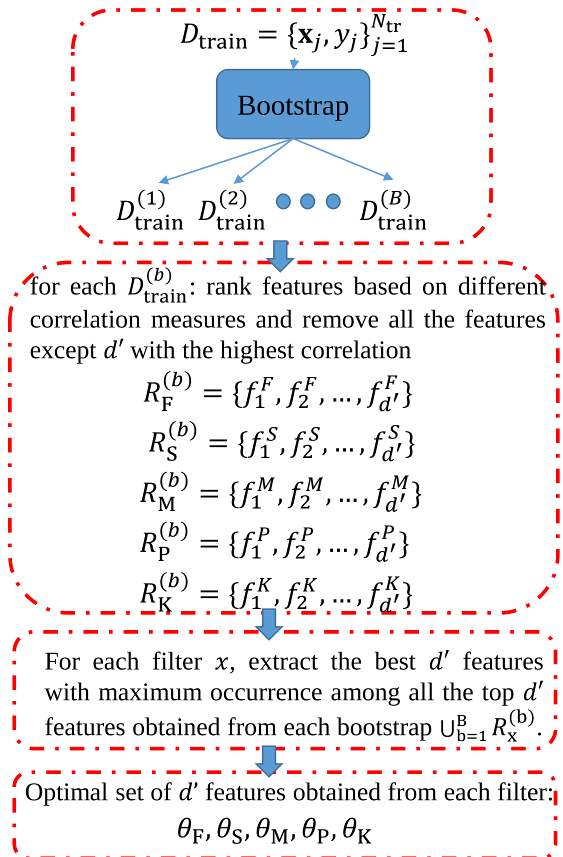

Fig. 2. Feature selection by bootstrap re-sampling from the original data set and based on different correlation measures. Note that, for each bootstrap b, $R_F^{(b)}$, $R_S^{(b)}$, $R_M^{(b)}$, $R_P^{(b)}$, and $R_K^{(b)}$ represent the ranking of features based on F-test, SIS, mutual information, Pearson correlation and Kendall correlation, respectively. Moreover, $R_x^{(b)}$ here represents ranked features obtained from each filter $x \in \{F, S, M, P, K\}$ and $\theta_{\text{F}}, \theta_{\text{S}}, \theta_{\text{M}}, \theta_{\text{P}}$, and θ_{K} are the optimal set of d' features after bootstrapping.

B. Wrapper Method

Unlike filter methods, wrapper methods use a specific classification method to select features, and then obtained set of selected features are used to train a model based on the same classification algorithm [14]. In this paper, a Sequential Forward Selection (SFS) is used for feature selection to reduce the dimension of feature space. SFS is an greedy search algorithm in which we start with an empty set of features $F = \emptyset$. In each iteration, we add one feature at a time which best improves the classifier performance until a feature subset of the desired size d' is reached [15]. Note that, due to the small number of samples, here we use a k-fold cross validation technique where we randomly divide the available set of train data into K equal-sized parts. We hold out one part for validation and we use the remaining $K - 1$ part to fit the model, and the fitted model is used to predict the responses for the observations in the validation set. This evaluation is done for each of K parts. Then, a candidate feature is

[2] Note that, residues magnitudes corresponding to each type of measurement are sorted in a descending order.

