



# **Anomaly Detection in Electric Power Systems using Machine Learning Methods**

**Ambreen Khurram**

**Submitted in accordance with the requirements for the degree  
of Doctor of Philosophy**

**The University of Leeds  
Faculty of Engineering and Physical Sciences**

**March 2023**

# Abstract

The electric power system is a complex nonlinear system that functions in a dynamic environment and is frequently subjected to a wide range of small and large disturbances. Small disturbances occur continuously due to load changes, while large disturbances are often caused by faults (such as equipment malfunction, human error, or attacks) and then propagate through the system. Depending on the system operating conditions, such disturbances can lead to stability issues and, in the worst case, to blackouts. The onset of instability in power systems causes fluctuations in different physical properties of the system, with the most critical being the voltage and frequency. Monitoring the variation of these physical properties over time allows for extracting information about the stability status of the system.

This thesis aims to tackle power system stability concerns by creating real-time detection algorithms that rely on Phasor Measurement Units (PMUs). These algorithms serve as early warning systems and are valuable inputs for stabilizing control techniques. The algorithms in question focus on two types of stability issues: short-term oscillatory stability, which pertains to low-frequency interarea oscillations, and long-term voltage stability, which is related to gradual voltage collapse.

In the thesis, the first section covers Low-Frequency Oscillations (LFO) in the power grid. While typically well-damped, under-damped LFOs can pose a significant threat to the grid's stability, making it crucial to detect them early for real-time monitoring. One important aspect of analyzing oscillatory stability is determining the frequency and damping of critical oscillatory modes, which can be challenging due to closely spaced and noisy natural modes in PMU signals. To address this issue, the thesis proposes a method for detecting LFO using the Empirical wavelet transform, which adaptively extracts different signal modes through a wavelet filter bank.

---

The second part of the thesis focuses on long-term voltage stability (LTVS) in electric power systems, which can gradually deteriorate over time due to the grid's inability to meet demand. Factors such as insufficient reactive resources, load characteristics, and tap changer response can contribute to LTVS, but the thesis primarily examines the stressed power system caused by high active power demand from excessive load. For the real-time assessment of long-term voltage stability (LTVS), this study proposes an approach that utilizes data mining and machine learning methods to evaluate long-term voltage stability (LTVS). The proposed technique employs a feature ensemble method to predict the voltage stability margin (VSM).

# Intellectual Property

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2023 The University of Leeds, Ambreen Khurram

Signed

A handwritten signature in black ink, appearing to read "Ambreen".

# Acknowledgements

I want to thank my supervisors, Dr. Arief Gusnanto and Dr. Petros Aristidou, for their thoughtful guidance in conducting the research. They have influenced my view of the research process and instilled in me the importance of producing quality research with potential impact.

Dr. Petros Aristidou has continually provided his expertise on the power system, life experiences, and responsibility throughout the Ph.D. process. Most of all, I am indebted to his valuable guidance, constructive feedback, and comments from the beginning until this Ph.D. is completed. Working with Dr. Aristidou has been a wonderful experience for the past four years. His vast knowledge, strong research enthusiasm, and hard-working attitude have inspired me during my Ph.D. and will profoundly affect my career forever. This Ph.D. research could not have been finished without his support, guidance, and encouragement. I am fortunate to have such an excellent supervisor, mentor, and exciting friend.

Dr. Arief Gusnanto and his assistance and suggestions in research inspired me. I value his honest opinions, critics, calmness and clarity of advice during difficult times, and patience and understanding over the past several years. His extensive understanding of Machine Learning and remarkable enthusiasm for research has helped me to overcome many challenging situations in the research environment. I am indebted to have had advisors that gave me all the resources, guidance, and support I could ever need during the period that led up to this research work.

I also thank my colleagues, Agnes Nakiganda and Shahab Dehghan. They shared their experience and knowledge during my Ph.D. and gave me their friendly support and assistance.

Finally, I thank my husband for being patient throughout my Ph.D. as I devoted my time to research work. Thanks for all your support, understanding, and encouragement throughout all these years.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Motivation and research objectives . . . . .	6
1.1.1	Unveiling the Significance of Oscillatory stability in Electric Power Systems	7
1.1.2	Unveiling the Significance of Long-Term Voltage Stability in Electric Power Systems . . . . .	8
1.1.3	Shortcomings of existing methods . . . . .	9
1.1.4	Unleashing the Potential of ML Techniques . . . . .	10
1.2	Monitoring of power system stability . . . . .	11
1.2.1	Supervisory Control and Data Acquisition . . . . .	11
1.2.2	Wide Area Measurement Systems . . . . .	13
1.3	Classification of Power System Stability . . . . .	13
1.3.1	Rotor angle stability . . . . .	15
1.3.2	Voltage stability . . . . .	16
1.4	Thesis Contributions . . . . .	17
1.4.1	List of published papers . . . . .	19
1.5	Outline of the Thesis . . . . .	19
<b>2</b>	<b>Voltage stability analysis</b>	<b>22</b>
2.1	Literature Review . . . . .	23
2.1.1	Methods Based on Modal Analysis . . . . .	23
2.1.2	Methods Based on Loading Margin . . . . .	23
2.1.3	Methods based on Machine Learning . . . . .	28
2.2	Voltage stability margin . . . . .	33
2.3	Problem formulation . . . . .	34

2.3.1	Framework for Real-Time Prediction of Loadability . . . . .	35
2.4	Preparation of Training and Testing Data . . . . .	39
2.4.1	Overview of Data Generation Process . . . . .	39
2.4.2	Operating point generation . . . . .	39
<b>3</b>	<b>Ensemble Feature Selection</b>	<b>45</b>
3.1	Introduction . . . . .	45
3.2	Feature Selection Techniques . . . . .	48
3.2.1	F-Test . . . . .	48
3.2.2	Variance threshold . . . . .	49
3.2.3	Least absolute shrinkage and selection operator . . . . .	49
3.3	Classification Algorithms for VSM prediction . . . . .	50
3.3.1	Support Vector Machine . . . . .	52
3.3.2	Decision Tree Algorithm . . . . .	54
3.3.3	K-Nearest Neighbor . . . . .	55
3.4	Hyperparameter Optimization . . . . .	55
3.5	Case Study . . . . .	58
3.5.1	Data generation . . . . .	59
3.5.2	Structure of the Data . . . . .	65
3.6	Model Performance Evaluation . . . . .	67
3.6.1	Phase-1: performance of the classifier without feature selection . . . . .	68
3.6.2	Phase-2: performance comparison of the proposed feature selection algorithms with similar feature selectors . . . . .	71
3.6.3	Phase-3: performance of the feature-ensemble . . . . .	75
3.6.4	Phase-4: Robustness of the feature ensemble against noisy data . . . . .	76
3.6.5	Comparative Analysis: Feature Ensemble Method vs. State-of-the-Art Machine Learning Techniques . . . . .	79
<b>4</b>	<b>Online application of Machine Learning Model</b>	<b>83</b>
4.1	K-means Clustering . . . . .	85
4.1.1	Database Update . . . . .	86
4.2	Case Study . . . . .	87
4.2.1	Identifying Secure and Insecure Operating Points . . . . .	89

<b>5 Power system oscillations</b>	<b>94</b>
5.1 Power System Electromechanical Oscillations . . . . .	95
5.1.1 Classification of Power System Oscillations . . . . .	95
5.2 Estimation of power system modes . . . . .	97
5.2.1 Model based method . . . . .	97
5.2.2 Measurement based method . . . . .	99
5.3 Motivation for Damping Estimation . . . . .	102
5.4 Interarea Oscillations . . . . .	107
5.4.1 Modal Parameter Estimation of Interarea Oscillations . . . . .	107
<b>6 Estimation of low frequency electromechanical oscillation</b>	<b>109</b>
6.1 Introduction . . . . .	109
6.2 Proposed Scheme . . . . .	110
6.3 Theoretical Background . . . . .	111
6.3.1 Event Detection . . . . .	111
6.3.2 Principal Component Analysis . . . . .	112
6.3.3 Empirical Wavelet Transform . . . . .	113
6.3.4 Proposed Enhanced Empirical Wavelet Transform . . . . .	117
6.3.5 Selecting the signal with highest power . . . . .	119
6.3.6 Hilbert Transform . . . . .	120
6.3.7 Average parameters . . . . .	121
6.4 Online consideration . . . . .	122
6.5 Numerical results . . . . .	122
6.5.1 System model and case study description . . . . .	122
6.5.2 Case study 1(Stable Model): . . . . .	124
6.5.3 Case study 2 (Unstable Model) : . . . . .	131
6.5.4 Detection of false alarm . . . . .	133
6.6 Performance evaluation and comparison . . . . .	137
6.7 Conclusion . . . . .	138
<b>7 Estimating interarea oscillation with machine learning</b>	<b>139</b>
7.1 Introduction . . . . .	140
7.1.1 Fundamentals and proposed algorithm . . . . .	140

7.2	Methodology . . . . .	141
7.2.1	Data Generation for machine learning . . . . .	142
7.2.2	Feature Selection with LASSO . . . . .	144
7.2.3	Training and evaluation . . . . .	144
7.2.4	Online Assessment . . . . .	147
7.2.5	Database update condition . . . . .	147
7.3	Numerical results . . . . .	149
7.3.1	System model and case study description . . . . .	149
7.3.2	Data Preparation . . . . .	150
7.3.3	Comparative Performance Analysis: Empirical Wavelet Transform vs. other adaptive methods . . . . .	153
7.3.4	Feature Selection . . . . .	158
7.3.5	Model training and testing . . . . .	158
7.3.6	Database update condition . . . . .	162
7.4	Chapter Conclusion . . . . .	164
<b>8</b>	<b>Conclusion and future work</b>	<b>166</b>
8.1	Advancements and Contributions . . . . .	167
8.2	Demonstrating Superiority: Unveiling the Strengths of the Proposed Approach Over Existing Methods . . . . .	168
8.3	Future Work . . . . .	170
<b>References</b>		<b>171</b>
<b>A appendix-A: Nordic Test System</b>		<b>192</b>
<b>B Appendix-B: Parameter setting for machine learning model</b>		<b>194</b>