[3] The remaining modes are associated with the noise in the signal

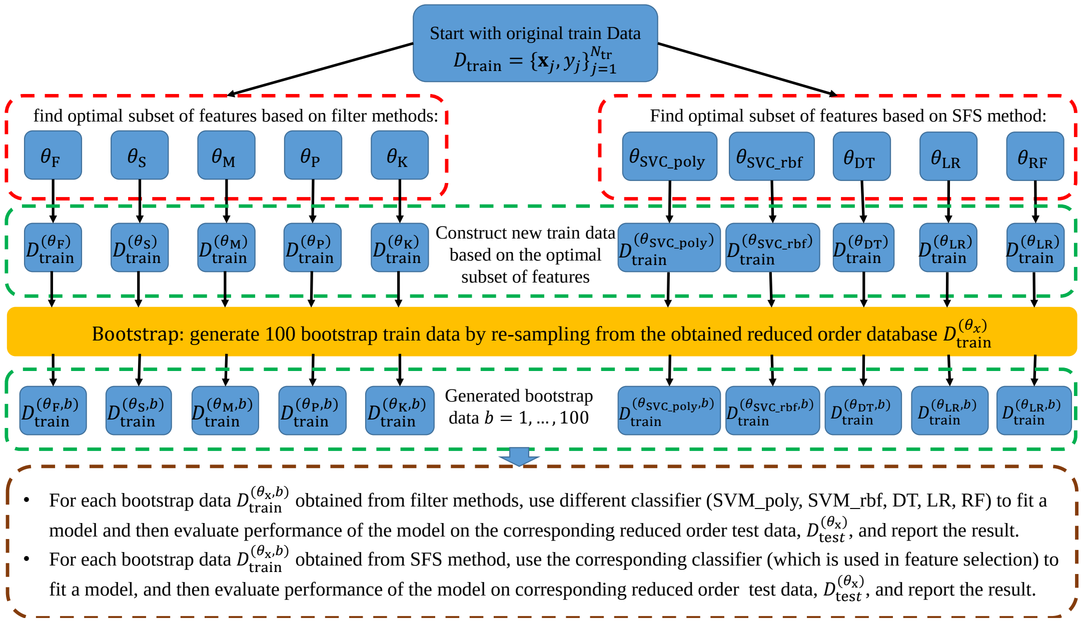

Fig. 3. Details of the proposed strategy to compare and validate the performance of each feature selection method using various classifiers.

evaluated based on its overall performance on hold out parts. To evaluate the performance of the classification we use the Area Under Curve (AUC) of Receiver Operator Characteristic (ROC) which illustrates the accuracy of a binary classifier based on various discrimination threshold. ROC plot shows the relation between true positive rate (also known as recall) against the false positive rate at various threshold settings. True Positive rate (TPR) and false positive rate (FPR) are defined as follow:

$$\text{TPR} = \frac{\text{\# of True Positives}}{\text{\# of True Positives + \# of False Negatives}} \quad (22)$$

$$\text{FPR} = 1 - \frac{\text{\# of True Negatives}}{\text{\# of True Negatives + \# of False Positives}} \quad (23)$$

ROC AUC value is bounded between 0 and 1. The closer AUC to 1, classifier has better ability to classify the positive and negative samples whereas an uninformative classifier yielding 0.5 AUC score.

For comparison purposes, five different classifiers, including SVM_poly, SVM_rbf, DT, LR and RF from off-the-shelf sklearn packages in Python are used to find the optimal subset of feature based on SFS method. It is clear that in this setting, using each classifier will result in a different subset of features. Obtained optimal set of d' features corresponding to each classifier are denoted as $\theta_{\text{SVM\_poly}}$, $\theta_{\text{SVM\_rbf}}$, θ_{DT}, θ_{LR}, and θ_{RF}. Selected features will be used to compare and evaluate the performance of each model on the test data D_{test}.

VI. Validation of Models Based on Different Feature Selection Methods

As explained in section V, we split our data into a train set D_{train} with $N_{\text{tr}} = 59$ samples and test data set D_{test} with $N_{\text{te}} = 11$ samples, and all the feature selection techniques are applied on the train data. Once the optimal subset of features are obtained, We us a bootstrap re-sampling technique for a fair evaluation of each model accuracy. Details of the proposed strategy is shown in Fig. 3. We start with the original train data set. Using various feature selection techniques explained in Section. V, we find the most important d' features ($d' << d$) as θ_{x} corresponding to the selected feature selection method, say x where x $\in$ {F, S, M, P, K, SVM_poly, SVM_rbf, DT, LR, RF}. Then, using the optimal obtained subsets of feature, we construct a reduced order $D_{\text{train}}^{(\theta_{\text{x}})}$ corresponding to each feature selection technique. Next, for each reduced order $D_{\text{train}}^{(\theta_{\text{x}})}$, we use bootstrapping to generate 100 bootstrap data set, denoted as $D_{\text{train}}^{(\theta_{\text{x}},b)}$ and $b = 1, \ldots, 100$. Finally, we use various classifiers to fit a model on each bootstrap data set and then we use the learned model to predict target variable in the corresponding reduced order $D_{\text{test}}^{(\theta_{\text{x}})}$.