# List of Figures

1.1	Architecture of WAMS; where PMUs collect the data from various sources of power systems. The collected data sets are delivered to a local Phasor Data Concentrator (PDC), which transmits them to a master database called super-PDC. The consolidated data sets collected by super-PDC are fed into analytic applications such as state estimation, stability assessments, data visualization, real-time monitoring, and control. . . . .	12
1.2	Classification of power system stability. In orange the sub-categories of interest in this thesis [1] . . . . .	14
2.1	The P-V curve determines the loading margin of a power system. The margin between the voltage collapse point ( $P_{max}$ ) and the current operating point ( $P_0$ ) is used as the voltage stability criterion. If the voltage is on the upper side of the red dotted line then total stability is under normal operating conditions. Below the red line the system cannot maintain voltage stability . . . . .	35
2.2	An overview of the proposed framework is shown in Fig. 2, in which the planned framework is illustrated with the part enclosed by the dotted lines. It comprises of three stages: 1)Data generation 2) Offline Training and 2) Online Prediction and Update . . . . .	37
2.3	DBSCAN: Hourly load profile based on electricity consumption of end users. . .	40
2.4	Cluster created by DBCAN with MinPts = 3, black dots represent the noise, blue dots represents the border points, and pink dots represent the core points .	41
2.5	Directly reachable points: Point $q$ is directly reachable from point $p$ . . . . .	41
2.6	Directly density reachable: Point $q$ is directly density reachable from point $p$ through a chain of points $q_1, \dots, q_n$ . . . . .	42

2.7 Density connected: Point $q$ is density connected to point $p$ through point $o$ as $p$ and $q$ are density reachable from $o$ . . . . .	43
2.8 Silhouette coefficient example; $a$ represents average intra-cluster distance i.e the average distance between each point within a cluster while $b$ represents average inter-cluster distance i.e the average distance between all clusters . . . . .	43
3.1 Feature-ensemble model with different base selector and same classifier. . . . .	46
3.2 Structure of the Decision Tree . . . . .	54
3.3 Steps for hyperparameter optimization . . . . .	56
3.4 a) Elbow plot for K selection, b) Silhouette Score . . . . .	60
3.5 Cluster of operating points . . . . .	60
3.6 The P-V curves are obtained by gradually increasing all loads in the Central area. The blue curve shows when the system operates without a contingency. The margin is higher as the power is transferred without endangering the limits of transmission lines and the bus itself. After the contingency is applied (red curve and green curve), the reduction of VSM occurs. As noticed, in the contingency case, the critical level is reached more quickly, resulting in the loss of voltage stability. . . . .	64
3.7 Variation of voltage magnitude across all buses . . . . .	64
3.8 PV curve generated for initial load operating point . . . . .	65
3.9 Comparison of classifier performance . . . . .	69
3.10 Comparison of classifier performance . . . . .	70
3.11 Residual plot for the classifier performance . . . . .	70
3.12 Comparison of RMSE of feature selection algorithms . . . . .	72
3.13 Time consumption by classifiers . . . . .	74
3.14 Comparison of root mean squared error for the classifier performance . . . . .	75
3.15 Training time with and without feature selection . . . . .	76
3.16 Noise with 80dB . . . . .	77
3.17 Noise with 60dB . . . . .	77
3.18 Noise with 40dB . . . . .	77
3.19 Root Mean Squared Error (RMSE) after introducing noise to the data. . . . .	78

3.20 Scatter Plot: Actual vs Predicted for Random Forest, AdaBoost, and XGBoost Models. . . . .	79
3.21 Comparison of the RMSE. . . . .	80
3.22 Residual Plot: Actual vs Predicted for Random Forest, AdaBoost, and XGBoost Models. . . . .	80
3.23 Comparison of the training time. . . . .	81
 4.1 Online VSM evaluation scheme . . . . .	84
4.2 Online monitoring . . . . .	88
4.3 change detection using k-means clustering . . . . .	88
4.4 operating point B, fault on line 4032-4044 cleared by opening line: distribution voltages . . . . .	90
4.5 operating point B, fault on line 4032-4044 cleared by opening line: distribution voltages . . . . .	91
4.6 Generator contingencies . . . . .	92
4.7 Online monitoring . . . . .	92
4.8 Online monitoring . . . . .	93
 5.1 Classifications of power system oscillations . . . . .	97
5.2 Plot of the eigenvalues of the Nordic test system. The system is small-signal stable as most of the eigenvalues have negative real parts , but it is poorly damped due to many of its eigenvalues being outside the 10% damping line shown in blue. . . . .	98
5.3 Ambient vs ringdown oscillation for the bus 4041 in Nordic test system [2] . . . . .	99
5.4 Growing oscillations at California Oregon border at Malin substation on August 10, 1996; <i>source</i> [2] . . . . .	102
 6.1 Flow diagram of the proposed methodology. . . . .	110
6.2 Detection of an event in the proposed methodology. . . . .	111
6.3 Variance analysis for different window sizes . . . . .	112
6.4 a) Voltage signal for all buses b) Voltage signal for generators g5,g6 and g7 . . . . .	113
6.5 Reconstructed signal from first three principal components . . . . .	114

6.6	Boundaries detected using the EWT in the Fourier spectrum (the black dashed lines represent the position of boundaries). Segmentation of the Fourier spectrum of the signal for model-C2 . . . . .	115
6.7	Segmentation of Fourier spectrum using SEWT, with the $\alpha = 0\%$ (blue), $\alpha = 10\%$ (green) and $\alpha = 25\%$ (red) of the maximum amplitude. . . . .	118
6.8	Case 1: Decomposition of PC1 voltage signal using SEWT. . . . .	123
6.9	Case 1: Voltage signal for three generator buses, g6,g7 and g8 after the disturbance is introduced at time $t = 50$ s . . . . .	124
6.10	Case 1: On the left voltage signal for generator buses g6, g7 and g8 with the SNR of 40 dB. On the right, PC1 and PC2 for the 20 generator buses from Nordic-32 test system. PC1 captures 83% of the variation, while the PC2 captures 8% of the total explained variance. . . . .	125
6.11	IA envelop fitted with the reconstructed EWT signal. . . . .	126
6.12	Case 1: Top panel shows IF $f(t)$ and the bottom panel shows IDR $\zeta(t)$ of extracted modes from PC1. . . . .	127
6.13	Case 1: Estimates for the average DR $\zeta_{avg}$ for different window sizes. . . . .	127
6.14	Excitation of event in the middle of the window . . . . .	128
6.15	Case 1: Estimates for the average DR. . . . .	129
6.16	Case 1: Estimation of Damping ratio for mode-2 and mode-3. . . . .	130
6.17	Case 1: Estimates for the average DR $\zeta_{avg}$ for different windows with different sample size. . . . .	131
6.18	Case 1: RMSE for different window size . . . . .	132
6.19	Case 2: Temporal evolution of voltage signals. . . . .	132
6.20	Case 2: Estimates of average DR $\zeta_{avg}$ and average frequency $F_{avg}$ from PC1. . . . .	133
6.21	Case 2: Estimates of average frequency . . . . .	134
6.22	Case 1: Estimates of average average frequency with different SNR . . . . .	135
6.23	Case 1: Estimates of average average frequency with different SNR . . . . .	136
6.24	Case 1: Estimates of average average frequency with different SNR . . . . .	137
7.1	Three different operating scenarios. Case-1 represents the unstable operating conditions, Case-2 represents the decreasing oscillations signifying the system is approaching stability . . . . .	141

7.2	Flow diagram of the proposed methodology . . . . .	142
7.3	Stratification of load profiles based on seasons of the year . . . . .	143
7.4	Feature extraction process. . . . .	143
7.5	K-nearest-neighbor . . . . .	146
7.6	C-1: Voltage signal from generator bus g10 with unstable oscillations. . . . .	149
7.7	C-1: Fourier spectrum with detected boundaries. . . . .	150
7.8	C-1: Extracted mono-components of the non-stationary voltage signal. . . . .	151
7.9	C-1: Signal power of extracted mono-components. . . . .	152
7.10	C-1: On the top Reconstructed signal with instantaneous amplitude envelop using SEWT, in the middle instantaneous frequency, the bottom panel shows instantaneous damping ratio. . . . .	153
7.11	Selected features for case study C-1, C-2, and C-3 . . . . .	154
7.12	Case 2: Left)Decomposition of the unstable signal using EWT, Right) Fourier frequency of the decomposed modes. . . . .	156
7.13	Case 1 Stable Scenario:Comparison of average damping ratio estimation by dif- ferent methods. . . . .	157
7.14	Case 2: Left)Decomposition of the unstable signal using EWT, Right) Fourier frequency of the decomposed modes. . . . .	157
7.15	Case 2: Left)Decomposition of the unstable signal using Variation Mode Decom- position (VMD), Right) Fourier frequency of the decomposed modes. . . . .	158
7.16	Case 2: Left)Decomposition of the unstable signal using Local Mean Decom- position (LMD), Right) Fourier frequency of the decomposed modes. . . . .	159
7.17	Case 2: Left)Decomposition of the unstable signal using Empirical Mode Decom- position (EMD), Right) Fourier frequency of the decomposed modes. . . . .	160
7.18	Root mean squared error. . . . .	160
7.19	Computational time . . . . .	163
7.20	Noise impact. . . . .	163
7.21	Root mean squared error. . . . .	164
A.1	Single line diagram of Nordic Test System: <i>Source</i> [3] . . . . .	193

# List of Tables

2.1	Comparison of different ensemble techniques . . . . .	36
3.1	Comparison of feature selection algorithms with respect to existing feature selection methods . . . . .	47
3.2	Comparison of Machine Learning Algorithms for Feature Ensembles . . . . .	51
3.3	Parameters of Nordic buses . . . . .	62
3.4	Parameters of Nordic buses . . . . .	63
3.5	Dataset information . . . . .	65
3.6	Structure of Data . . . . .	66
3.7	Performance of Individual Classifiers . . . . .	68
3.8	Performance of Feature Selection methods. . . . .	72
3.9	Performance of feature ensemble . . . . .	75
3.10	Performance of feature ensemble under the influence of noise . . . . .	78
3.11	Comparison of Feature Ensemble with Random Forest, Adaboost, and XGBoost	79
6.1	Small-signal stability analysis. . . . .	122
6.2	Average frequency without noise . . . . .	134
6.3	Comparison of Damping Ratio and Damped Natural Frequency Metrics with Various SNR Levels . . . . .	136
6.4	Mean squared error . . . . .	138
7.1	Small-signal stability analysis of the Nordic system. . . . .	150
7.2	RMSE FREQUENCY . . . . .	155
7.3	RMSE DAMPING RATIO . . . . .	155
7.4	Feature Score . . . . .	161

7.5	Value of RMSE . . . . .	161
7.6	Value of $R^2$ . . . . .	162
7.7	Training and testing time . . . . .	162
7.8	Noise Impact . . . . .	162
B.1	Best score for hyperparameter obtained through Bayesian optimization . . . . .	195

# List of Abbreviations

1. ANN - Artificial Neural Network
2. BHO - Bayesian Hyperparameter Optimization
3. CART - Classification And Regression Trees
4. CPF - Continuation Power Flow
5. CWT - Continuous Wavelet Transform
6. DBSCAN - Density-Based Spatial Clustering of Applications with Noise
7. DNF - Damped Natural Frequency
8. DR - Damping Ratio
9. DT - Decision Tree
10. DWT - Discrete Wavelet Transform
11. ELM - Extreme Learning Machine
12. EMD - Empirical Mode Decomposition
13. EPS - Electric Power System
14. ERA - Eigen-system Realization Algorithm
15. EMT - Electromagnetic Transients
16. EO - Electromechanical Oscillations
17. EWT - Empirical Wavelet Transform
18. FFT - Fast Fourier Transform

19. F-test - F-test statistic (used as a feature selection method)
20. FVSI - Fast Voltage Stability Index
21. GA - Genetic Algorithm
22. HHT - Hilbert Huang Transform
23. HT - Hilbert Transform
24. HPO - Hyperparameter Optimization
25. IA - Instantaneous Amplitude
26. IF - Instantaneous Frequency
27. IMFs - Intrinsic Mode Functions
28. KNN - K-nearest neighbours
29. Lasso - Least Absolute Shrinkage and Selection Operator
30. LM - Load Margin
31. LTC - Load Tap Changer
32. MC - Monte Carlo
33. MI - Mutual Information
34. ML - Machine Learning
35. MLM - Multivariate Machine Learning Model
36. MSE - Mean Squared Error
37. NF - Natural Frequency
38. PCA - Principal Component Analysis
39. PDC - Phasor Data Concentrator
40. PMU - Phasor Measurement Unit
41. PSS - Power System Stabilizer
42. RELIEF - ReliefF Algorithm (a feature selection technique)

- 43. R2 - Coefficient of Determination
- 44. RBF - Radial Basis Function
- 45. RFE - Recursive Feature Elimination (a feature selection algorithm)
- 46. RMSE - Root Mean Squared Error
- 47. SCADA - Supervisory Control and Data Acquisition
- 48. SEWT - Sliding Window Based Empirical Wavelet Transform
- 49. SFE - Sequential Feature Elimination
- 50. SSO - Sub-Synchronous Oscillation
- 51. SVD - Singular Value Decomposition
- 52. SVR - Support Vector Regressor
- 53. VSA - Voltage Stability Analysis
- 54. VSI - Voltage Stability Index
- 55. VSM - Voltage Stability Margin
- 56. VAR - Variance Threshold
- 57. WAMS - Wide-Area Measurement Systems
- 58. WT - Wavelet Transform

# Chapter 1

## Introduction

The electric power system comprises a complex network of generators, transformers, loads, transmission lines, and distribution lines. The generators convert mechanical energy into electrical energy, which is then transported to consumers (loads) through the transmission and distribution systems. The main objective of the transmission and distribution systems is to ensure uninterrupted supply of high-quality electricity.

The global energy demand has been rising due to rapid urbanization and population growth, leading to significant changes in the power system. In the past, electric utilities were vertically integrated and operated within specific geographic regions. However, the deregulation of the electricity market now allows utility companies to sell power across borders and compete for customers.

To meet the increasing demand for energy, the power system has expanded significantly, with decentralized and renewable sources such as wind, solar, and hydro-power being integrated into the system. The regional grids are also interconnected to enable power exchange over a broader area. However, this complexity also increases the likelihood of contingencies and the severity of their impact. In the event of a disruption such as loss of generation or a short-circuit, the generator's rotor angle separation, bus voltages, and system frequency could be disturbed, leading to cascading failures [4].

At the same time, extensive utilization of power electronic devices and the presence of variable demands, such as those from electric vehicle charging stations and electrified railways, contribute power quality anomalies to electrical grids, including harmonics and rapid fluctuations

in frequencies. In one way or another, these new sources have affected the power system operating conditions by degrading the quality of supply and forcing the system to work under stressful conditions. Consequently, the modern power system encounters numerous challenges that significantly jeopardize its stability parameters. Among these challenges, inter-area oscillations and voltage stability hold particular importance.

Inter-area oscillations belong to the so-called Small Disturbance, or Small Signal Stability problem in power systems [5]. Inter-area oscillations frequently appear among different parts of the power system due to continuing growth in the interconnections. Another contributing factor to these oscillations is the high electricity demand, due to which transmission corridors operate closer to their small-signal, and transient stability limits [6]. If not damped, these oscillations can increase in magnitude and result in system separation, synchronization loss, or even blackouts as a worst-case scenario [7]. Many grids have experienced these oscillations, including South China, the USA, South America, Africa, and Scandinavian countries [8].

On the other hand, voltage stability problems are linked to the escalation of loading in the transmission lines, scarce reactive supplies, and long-distance power distribution. A sequence of events accompanying voltage stability can drive the system to voltage collapse, characterized by an initially slow progressive decline in the voltage magnitude of the power system buses and a final rapid decrease in the voltage. Several incidences of voltage collapse across the world were reported in [9], including the blackouts that happened in Belgium (Aug 1982), Sweden (Dec. 1983), and Tokyo (July 1987). Another incidence of blackout in the United States and Canada (Aug 2003) has proven to be the most significant [10]. Approximately 63 GW of the load was lost during the outage, and about 50 million people were affected. Yet another major collapse took place in Southern Sweden (Sep 2003) that impacted up to 2.4 million customers [11]. These instances prove that any disruption in the power system would lead to cascading outages and dire consequences due to other major infrastructures heavily relying on electricity, such as communication, traffic, water and gas supply, etc.

Therefore, a power system with continuous monitoring and stability prediction is urgently needed. Furthermore, the centralized computational approaches, traditionally used in power systems, face severe challenges in analyzing modern power systems. Due to extensive interconnections, the degree of complexity of the power system has increased to the extent that planning and operation are virtually unmanageable without comprehensive and robust analysis methods.

Research in this area aims to predict voltage collapse to reduce its risk on the power stability networks and identify inter-area modes by continuously capturing and processing power oscillations on wide-area measurement system (WAMS) and its corresponding frequency and damping. Large-scale blackouts such as the 1996 blackout in the U.S. [12], the 2003 blackout in North America [13], the 2003 blackout in Italy [14], and the 2006 blackout in Europe. Vleuten et al.[15] highlight the risks associated with lack of reliability in electric energy infrastructure and the economic impacts of blackouts. Analysis of these blackouts shows that a sequence of cascading events involving line tripping, overloading of other lines, malfunctions of protection systems, power oscillations, voltage stability, and system splitting and collapse caused these outages [16].

## 1.1 Motivation and research objectives

The primary focus of this thesis revolves around two crucial aspects in the power system domain:

1. Long-term voltage stability, and
2. Interarea oscillations.

The motivation for studying these phenomena arises from numerous instances of voltage collapse experienced globally. While both long-term voltage stability and interarea oscillations are associated with power system stability, they each tackle different aspects. Long-term voltage stability focuses on maintaining stable voltage levels at all busses, whereas interarea oscillations pertain to slower oscillations that occur between distinct regions within the power system.

The slow nature of these two instabilities presents significant opportunities for applying machine learning (ML) algorithms. Their gradual processes make them suitable candidates for leveraging ML techniques, which is why we have chosen to investigate these phenomena in our research. By harnessing ML's capabilities, we aim to gain deeper insights into these complex stability issues and develop effective strategies for enhancing the stability of power systems.

In the following subsections, we discuss the implications presented by these phenomena and the driving factors that have sparked our research endeavor.

### 1.1.1 Unveiling the Significance of Oscillatory stability in Electric Power Systems

Interarea oscillation involves slower and sustained oscillations that occur between different regions or areas of a power system. These oscillations typically have frequencies in the range of a few cycles per second and are observed in the large system with groups of generators, or generating plants connected by relatively weak tie lines. These oscillations involve groups of generators, or generating plants, on one side of the tie oscillates against groups of generators on the other side of the tie[17]. Interarea oscillation is an area of interest and concern for researchers due to the following reasons:

- **Widespread Impact:** Interarea oscillations are low-frequency oscillations that can propagate across large geographical regions, affecting multiple interconnected areas of the power system. These oscillations can lead to widespread disturbances and impact the stability of the entire grid[18].
- **Cascading Effects:** Interarea oscillations can trigger cascading failures, where the oscillations interact with various system components, leading to a chain reaction of failures. Cascading events can result in extensive blackouts and severe disruptions in the power supply[19].
- **Resonance Phenomena:** The process of interarea oscillations can be triggered by resonance phenomena, where the natural frequencies of different parts of the power system align, leading to amplified oscillations. Understanding these resonance mechanisms is crucial to mitigate their adverse effects[20].
- **Complex System Dynamics:** Power systems are complex and interconnected networks with multiple generation sources, transmission lines, and loads. Interarea oscillations are challenging to analyze and control due to the complexity of the system dynamics [21].
- **Renewable Integration:** With the increasing integration of renewable energy sources, power systems become more susceptible to interarea oscillations. The variability and unpredictability of renewable generation can introduce additional uncertainties in the system, impacting stability[22].
- **Transmission Line Dynamics:** Interarea oscillations can interact with the dynamics of long-distance transmission lines, affecting the system's behavior over large distances [23].

- **Interconnected Grids:** Interarea oscillations can be exacerbated in interconnected power grids, where power transfers between different regions can influence oscillation dynamics [24].

Detecting and mitigating these oscillations in real-time require advanced monitoring and control strategies. Understanding the dynamics of these oscillations and developing effective control strategies are essential to safeguarding the smooth operation of interconnected power grids.

### 1.1.2 Unveiling the Significance of Long-Term Voltage Stability in Electric Power Systems

Long-term voltage stability, as a gradual process, emerges when the power network encounters challenges in delivering sufficient reactive power support, particularly in specific network nodes or areas of the power system [25]. It involves assessing the system's ability to withstand varying load conditions, generation patterns, and other factors that may impact voltage levels. Long-term voltage stability is an area of concern for researchers in electric power systems due to several reasons:

- **Power System Complexity:** Modern power systems are becoming increasingly complex, incorporating various sources of generation, flexible loads, and interconnected transmission networks. As the complexity of the power grid increases, it becomes challenging to ensure stable voltage levels over extended periods [26].
- **Renewable Energy Integration:** The integration of renewable energy sources, such as solar and wind, introduce intermittent and variable generation patterns into the power system. These fluctuations can impact voltage stability, especially during periods of high renewable energy penetration[27].
- **Reduced Stability Margins:** Power systems are often operated with reduced stability margins to optimize asset utilization and accommodate diverse demand patterns. Operating with smaller stability margins makes the system more susceptible to voltage instability and collapses [28].
- **Risk of Blackouts:** Voltage instability can lead to cascading failures, where a disturbance in one part of the system triggers a chain reaction of failures, eventually resulting in blackouts. Such blackouts can have severe consequences, including economic losses and

disruption of essential services [29].

- **Demand Growth:** Growing electricity demand and changing load patterns can put additional stress on the power system, affecting voltage stability. Researchers need to understand how the changing demand can impact the long-term voltage stability of the grid [30].