VII. Simulation Results

In order to evaluate the performance of the proposed framework for event identification, we use real data [4] collected from ≈ 500 PMU, where each PMU provide time-synchronize VPm, VPa, IPm, IPa, f, and DF from various points of power system with sampling rate of 30 samples/second. For each historical line trip and generation loss event, specific recorded measurement (VPm, VPa, etc.) from multiple PMUs and for 100 samples after the exact start of the event, are collected in a matrix M. For instance, matrix M for VPm measurement can be denoted as M_{VPm} where each element $m_{i,j}$ ($i'th$ row and $j'th$ column) is the $j'th$ VPm samples obtained from $i'th$ PMU. Before any modal analysis, first we remove the mean or DC offset from each measurement i in the above matrix. Then using the constructed matrix, we apply multi signal Matrix Pencil method (the details are explained in appendix III) to obtain modal analysis result as shown in Eq. (19). Next, for each event j, similar approach is applied on other

[4]Due to confidentiality agreements, data files are not available for public release.

measurement matrices (i.e. $M_{\mathrm{VPa}}, M_{\mathrm{IPm}}, M_{\mathrm{IPa}}, M_{\mathrm{f}}$, and M_{DF}) and based on the modal analysis results on each of these matrices, we can describe event j as a set of features (see Eq. (21)). After extracting features for all of the 70 historical events, we construct our data set as $\mathbf{X}$ with corresponding event labels $\mathbf{y}$ (as explained in Section IV). Note that, as the last preprocessing step, our data set $\mathbf{X}$ is normalized based on Z-score normalization.

In order to estimate the number of required features, first we use a sequential forward selection (SFS) technique where based on the classifier performance, we add one feature at a time. Simulation results of SFS using LR classifiers for $d' = 1, \dots, 25$ are shown in Fig. **??** (Due to limited space, simulation results of SFS using other classifiers are shown in Appendix **??**). Note that, shaded blue area in the plots shows the confidence interval of ROC-AUC score based on K-fold cross validation. These result suggest that using only $d' \approx 15$ most important features would result in an acceptable accuracy in terms of ROC-AUC score. Therefore, for filter based feature selection methods, we aim to find $d' = 15$ most important features (as described in Section V). Using the obtained optimal set of d' feature from each feature selection technique, first we obtain the reduced order $D^{(\theta_{\mathrm{x}})}_{\mathrm{train}}$, and then we use bootstrapping to generate 100 bootstrap data set. Next, We use various classifiers to fit a model on each bootstrap data set and based on the learned model, evaluate the accuracy of model on $D^{(\theta_{\mathrm{x}})}_{\mathrm{test}}$ in terms of ROC-AUC score.Performance of each classifier based on the various feature selection method are summarized in Table I. Note that, Accuracy of models are evaluated based on ROC-AUC score and the 5% and 95% confidence interval. The best feature selection technique for each classifier is highlighted in bold font.

TABLE I
PERFORMANCE OF EACH CLASSIFIER BASED ON THE VARIOUS FEATURE SELECTION METHOD AND USING THE 1ST FEATURE EXTRACTION METHOD.