Long-term voltage stability analysis helps power system planners and operators anticipate potential voltage stability issues. By analysing voltage stability over the long term, power system operators can identify potential vulnerabilities and risks associated with different operating scenarios.

### 1.1.3 Shortcomings of existing methods

:

Techniques used for interarea oscillation analysis often include modal analysis, eigenvalue analysis, and frequency response analysis. These methods help identify the critical interarea oscillation modes and assess the stability of the power system concerning these oscillations. Long-term voltage stability analysis requires a different set of techniques and tools, such as load flow studies, voltage stability indices, and voltage stability margins. These methods help identify potential voltage stability issues in the power system and assess its ability to withstand large disturbances, such as heavy load demands or equipment failures.

Analyzing long term voltage stability and interarea oscillation in power systems using traditional methods has some shortcomings, and ML offers a better choice to address these limitations. Here are some of the shortcomings of traditional analysis:

- **Complexity and Nonlinearity:** Power systems are complex and nonlinear, and traditional analysis methods may struggle to capture the intricate relationships between variables accurately.
- **Sensitivity to Assumptions:** Traditional methods often rely on simplifying assumptions, which may not accurately represent real-world conditions, leading to potential inaccuracies in stability assessments.
- **Limited Data Utilization:** Traditional approaches might not effectively utilize the vast amount of high-resolution data available from PMUs and other sensors, missing valuable

insights.

- **Time-Consuming:** Simulation-based methods for stability analysis can be computationally intensive and time-consuming, making real-time monitoring and decision-making challenging.
- **Difficulty in Handling Missing Data:** Traditional methods may struggle to handle missing or incomplete data, leading to information loss and reduced accuracy.

#### 1.1.4 Unleashing the Potential of ML Techniques

:

Recently, ML techniques offer a paradigm shift in long-term voltage stability analysis by leveraging the power of data-driven insights and advanced computational capabilities. The inherent challenges faced by traditional methods in accurately addressing voltage stability and oscillation issues have paved the way for the motivation behind the utilization of ML techniques. These algorithms are gaining popularity due to several reasons:

- **Data-Driven Insights:** ML leverages data from real-world scenarios to reveal complex relationships and patterns that traditional methods might overlook.
- **Nonlinear Dynamics:** Long-term voltage stability involves nonlinear interactions. ML techniques can capture these non-linearities more effectively.
- **Model Complexity:** Power systems are intricate and constantly evolving. ML can handle large datasets and adapt to changing system dynamics.
- **Enhanced Prediction:** ML algorithms can predict voltage stability margins based on historical data, offering real-time assessment and early warning capabilities.
- **Automation:** ML-based tools can automate the analysis process, reducing the need for manual parameter tuning and simplifying complex calculations.
- **Adaptability:** ML algorithms can adapt to changing system conditions and incorporate new data for continuous improvement.
- **Early Detection:** ML-based monitoring can offer early detection of potential voltage instability, allowing grid operators to take proactive control actions.

Overall, this research aims to provide valuable insights into long-term voltage stability and interarea oscillations in electric power systems. By leveraging PMU data and employing data-driven approaches with ML techniques, we seek to enhance stability assessments and contribute to the overall reliability and efficiency of power systems.

## 1.2 Monitoring of power system stability

Modern power systems are comprised of two sets of measurement systems,

- SCADA and
- WAMS [31].

Both of these systems have their pros and cons. Due to longer tenure, the SCADA system has amassed a variety of operational experiences, while WAMS offers accurate and rapidly updated measurements [32].

### 1.2.1 Supervisory Control and Data Acquisition

The traditional monitoring approach uses a SCADA tool to gather the system's information. SCADA is a computer system for assembling, analyzing, and monitoring real-time data. SCADA systems monitor and control plants or equipment and gather information from meters, transducers, and similar devices. The collected data contain information on the real and reactive power flows, voltage magnitudes, and breakers and switches' status [33]. Simultaneously, the system status measured from the substations is broadcast to the control centre. In the control centre, the SCADA system facilitates the operators to monitor or control the entire power system by transmitting the data to the principal computer facility, and presenting the information to the operator through the human-machine interface [34].

However, with the rapid economic growth and increasing electricity demand, power systems often operate closer to their stability limit and are more endangered by fast-evolving dynamic events. In this scenario, the SCADA applications are inadequate to assist power systems in performing firmly and securely; the blackout in North America and Canada in 2003 is such an example. Since the SCADA system provides low sampling density and non-synchronous information about the network, the control center cannot know the dynamic operation states of the system precisely [35].

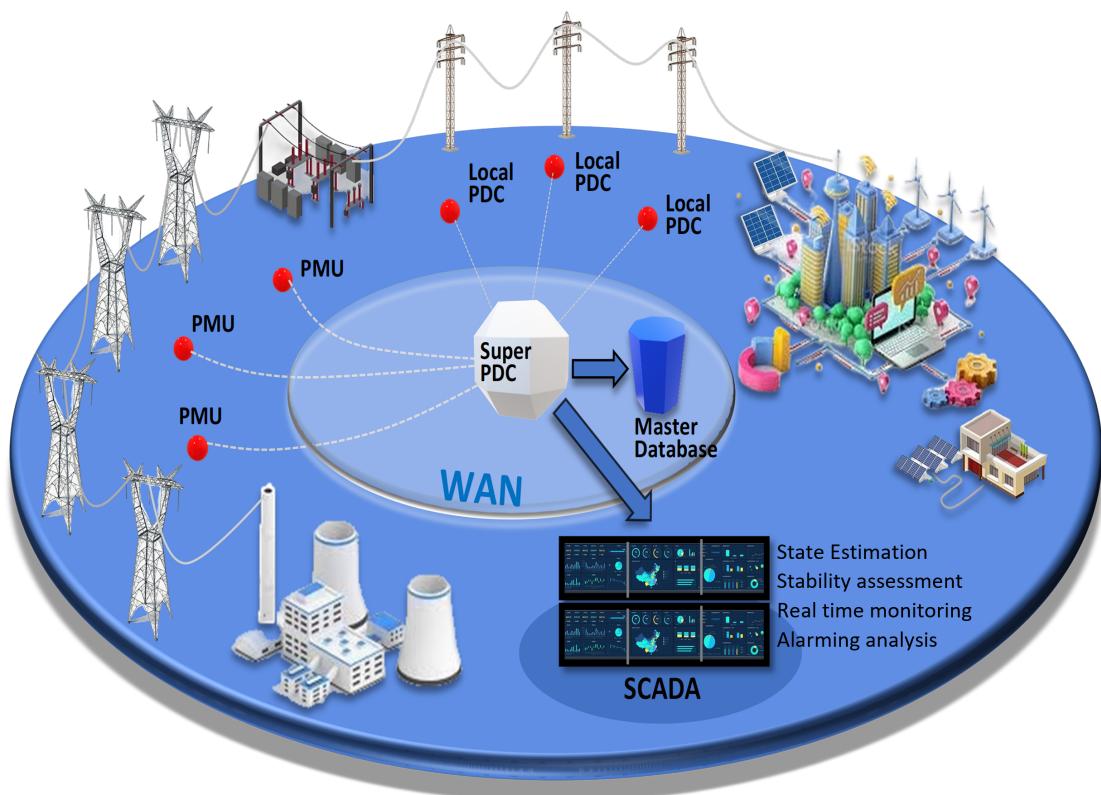


Figure 1.1: Architecture of WAMS; where PMUs collect the data from various sources of power systems. The collected data sets are delivered to a local Phasor Data Concentrator (PDC), which transmits them to a master database called super-PDC. The consolidated data sets collected by super-PDC are fed into analytic applications such as state estimation, stability assessments, data visualization, real-time monitoring, and control.

### 1.2.2 Wide Area Measurement Systems

Due to its asynchronous and slow nature, SCADA does not provide power system information at sub-second time frames to the state estimator and, therefore, does not provide dynamic state estimations [36]. WAMS comprising several phasor measurement units (PMUs) provide enhanced capabilities for accurate and real-time monitoring of the system's stability [37]. WAMS complements the data acquisition functions of SCADA and has higher precision, shorter update cycle, lower transmission delay, and sensitivity with the system status variation.

Over the past two decades, WAMS has been widely deployed to monitor power system stability and get real-time measurements of voltage magnitude, frequency, phase-angle, active and reactive-power variations, and waveforms with high-quality visualizations. Due to its high sampling rates, WAMS are capable of capturing data with a level of granularity that significantly reduces online estimation errors. This enhanced frequency of data acquisition allows for a more accurate representation of the dynamic behavior of the power system, ensuring that real-time measurements closely align with the actual system conditions [38].

Fig. 1.1 shows the architecture of a typical WAMS. Widely distributed PMUs dispatch the measurements to PDC. Generally, a specified number of PMUs are dealt with by a single PDC. The PDC retains the local applications' data and transmits the rest to a super PDC for advanced application. In super PDC, the measured data from the whole power system is synchronized using timestamps. The three layers of WAMS operations can be classified into data acquisition, management, and applications. The WAMS technology helps prevent blackouts, improve state estimation, and better utilize transmission networks [39].

## 1.3 Classification of Power System Stability

For a given initial operating condition, power system stability refers to the ability of an electric power system (EPS) to regain a state of operating equilibrium after being subjected to a physical disturbance, with most system variables bounded so that practically the entire system remains intact [40]. Fig. 1.2 gives an overall picture of the power system stability classification, identifying its categories and subcategories. This thesis focuses on two of the power system stability categories [41]:

1. Angle Stability or Rotor Angle Stability

## 2. Voltage Stability

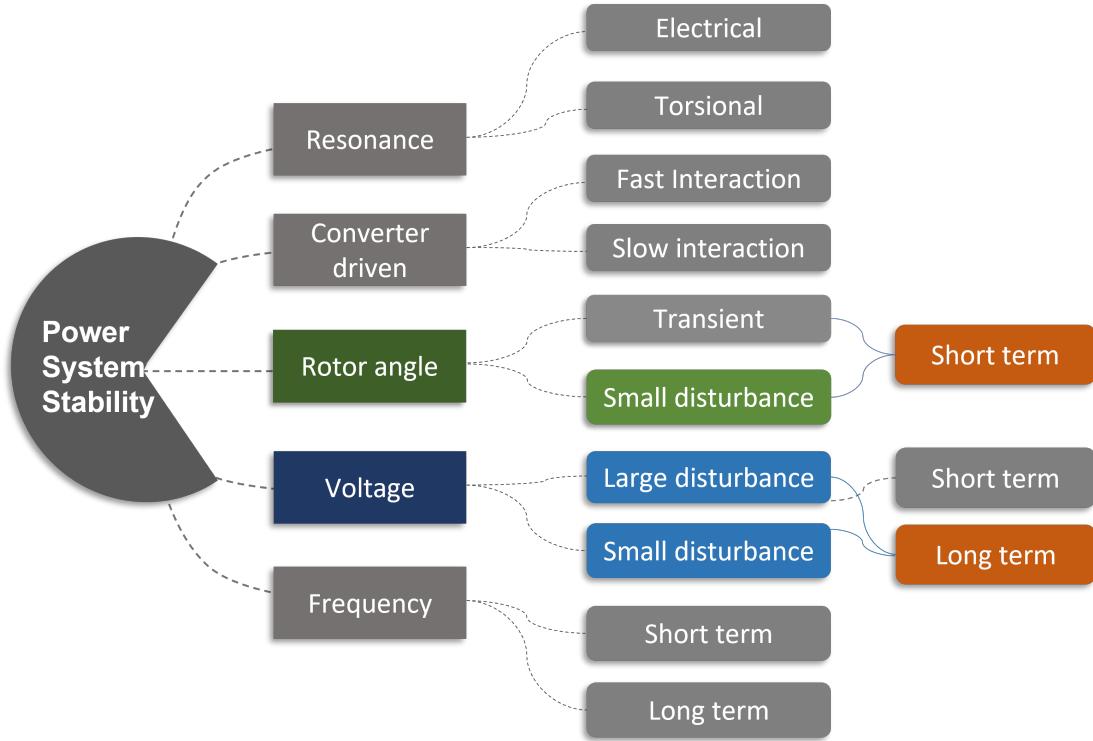


Figure 1.2: Classification of power system stability. In orange the sub-categories of interest in this thesis [1]

The selection of Angle Stability (Rotor Angle Stability) and Voltage Stability as the primary focus for the thesis is based on their characteristics as slow-moving stability issues. These types of stability problems generally evolve over a more extended period compared to other types of instability in power systems, which can provide better opportunities for the training of ML algorithms.

We present here several justifications for the pertinence of incorporating slow-moving stability concerns into ML training, and we elucidate the rationale behind selecting them as the central focal point:

- **Data Availability and Quality:** Slow-moving stability issues allow for more frequent and continuous measurements of system variables, such as angles and voltage levels, over time. This availability of high-quality data can be crucial for effectively training ML algorithms, as large datasets can enhance the model's accuracy and generalization.
- **Feature Engineering:** In slow-moving stability problems, the system dynamics change gradually, allowing for more effective feature engineering. ML algorithms can benefit

from well-engineered features that capture the system's behavior and changes, leading to improved model performance.

- **Resilience Analysis:** The study of slow-moving stability issues can help identify critical vulnerabilities and potential instability threats in power systems. ML algorithms can be applied to assess the resilience of the system to varying conditions and disturbances.
- **Decision Support Systems:** Slow-moving stability issues demand effective decision support systems that can provide timely warnings and actionable insights to power system operators. ML algorithms can be used to develop intelligent decision support tools that enhance situational awareness and aid in stability management.

The research can play a vital role in advancing the state-of-the-art in power system stability analysis and control, benefiting both the power industry and society as a whole. A brief summary of the two stability aspects mentioned above is presented as follows:

### 1.3.1 Rotor angle stability

Angle stability (or Rotor Angle stability) is defined as '*the ability of interconnected synchronous machines of a power system to remain in synchronism*'. The category of angle stability can be considered in terms of two main subcategories [1]:

**Transient stability** results from the inability to maintain synchronism after large disturbances such as system faults and /or equipment outages. Transient stability studies aim to determine if a system's machines will return to a steady synchronized state following a large disturbance.

**Small signal stability** results from the inability to maintain synchronism and/or dampen out system transients and oscillations caused by small system changes, such as continual changes in load and /or generation. This stability problem involves the study of electromechanical oscillations (EO) inherent in power systems. EO was observed in the power system as soon as synchronous generators were interconnected to provide more power capacity and reliability. Formerly interconnected generators were close to each other, and oscillations were in a frequency range of 1 to 2 Hz [42],

As more utilities connected, oscillations in different frequency ranges were observed. Nowadays, EO can be classified by specific types, with each type exhibiting a particular range of oscillation frequency when the phenomenon occurs.

In general, EO can be grouped into four broad classes:

- Local plant mode oscillations occur when synchronous generators swing against the extensive power system in a specific power station. This type of oscillation occurs within a range of 0.7 to 3 Hz.
- Inter-area mode occurs when a group of generators in one area swings against a group of generators in another area of the power system. The frequency range of inter-area oscillation mode is in the range of 0.1 to 0.7 Hz.
- Interplant mode occurs when two or more synchronous generators in the same power plant or nearby power plant swing against each other between 1.5 to 3 Hz.
- Torsional mode oscillations occur due to the interaction of the mechanical turbine generator and a system connected through the series compensated line. The characteristic frequency of torsional mode oscillations is 10Hz to 46 Hz.

From an operational perspective, oscillations are permissible as long as they decay. However, unstable EOs can be even triggered by regular small changes in the system load. In this case, there might be no warnings to the operator.

### 1.3.2 Voltage stability

Voltage stability can be defined as '*the ability of a system to maintain steady acceptable voltages at all buses following a system contingency or disturbance*'. Voltage stability arises from the inability of the transmission and generation system to supply the power demanded by loads. A system can reach a state of voltage stability when the increase in load demand or shift in system condition yields a sudden and unmanageable drop in voltage. An EPS is small disturbance voltage stable at a given operating state if voltages near the loads do not alter or maintain pre-disturbance values after being subject to any small disturbance. The notion of small disturbance voltage stability is connected to steady-state stability and can be examined using the system's small-signal (linearized) model [41].

The literature of this thesis will focus on the steady-state/dynamic stability subcategory related to inter-area oscillations and small disturbance voltage stability related to voltage stability. Both of these subcategories are indicated by orange boxes in Fig. 1.2.

## 1.4 Thesis Contributions

This thesis proposes two methods for detecting power system stability problems, which can help provide preventive actions for system operators. The first method is designed to monitor long-term voltage stability, while the second method is intended to observe low-frequency inter-area oscillations (LFIO). These methods are outlined in detail in the following papers

- A feature-subspace-based ensemble method for estimating long-term voltage stability margins.
- Detection of Oscillatory Modes in Power Systems using Empirical Wavelet Transform

The main contributions of this thesis are summarised below:

### **Contributions related to voltage stability**

- Previous research has introduced various machine learning (ML)-based techniques like artificial neural networks (ANN), support vector machines (SVM), and Classification and Regression Trees (CART) to predict voltage stability/loadability margin. This thesis proposes two novel concepts to enhance precision and robustness: (i) an automated feature selection method for improved long-term voltage stability (LTVS) predictions and (ii) a feature ensemble ML models for robust predictions under changing topological conditions.
- Feeding numerous features into an ML model exponentially increases the search space, leading to challenges in generalization due to the curse of dimensionality. Feature selection becomes crucial, especially for large, high-dimensional datasets like power systems, as it optimizes the learning complexity of ML models. This reduces redundancy, computational complexities, memory usage, and computation time.
- A significant contribution of this thesis is the introduction of a novel feature ensemble approach for online voltage stability monitoring, surpassing conventional methods. Despite various available feature selection techniques, effective feature selection remains challenging. Initial experiments reveal discrepancies in feature sets produced by current methods under the same conditions. Unlike existing techniques, this approach aggregates outputs from multiple feature selectors to handle high-dimensional data and improve generalization. It outperforms regression-based [43] and attention mechanism [44] approaches in terms of enhancing feature extraction, model performance, and interpretability. Indi-

vidual techniques such as RELieff [45] and mutual information [46] exhibit inconsistent performance, making reliance on a single method risky for building models with selected features. Automation is difficult due to the need for domain expert analysis. The thesis addresses these challenges through the ensemble feature selection technique.

- In the field of electric power systems (EPS), many ML studies use Monte Carlo (MC) simulation for data generation and operating space selection. However, this traditional MC approach is computationally expensive, particularly for fast assessment applications and identifying low-probability high-impact outage events. The study introduces a Cluster-based sampling approach to alleviate computational burden and capture diverse load features. Clustering identifies similarities in load profiles collected from distributed Phasor Measurement Units (PMUs), enabling understanding of consumption patterns across different consumer types and time scales.

### **Contributions related to oscillatory stability**

- The central objective of this thesis is to design an adaptive data-driven signal processing framework for determining modal parameters, specifically focusing on leveraging the Empirical Wavelet Transform (EWT). However, one notable challenge with EWT arises from its requirement to pre-specify the number of modes—a task complicated by closely spaced modes in power systems and signal noise.
- To overcome this challenge, a pioneering sliding window-based empirical wavelet transform (SEWT) methodology is introduced in the thesis. This SEWT approach excels in automatically detecting modes without the need for a predefined mode count. It employs a sliding window-driven segmentation strategy, defining empirical boundaries that enable signal decomposition into mono-components. These components hold potential anomaly insights. By leveraging the Hilbert transform, modulation information is extracted from the mono-components.
- A primary goal of this thesis is to attain precise damping estimations for Inter-area modes, renowned for their susceptibility to noise interference. The innovative aspect of this approach lies in the integration of SEWT with machine learning (ML) models through a multivariate paradigm. Through the fusion of adaptive signal processing and ML techniques, the proposed method effectively avoids limitations and achieves heightened accuracy in

estimating Damping Ratios (DRs) and modal frequencies.

- A distinct novelty emerges from the integration of multivariate machine learning into the methodology. This integration significantly boosts the approach's ability to estimate signals originating from multiple buses. The incorporation of this multivariate approach efficiently captures intricate inter-dependencies among variables, a capability that sets it apart from conventional wavelet and Empirical Mode Decomposition (EMD) techniques—methods that often grapple with unravelling complex multivariate relationships.

#### 1.4.1 List of published papers

- Ambreen Khurram, and Arief Gusnanto, and Petros Aristidou, “ Detection of oscillatory modes in power systems using empirical wavelet transform”, published in IEEE Madrid PowerTech, 2021
- Ambreen Khurram, and Arief Gusnanto, and Petros Aristidou, “ A feature-subspace-based ensemble method for estimating long-term voltage stability margins”, published in EPSSs Research, 2022
- Ambreen Khurram, and Arief Gusnanto, and Petros Aristidou, “ Estimation of interarea modes in power system using ensemble learning.”, to be submitted.

## 1.5 Outline of the Thesis

The thesis is organized as follows:

In **Chapter 2**, the focus is on power system stability issues and a comprehensive analysis of the related literature. The long term voltage stability problem formulation is explained and various techniques for generating relevant data are discussed. It is emphasized that accurate modeling of power system loads is crucial for understanding the voltage stability phenomenon. Furthermore, dynamic load models are found to be more effective than static load models for studying voltage stability. To achieve this, clustering-based algorithms are presented as a means of generating realistic load operating data.