Feature Selection Method			Filter methods					Wrapper
			F	S	M	P	K	SFS
Classifier	SVM _poly	AUC	0.69	0.61	0.63	0.62	0.69	**0.72**
		5% CI	0.25	0.12	0.33	0.16	0.29	**0.48**
		95% CI	0.96	1.0	0.79	0.95	0.75	**0.83**
	SVM _rbf	AUC	0.53	0.69	0.70	0.71	0.69	**0.74**
		5% CI	0.25	0.16	0.45	0.25	0.53	**0.5**
		95% CI	0.75	1.0	0.83	0.75	0.87	**0.91**
	DT	AUC	0.60	0.62	0.64	0.61	**0.66**	0.60
		5% CI	0.375	0.37	0.43	0.37	**0.43**	0.37
		95% CI	0.83	0.87	0.83	0.83	**0.87**	0.83
	LR	AUC	0.73	**0.93**	0.70	0.73	0.70	0.83
		5% CI	0.62	**0.78**	0.58	0.625	0.66	0.70
		95% CI	0.83	**1.0**	0.83	: 0.83	0.75	0.95
	RF	AUC	**0.79**	0.74	0.78	**0.79**	0.77	0.76
		5% CI	**0.56**	0.45	0.62	**0.56**	0.64	0.62
		95% CI	**0.95**	0.91	0.91	**0.95**	0.87	0.91

TABLE II
PERFORMANCE OF EACH CLASSIFIER BASED ON THE VARIOUS FEATURE SELECTION METHOD AND USING THE 2ND FEATURE EXTRACTION METHOD.

Feature Selection Method			Filter methods					Wrapper
			F	S	M	P	K	SFS
Classifier	SVM _poly	AUC	0.69	0.68	0.51	0.65	0.66	**0.75**
		5% CI	0.18	0.24	0.2	0.16	0.2	**0.54**
		95% CI	0.87	0.75	0.75	0.83	0.83	**0.87**
	SVM _rbf	AUC	0.59	0.54	0.61	0.59	0.55	**0.64**
		5% CI	0.33	0.25	0.25	0.29	0.2	**0.18**
		95% CI	0.79	0.79	0.75	0.75	0.81	**0.81**
	DT	AUC	0.63	0.61	0.58	0.64	0.71	**0.72**
		5% CI	0.37	0.41	0.41	0.37	0.41	**0.37**
		95% CI	0.77	0.77	0.87	0.81	0.83	**0.93**
	LR	AUC	0.77	**0.89**	0.66	0.76	0.85	0.81
		5% CI	0.62	**0.71**	0.45	0.62	0.62	0.68
		95% CI	0.87	**1**	0.87	0.87	0.79	0.94
	RF	AUC	0.71	0.75	0.77	0.71	**0.87**	0.82
		5% CI	0.6	0.49	0.54	0.6	**0.66**	0.62
		95% CI	0.87	0.89	0.87	0.87	**0.91**	0.87

VIII. CONCLUSION

This paper, proposed a novel framework for event identification in power system based on modal analysis of PMU measurements using multi signal Matrix Pencil method. Modal analysis is conducted on various PMU measurements. The result of modal analysis is used to extract features from time series PMU data. Various feature selection techniques are implemented to find the optimal subset features. A bootstrap based method is used to compare and validate the performance of each feature selection method using various classifiers. Based on the simulation result, and the selected features from SIS filter, LR classifier have the best identification accuracy. Another interesting observation is that, in most cases, angular frequency and first few residues magnitudes corresponding to the first mode of VPm and f signal are included in the set of optimal features obtained from various methods. Future work will focus on Ensemble Feature Selection techniques which makes use of multiple feature selection methods.

REFERENCES

[1] S. A. R. Konakalla and R. A. de Callafon, “Feature based grid event classification from synchrophasor data,” *Procedia Computer Science*, vol. 108, pp. 1582–1591, 2017.

[2] M. L. Crow and A. Singh, “The matrix pencil for power system modal extraction,” *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 501–502, 2005.

[3] L. L. Grant and M. L. Crow, “Comparison of matrix pencil and prony methods for power system modal analysis of noisy signals,” in *2011 North American Power Symposium*, 2011, pp. 1–7.

[4] W. Trinh and T. Overbye, “Comparison of dynamic mode decompositon and iterative matrix pencil method for power system modal analysis,” in *2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*, 2019, pp. 1–6.

[5] C. M. HWEE, “Matrix pencil method as a signal processing technique performance and application on power systems signals,” Ph.D. dissertation, 2013.