In **Chapter 3**, the Ensemble Feature Selection approach is introduced as a means of improving the determination of long-term voltage stability. The chapter details the proposed scheme and explains its significance in this context. Additionally, the chapter provides a background on

various feature selection methods and classification algorithms and explains how these techniques can be applied to predict VSM. The chapter includes case studies and describes dynamic simulations used to generate load operating points. The feature ensemble technique is then applied to select the most pertinent features, and a ML model is trained offline. Finally, the chapter evaluates the performance of the classifier on the generated offline data.

**Chapter 4** is dedicated to the application of a trained MLM for estimating voltage stability margin (VSM) in real time, with the primary goal of providing qualitative information about stability margins to the transmission operator. However, during the online process, statistical classification tasks face additional challenges due to potential differences between the training and target sample distributions over time, caused by changes in the system's topology or various fault conditions. To address this challenge, a k-means clustering-based approach is utilized to detect changes in operating conditions by measuring the distance between the cluster centroid. The chapter highlights the importance of the database update stage in improving the feature ensemble scheme's generalization ability and robustness under complex operating conditions. Moreover, the study explores whether the distance measured from the k-means clustering indicates the need to update the database.

In **Chapter 5**, interarea oscillation is introduced, and a detailed review of low-frequency oscillation phenomena in power systems is presented. The chapter provides a clear definition of low-frequency oscillation and its categorization. It also includes examples of power system blackout incidents caused by low-frequency oscillation. The chapter presents widely used methodologies for studying low-frequency oscillation among researchers.

In **Chapter 6**, a technique is presented for estimating interarea modes in real-time using PMU measurement data. The algorithm proposed in this chapter is designed to identify the DNF and DR of the interarea oscillatory modes that are present in the power system signal. To reduce the multidimensional PMU data, principal component analysis is used initially. Then, SEWT approach is introduced that employs moving window segmentation for detecting boundaries in the Fourier spectrum. By decomposing the low-frequency electromechanical oscillation signal into a series of mono-components, the SEWT approach effectively applies the Hilbert transform to the IMF to acquire the instantaneous parameters of the signal. The DNF and DR of each mode are then determined by computing the average of the instantaneous parameters. The proposed method achieves accurate segmentation even in noisy and non-stationary signals.

Additionally, simulated and experimental signals are utilized to confirm the efficacy of the proposed method.

In **Chapter 7**, a ML based approach is presented for estimating an inter-area dominant mode. Initially, the oscillatory signal's features are extracted through SEWT, and the dominant inter-area modes are evaluated. Subsequently, the Hilbert transform (HT) is employed to estimate the instantaneous amplitude and instantaneous frequency from the decomposed monocomponents signal. The ML algorithm is then utilized to process the extracted features, and the trained model is implemented to estimate the interarea modes from multiple busses in real-time. The proposed method's effectiveness is demonstrated by applying it to the Nordic test system, which shows improved accuracy of inter-area mode estimation as compared to the traditional approach.

**Chapter 8** concludes the work and shows the future scope of work.

## Chapter 2

# Voltage stability analysis

Long-term voltage stability is usually related to a gradual or uncontrollable drop in voltage magnitude after the system is subjected to a disturbance, a rise in load demand, or an inability to meet the reactive power demand [29]. Voltage stability may result in a black-out in parts of the system. When the power system operates with an inadequate Voltage Stability Margin (VSM), it becomes susceptible to voltage collapse. As power systems are operated under increasingly stressed conditions, the ability to maintain voltage stability becomes a growing concern. Voltage stability analysis mitigates the risk of voltage collapse and should be considered during the power system's planning and real-time operating stages.

Voltage stability assessment methods can be divided into offline and online [29]. Offline studies are conducted during the system planning stage, and online assessment is performed during the system operations (in real-time). In contrast to offline planning, where the computational speed may not be critical, online analysis tools are of great importance for assessing the voltage stability of the power system in real-time and determining proximity and potential stability mechanism [47]. Proximity gives a measure of voltage security and uncovers how close the system is to voltage collapse, whereas a mechanism provides information helpful in determining system modifications or operating strategies that could be used to prevent voltage stability.

Over the past three decades, various methods have been proposed for voltage stability analysis. In this Chapter, an overview of the most popular and successful analysis techniques is presented.

## 2.1 Literature Review

Voltage stability analysis methods proposed in the literature can be broadly classified into three categories:

- Methods based on Modal Analysis
- Methods based on Loading Margin
- Methods based on Machine Learning

These methods are briefly described in this section.

### 2.1.1 Methods Based on Modal Analysis

The modal analysis method [47] is a mathematical technique, mainly used to predict voltage collapse in complex power systems. The approach computes the smallest Eigenvalue of the reduced power flow Jacobian matrix to estimate how close the system is to the voltage collapse. However, the modal analysis results are valid for incremental (small-signal) changes only. Consequently, the strength of this methodology lies in providing information on system trends rather than estimating the actual numerical values of system variables following changes [48].

### 2.1.2 Methods Based on Loading Margin

One of the practical approaches to assess voltage stability is by calculating the loadability margins. Methods based on the calculation of the loadability margins can be further divided into the following methods:

- Continuation power flow (CPF) method
- P–V and Q–V curves and
- Voltage stability index (VSI)

#### Continuation Power Flow technique

A major limitation of conventional power flow is its inability to converge to a maximum loading point and the problem of reaching the singularity point of the Jacobian matrix. To avoid this issue, the power flow equations can be reformulated by applying a locally parameterized continuation technique. Continuation methods have four basic elements: i) parameterization, ii)

predictor, iii) step length control, and iv) corrector [49]. The continuation technique estimates a new solution in each predictor step by increasing the load in the conventional power flow equation. The exact solution is computed in the subsequent corrector step. The process resumes till the tangent vector becomes zero at which point the maximum loading point is achieved [50].

One of the early contributions to computing the turning points appeared in [51]. However, much of the analytical development of robust path-following techniques is due to Keller [52]. An outstanding explanation of a locally parameterized continuation algorithm was provided by Rheinboldt in [53]. Some good additional resources for turning points computation were mentioned in [54] and [55]. Lately, [56] applied geometric parameterisation-based CPF to eliminate the Jacobian matrix singularity by the addition of the line equations which pass through the points in the plane determined by the variables loading factor and the sum of nodal voltage magnitudes, or angles, of all system buses. Ju et al. [57] suggested a local geometric parameterisation technique that can trace the P–V curve with a fixed step length. The technique focuses on tracing the segment near the saddle-node bifurcation (SNB) point and the lower part of the P–V curve.

However, [58] has classified two types of failure in applying CPF for real-time voltage stability monitoring. The first failure is related to the conventional global parameterisation that cannot overcome the Jacobian matrix singularity for stability cases with strong local characteristics. The second failure arises when the extended Jacobian matrix in the corrector procedure becomes singular because of improper parameterisation.

### P-V and Q-V Curves

The P-V curve portrays the relationship between voltage (V) and active power (P), while the Q-V curve shows the associations between the V in a bus and its reactive power (Q) [59]. Both curves show the variation (increase or decrease) of the voltage as the load demand varies. The nose point of the P-V curve is the voltage collapse/critical point [50]. The upper part of the P-V curves indicates a stable region, while the lower part shows unstable region. Whereas in the Q-V curve, if the operating point is on the right side of the curve, the system is considered stable and if the operating point is on the left side of the curve the process is deemed to be unstable. The voltage collapse occurs when the system load (P and/or Q) increases beyond a specific limit. The VSM for a given operating point can be easily found if the limiting values

of P and Q are known [60].

[61] investigated the impact of grid-connected photovoltaic (PV) system on static voltage stability using P-V curve and improved VSI. [62] proposed a P-V curve-based method using one of the load bus voltages as a parameter by regularizing the power flow solution around the maximum loading point. The method monitors Q limits on P-V buses. P-V curves were traced with and without voltage-dependent loads. [60] combined P-V and Q-V curves into P-Q curves to determine the voltage stability limit. The boundary of the voltage stability region is first determined and then presented in the P-Q plane. The voltage stability margin is determined from the stability boundary in the P-Q plane. Motivated by [60] work, [63] presented a method of determining the voltage stability boundaries for a power system with voltage-sensitive loads using P-Q curves by incorporating the load characteristics such as constant current, constant impedance and mixed load. The limiting or critical values of P and Q at the voltage collapse point are first determined for each type of load and then used to plot the voltage stability boundary in the P-Q plane.

The P-V, Q-V curve involve performing power flow calculations at different voltage levels to obtain the relationship between active/reactive power and voltage. However, these methods have limitations for online applications due to the following reasons:

1. Computational Complexity: The P-V and Q-V curve methods require solving power flow equations at multiple voltage levels, involving a large number of nonlinear equations and iterative calculations. This computational complexity increases exponentially with the size of the power system, making it impractical for real-time or online applications where quick and efficient analysis is required.
2. Limited Scalability: Power systems are becoming increasingly larger and more complex, with the integration of renewable energy sources, distributed generation, and advanced control technologies. The P-V and Q-V curve methods struggle to scale effectively with such complex systems due to their computational limitations. As the size of the system grows, the computational burden becomes even more significant, hindering their practicality for online applications.
3. Lack of Dynamic Information: The P-V and Q-V curve methods provide static information about the power-voltage relationship at different operating points. However, they do not

capture the dynamic behavior and transient effects that can occur in power systems, such as voltage fluctuations, disturbances, or rapid load changes. Online applications often require real-time monitoring and analysis of dynamic system conditions, which cannot be adequately addressed by the static P-V and Q-V curve methods.

### **Voltage Stability Index**

Voltage collapse is marked by gradual variations in the system's operating point due to an increase in the loads such that the voltage magnitude slowly decreases until a sharp decline occurs. It has been found that voltage magnitudes do not give a good indication of proximity to the voltage stability limit. The problem of voltage collapse may be explained as the inability of the power system to supply reactive power. An effective VSI indicates how far the current operating condition is from voltage collapse, what weak lines exist, which buses are the most vulnerable, and which ones will go a long way in helping power system operators [64]. Many VSIs have been proposed in the literature. These indices can be broadly classified as:

- Jacobian matrix based VSIs
- Bus, line and overall VSIs

In this review, we will refer only to the techniques that have been applied widely in the literature for the case study of online electric power stability (EPS) assessment.

#### **Jacobian matrix based voltage stability indices**

VSIs based on the Jacobian matrix utilize the notion of the singularity of the power flow Jacobian matrix [65, 66, 67, 68]. The Jacobian matrix near the point of voltage collapse is close to the singularity, hence the minimum singular value or eigenvalue index could measure the distance of the current state from the voltage collapse point [69]. However, the computation time of the VSI based on the Jacobian matrix is high and is not suitable for real-time monitoring as any topological change leads to a change in the Jacobian matrix and this matrix must be recalculated.

#### **Line voltage stability indices**

The line stability indicators set the discriminant of the voltage quadratic equation to be greater than or equal to zero to achieve stability [70]. The theoretical foundation of most line VSIs is the same except for the assumptions made by each index.

[71] put forward a line stability index,  $L_{mn}$  founded on the principle that there exist solutions to the quadratic voltage equation; the index equal to the maximum value of one indicates the voltage stability is collapsing, and a value of zero indicates there is no load in the system.  $L_{mn}$  assumes zero shunt resistance and that real power does not affect voltage stability. Its primary advantage is its insensitivity to transmission lines' resistance/reactance ratio, as shown by the authors in [64]. However, the assumption that real power does not affect voltage stability might cause the index to be inaccurate under certain operating conditions [72].

[73] formulated the fast voltage stability index (FVSI), which measures the stability related straight to reactive power and indirectly to active power across the lines. For a system to remain stable, FVSI must be below 1. Otherwise, the system will experience a sudden voltage drop, and voltage collapse will occur. FVSI assumes that the voltage angle difference between sending end and receiving end buses is approximately zero. This assumption is a significant drawback as large voltage angle differences are considered a precursor to the voltage collapse, as highlighted by authors in [74]. Another disadvantage is the index's sensitivity to the resistance-reactance ratio of the transmission line [75].

[76] introduced the line stability factor (LQP) on the same principle as  $L_{mn}$  and FVSI. To maintain stability, LQP must be less than 1. This index ignores shunt admittances and assumes that lines in the power system are lossless, which could cause it to be inaccurate under certain operating conditions. The advantage of this index is it's insensitive to transmission lines' resistance/reactance ratio of transmission lines [75].

[71] proposed Line Stability Index based on the same concept as the last VSIs. Voltage collapse can be assessed based on the value of the developed index. If the system yields a value exceeding its maximum limit of stability index one, it indicates a voltage collapse situation. In this index, the effect of reactive power on voltage stability and line shunt admittance is neglected, and it is assumed that only the active power affects the line voltage stability.

### **Bus voltage stability indices**

The VSIs based on buses assess the voltage stability of system to determine critical buses or estimate the voltage stability margins of the EPS [77]. Bus VSIs are widely used in power system analysis to assess the voltage stability of individual buses in a power system. These indices are typically based on voltage magnitudes and angles at the buses and are useful for

identifying buses that are experiencing or approaching voltage instability. However, they do not directly provide information about the components such as transmission lines, transformers, or generators with limited capability, that can contribute to voltage problems [70]. Some VSIs are briefly described in this section.

[78] offered Index L, which varies in the range of 0 (no-load of the system) and 1 (voltage collapse) to determine the voltage stability or the proximity of a collapse. The L-index estimates the closeness of the actual state of the system to the stability limit [77]. The S-difference-criterion (SDC) presented in [79] is established because when the system approaches collapse, the total increase in apparent power loading is due to the transmission losses. [80] suggested the Local Identification of Voltage Emergency Situations (LIVES) index based on the Load Tap Changer (LTC) controllers of bulk power delivery transformers. Furthermore, an extension of the aforementioned is presented as NLI (New LIVES Index) by Vournas, Lambrou, and Mandoulidis [81], based on observing the controlled voltage of bulk power delivery LTC transformers on the distribution side. [82] formulated the Voltage Collapse Proximity Index (VCPI) by using the voltage phasor and angle information and the network admittance matrix to assess the system voltage stability at a given bus. The index is derived from the fundamental power flow equation, and its value varies between 0 and 1, with 1 being the collapse point.

One major drawback of many VSIs is not providing enough intuitive information as provided by the offline analysis tools such as P-V curves or CPF for making a proper decision. Moreover, these VSIs show different levels of accuracy under different conditions and load models [64]. The practical aspects such as measurement errors and incorrect network topology information can also affect the accuracy of VSIs estimation in different ways and it is difficult to single out one VSI which is more reliable [28].

### 2.1.3 Methods based on Machine Learning

As mentioned above, numerous tools have been developed to conduct a comprehensive analysis of the voltage stability assessment. However, most of these methods cannot be used in real-time as they are computationally time-consuming and rely on the complex mathematical modelling of the EPS. The enormous computational requirements could be resolved by utilizing machine learning techniques. Machine learning provides computers with learning capabilities without being programmed [83]. Specifically, machine learning is the subset of artificial intelligence (AI)

which extracts knowledge by analyzing and manipulating data compiled from real-world use cases [84]. Machine learning comprises numerous approaches such as artificial neural network (ANN), decision trees (DTs), fuzzy logic (FL), adaptive neuro-fuzzy inference systems (ANFIS) and support vector machines (SVMs).

### **Artificial Neural Network**

ANNs are inspired by biological neural networks. ANN can learn to solve many types of problems by mimicking the processes of real neurons in the brain [85]. The basic computational units are called neurons or nodes. A simplified ANN incorporates an input layer, a hidden layer, and an output layer. These layers are interconnected. The essential operation of ANN is multiplication, summation and activation. Every input value is multiplied by the individual weight. The Hidden layers of the ANN then sum the weighted inputs, add bias, and process the sum with a transfer function. An artificial neuron passes the processed information via the output layer.

Within the context of voltage stability, ANN can investigate and predict the long-term voltage stability margin represented by the Loadability Margin [28]. ANN suffer from the amount of training time and the scores of the learning parameters. Because of the slow gradient-based learning algorithms with all the parameters tuned iteratively, the training time of feed-forward neural network is in general higher

Below we provide a review table for some of the voltage stability analysis based on ANN.

<b>Proposed</b>	<b>Type</b>	<b>Input</b>	<b>Output</b>	<b>Configuration</b>	<b>Features Selection</b>
Ref [[86]]	MLP	Vol. mag., phase angles and injected active and reactive powers	VSM and real part of critical eigenvalues	A single ANN for different configurations or A separate ANN for each configurations	Self organizing map (SOM)
Ref [[87]]	RBF	Load active and reactive powers	Vol. performance index	A separate ANN for each cluster of input pattern	Class separability index and corr. conditions
Ref [[88]]	RBF	Vol.profile extracted by wavelet transform	VSM	A single ANN for different configurations	Principal component analysis . (PCA)
Ref [[89]]	RBF	Active and reactive line flows	L-index	A separate ANN for each cluster of input pattern	Mutual information
Ref [[90]]	MLP	Vol. mag. Active power of the slack, P-V buses, system loads, & system generators	VSM	A single ANN for different configurations	Gram– Schmidt orthogonalization process
Ref [[91]]	FF BP NN	Complex voltage phasors, Real & Imaginary parts of Bus Vol.	VSM	CPF	Ward reduction
Ref [[92]]	FF BP NN		VSM		Active learning
Ref [[93]]	MLP	Vol. mag., Vol. angles. Rate of active power change	VSM	A single ANN for different Configurations	Actual VSM obtained using: CPF
Ref [[94]]	ANN, SVM, ELM	Vol. mag.	VSM	A single ANN for different configurations	Actual VSM obtained using: CPF

### Support Vector Machine

Support Vector Machine (SVM) is a supervised classification and regression algorithm that use machine learning theory to maximize predictive accuracy without data over-fitting. The goal of SVM is to create the best line or decision boundary (the best decision boundary is called hyperplane) that can separate the data into classes and extend this to non-linear boundaries using kernel trick [95]. The kernel function implicitly maps the training data into a higher-dimensional space where the data is linearly separable. Two similar hyperplanes are constructed on both sides of the hyperplane that separates the data. The classifier generalizes well if the margin or distance between these parallel hyperplanes is larger [96].

Proposed by	Kernel type	Opt. Algorithm	Input	Output	Config.	Feature Selection
Ref [46]	Least Square		Real power & reactive power	VSM	CPF	PCA, MI
Ref [97]	Gaussian (RBF)		Vol. mag. & vol. Angle	VSM	CPF	
Ref [98]	RBF	PSO GA	Vol. mag. & vol. Angle	VSM	CPF	
Ref [99]	Gaussian (RBF)	GA	Vol. mag. & reactive power	VSI	VSI	
Ref [100]	RBF		Vol. mag. & vol. Angle, active, reactive power	VSI	VSI	PCA

Even though SVM has superior features but inappropriate selection of hyper parameters may lead to over fitting or under fitting of SVM model. Various optimization techniques have been used to determine these parameters such as Grid Search method and Genetic Algorithm (GA).

### Classification and Regression Trees

Classification And Regression Trees (CART) [101] have greatly increased in popularity during the past several years. CART represent a methodology for the analysis of large data sets via binary partitioning procedure. This consists of a recursive division on which a response variable and a set of predictors are observed. Such partitioning procedure is known as a regression tree when the response variable is numerical, and as a classification tree when the response variable is categorical [102]. Decision trees can capture the non-linear relationship between the data

with simplicity by clearly identifying the feature importance. However, a DT tends to overfit, and slight variations in the data can generate a completely different tree.

Recently, [103] explored the use of CART for fast evaluation of oscillatory and voltage stability and investigated the impact of the DT growing method and node-set on the classification accuracy. [104] proposed DT utilization for online status appraisal of power grid model by using knowledge database that covers all possible pre-fault operating conditions and decision rules in the form of hierarchical trees for online assessment. Krishnan and McCalley [105] proposed a decision tree-based power system security assessment for multiple contingencies. A contingency grouping technique was used to produce a reduced number of DTs for multiple contingencies. The contingency grouping is based on a newly devised metric that finds the overlap of class boundary progression of various contingency's training databases. [106] developed a combined method for online voltage security assessment by using PCA to reduce the dimensions of PMU measurement data. In [107] the DT-based PCA method is combined with two optimization algorithms, namely biogeography-based optimization and invasive weed optimization, to assess the voltage stability. [108] proposed a new approach based on fuzzy decision trees to assess the voltage security of the power system. The proposed approach's objective is to analyze power system parameters and locate the probable area contributing to voltage collapse.

### **Ensemble Learning**

Ensemble learning algorithms construct a set of base classifiers and then classify new data points by taking a vote of their predictions. The generalization capability of an ensemble is usually more robust than that of base learners [113]. The base learners are generally constructed by perturbing the original training data. Voting is the second stage of ensemble methods which combines the base models built in the previous step into the final ensemble model. Several ensemble machine-learning techniques have been proposed for the online monitoring of voltage stability.