[6] D. Bian, Z. Yu, D. Shi, R. Diao, and Z. Wang, “A robust low-frequency oscillation detection and analysis (lfoda) system with innovative ensemble filtering for real-time grid operation,” *CSEE Journal of Power and Energy Systems*, 2019.

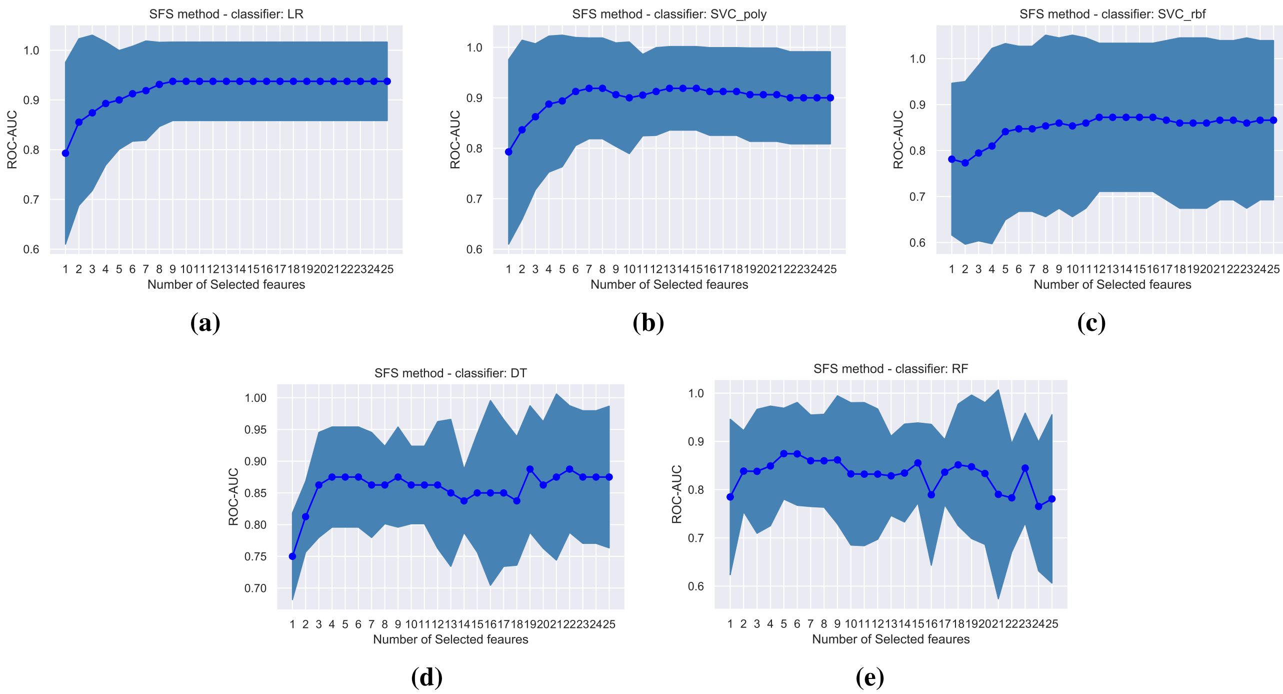

Fig. 4. Simulation results of SFS using **(a)** Logistic regression **(b)** SVC_poly **(c)** SVC_rbf **(d)** Decision Tree **(e)** Random Forest classifier and using first method for feature extraction

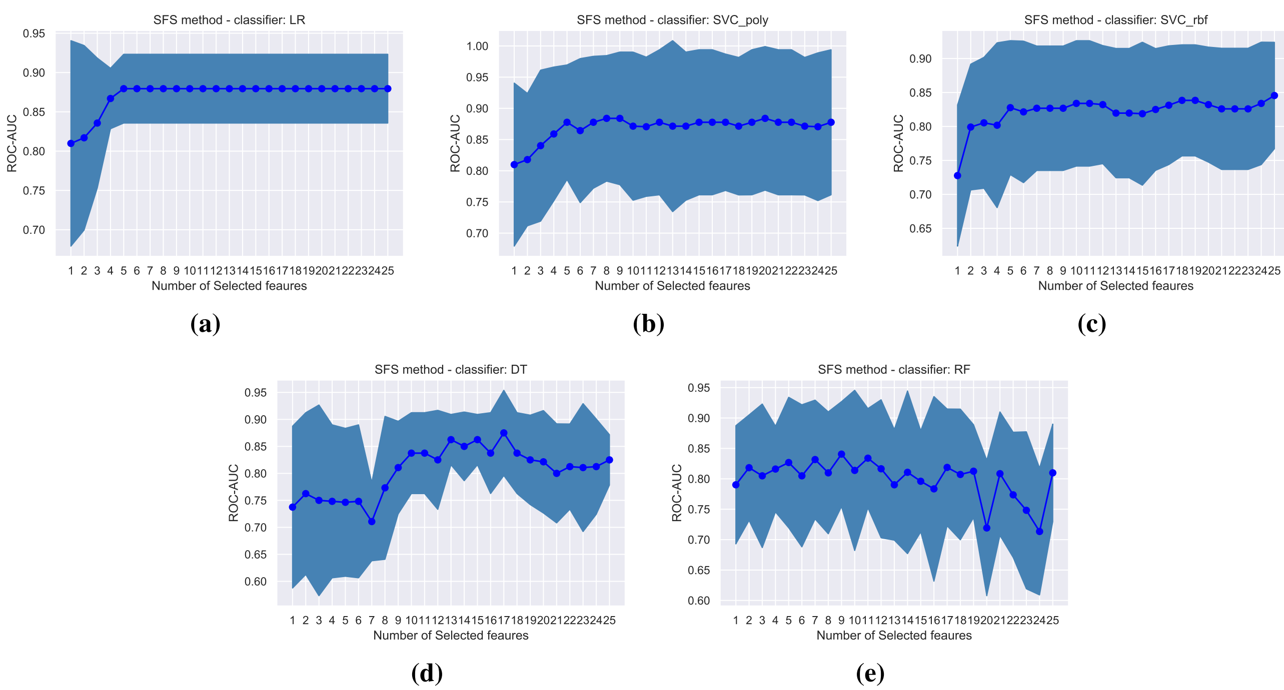

Fig. 5. Simulation results of SFS using **(a)** Logistic regression **(b)** SVC_poly **(c)** SVC_rbf **(d)** Decision Tree **(e)** Random Forest classifier and using second method for feature extraction

[7] J. Fan and J. Lv, “Sure independence screening for ultrahigh dimensional feature space,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 5, pp. 849–911, 2008.
[8] M. Sheikhan, M. Bejani, and D. Gharavian, “Modular neural-svm scheme for speech emotion recognition using anova feature selection method,” *Neural Computing and Applications*, vol. 23, no. 1, pp. 215–227, 2013.
[9] L. Yu and H. Liu, “Feature selection for high-dimensional data: A fast correlation-based filter solution,” in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 856–863.
[10] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys, “Robust biomarker identification for cancer diagnosis with ensemble feature selection methods,” *Bioinformatics*, vol. 26, no. 3, pp. 392–398, 2010.
[11] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, “Threshold-based feature selection techniques for high-dimensional bioinformatics data,” *Network modeling analysis in health informatics and bioinformatics*, vol. 1, no. 1-2, pp. 47–61, 2012.
[12] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*. McGraw-hill New York, 1994, vol. 7.
[13] S. Zhang, Y. Hao, M. Wang, and J. H. Chow, “Multichannel hankel matrix completion through nonconvex optimization,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 4, pp. 617–632, 2018.
[14] T. van der Ploeg and E. W. Steyerberg, “Feature selection and validated predictive performance in the domain of legionella pneumophila: a comparative study,” *BMC Research Notes*, vol. 9, no. 1, pp. 1–7, 2016.
[15] M. Kudo, P. Somol, P. Pudil, M. Shimbo, and J. Sklansky, “Comparison of classifier-specific feature selection algorithms,” in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2000, pp. 677–686.