[114] proposed a new online voltage security assessment method based on wide-area measurements and evaluated several machine learning methods for the voltage stability analysis. The study comprised of an evaluation of two groups of machine learning methods – single and Ensemble learning classifiers. Ensemble classifiers include Bagging, XGBoost, Random Forrest and AdaBoost. It was shown that AdaBoost achieved the highest classification accuracy. [115]

Proposed by	DT type	Inputs	Outputs	Config.	Feature Selection
Ref [103]	CART 16	Vol.mag., vol. angle, current Mag., Current Angle	VSM	CPF	
Ref [109]	C4 .5	Vol. mag., vol. Angle, Real Power	Voltage stability boundary	P-V curve	Eigenvalue Analysis, RELIEF
Ref [110]	Classification rules	Vol. mag., vol. Angle, Active Power	Voltage stability boundary		Eigenvalue Analysis, Direct method using CPF
Ref [111]	CART	VSI, Power flow for load consumption	Voltage stability boundary	CPF	
Ref [112]	CART	Active power flows and PMU-based voltage	Voltage stability boundary	CPF	
Ref [104]		Active power flows and PMU-based voltage	Voltage stability boundary		

designed an ensemble by combining a model of Extreme Learning Machine in parallel. The model was initialized by randomly assigning a hidden node number to each Extreme Learning Machine within an optimal range, subjected to a pre-tuning procedure. In the training stage, the single ELMs produce their prediction output. The final prediction result is taken as the average value of individual ELMs.

## 2.2 Voltage stability margin

The additional power that can be transmitted before reaching the voltage collapse point from the current point of operation is usually referred to as the VSM or load margin (LM). Since the LM is easily understandable and because it reflects the proximity to voltage collapse in terms of a measurable and controllable quantity, the system operators can be alerted and automatic remedial actions can be initiated when the LM drops below pre-determined set of critical values.

Although the LM is a good indicator of voltage stability, it is difficult to compute the LM

in real-time using the CPF for a large network, due to the iterative computations involved. Furthermore, the topology and parameters of the network at the current time are required for CPF. Therefore, an alternative strategy is required to predict the LM using real-time power system measurements. The approaches based on machine learning techniques are potential candidates.

## 2.3 Problem formulation

The main objective of online voltage stability analysis (VSA) is to determine whether the current operating point of power system is stable, meeting various operational criteria [116]. Voltage stability is often assessed through Power Voltage analyses [117]. The P-V curve of the system (see Fig. 2.1), in combination with the current operating conditions can be used to obtain the VSM. The P-V curve, as shown in Fig. 2.1 is an important tool in voltage stability analysis, representing the relationship between active power and voltage magnitude in the power system. By simulating load growth from a predefined operating point, the P-V curve is constructed. The curve resembles a human nose and is commonly referred to as the "nose curve". The nose point on the curve indicates the maximum power the system can deliver. As the load increases, the power-voltage point moves along the curve towards the tip of the nose. At the tip, the maximum power that the system can provide is reached. Beyond this point, additional loads cause a drop in voltage and power, and the curve extends towards the lower left corner of the plot, corresponding to the uncontrollable region.

The decline in voltage at the nose point occurs due to the escalating demand for reactive power in the system. As active power increases, the demand for reactive power also rises. However, the sources of reactive power in the system may reach their limits in providing the necessary reactive power. Reactive power is crucial for sustaining voltage levels and supporting inductive loads. Consequently, the system experiences a decline in voltage because it becomes incapable of generating or supplying sufficient reactive power to meet the growing demand. The decline in voltage at the nose point indicates that the system is nearing its stability limit. If the power demand continues to increase beyond this point, the voltage may further decrease, potentially resulting in voltage collapse or instability. System operators monitor the P-V curve, especially the nose point, to evaluate the system's stability margin.

In Fig. 2.1 the nose point or the maximum loading is represented as  $P_{max}$ . The initial state is

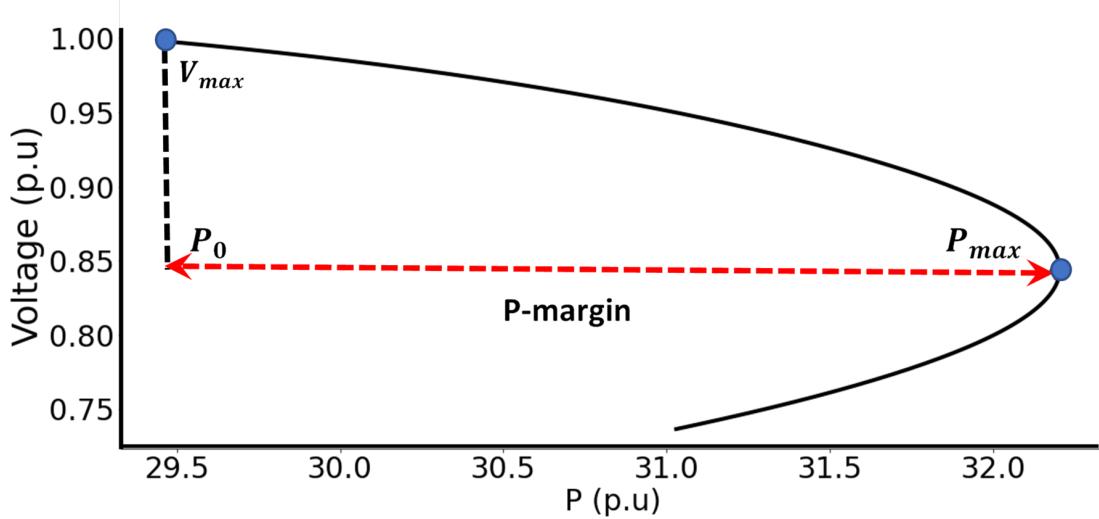


Figure 2.1: The P-V curve determines the loading margin of a power system. The margin between the voltage collapse point ( $P_{max}$ ) and the current operating point ( $P_0$ ) is used as the voltage stability criterion. If the voltage is on the upper side of the red dotted line then total stability is under normal operating conditions. Below the red line the system cannot maintain voltage stability

represented in the figure by  $P_0$ . The Voltage Stability Margin (VSM) [118] can be expressed as follows:

$$\lambda = \frac{P_{max} - P_0}{P_{max}} \geq 0 \quad (2.1)$$

A larger  $\lambda$  value indicates a stable system while a decreasing value suggests closeness towards voltage instability [119].

It should be noted that different system topologies and operating conditions produce different P-V curves. In addition, faults occurring in the system that can lead to topology changes (e.g., line tripping) or changing operating conditions (e.g., generator tripping) can significantly affect the P-V curve and the value of  $\lambda$  for the current operating condition.

### 2.3.1 Framework for Real-Time Prediction of Loadability

Calculating the VSM using conventional methods, like the CPF, can be time-consuming (especially in large-scale systems) and thus ineffective for *online* monitoring solutions. On the contrary, machine learning-based techniques can be used to estimate the VSM almost instantaneously but require heavy *offline* computations for model training, feature selection, and parameter tuning.

	<b>Bagging</b>	<b>Boosting</b>	<b>Feature Ensemble</b>
<b>Description</b>	Builds multiple independent models by training on different subsets of the training data.	Trains multiple models sequentially, where each subsequent model corrects the mistakes of the previous ones.	Selects subsets of features and trains models on those subsets to capture different aspects of the data.
<b>Usage</b>	Classification, Regression, Ensemble Learning	Classification, Regression, Ensemble Learning	Feature Selection, Dimensionality Reduction,
<b>Advantages</b>	Reduces variance, improves stability, handles outliers , improve generalization	Reduces bias, handles complex datasets, improves accuracy	Enhances diversity, improves interpretability
<b>Data Sampling</b>	Random sampling with replacement (bootstrap)	Weighted sampling based on misclassifications	Random subset of features
<b>Model Training</b>	Independent models trained on separate data subsets	Sequential models trained iteratively	Independent models trained on separate feature subsets
<b>Combining Predictions</b>	Averaging or majority voting of individual model predictions	Weighted combining based on model performance	Combining predictions or decisions

Table 2.1: Comparison of different ensemble techniques

This thesis proposes a machine learning approach that shows promise in predicting the VSM during online operation. The technique employs ensemble learning to fit a model on distinct sets of randomly selected features within the training dataset. The objective of ensemble feature selection is to capture the relationship between the input and output pairs obtained from offline simulations. In this case, the inputs correspond to the system measurements provided by the PMUs, such as the bus voltage magnitude and phase angle, while the output is the VSM indicator  $\lambda$ . Subsequently, the method is applied to estimate the VSM during online operation.

In contrast to conventional ensemble methods such as bagging or boosting, which primarily concentrate on forming a diverse subsets of training data or adjusting sample weights to enhance overall model performance, the feature ensemble approach randomly selects distinct sets of features from the training dataset. It's important to emphasize that, unlike boosting and bagging which generate ensemble predictions for classification tasks, the outcome of the ensemble feature selection technique is a subset of features.

In bagging, multiple bootstrap samples are derived from the original training dataset. This

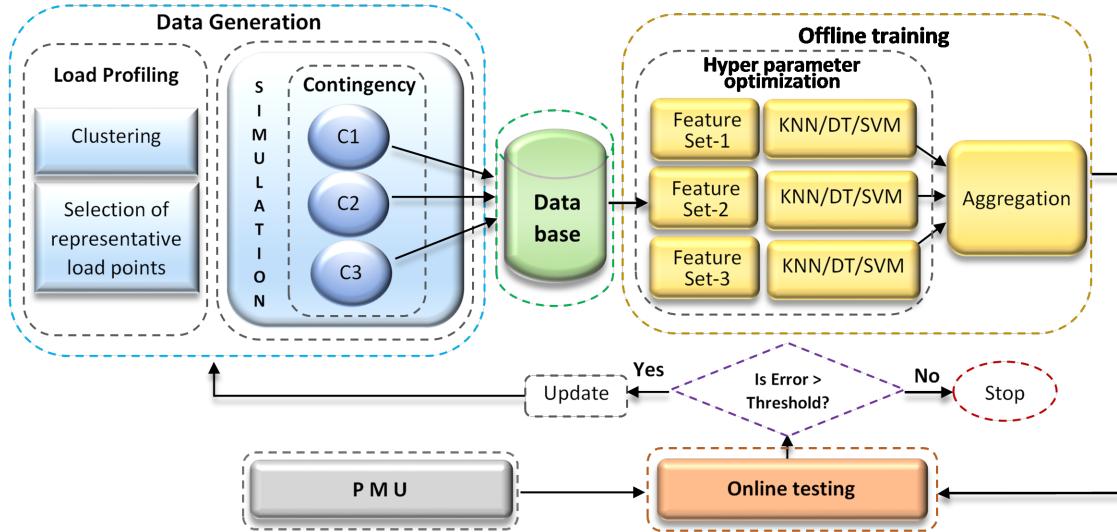


Figure 2.2: An overview of the proposed framework is shown in Fig. 2, in which the planned framework is illustrated with the part enclosed by the dotted lines. It comprises of three stages: 1) Data generation 2) Offline Training and 2) Online Prediction and Update

entails employing homogeneous weak learner models that autonomously learn from each other in parallel, and later combine their insights to establish a model average. In contrast, Boosting strives to construct a robust classifier through a series of weak learners. These weak learners are designed to minimize errors or misclassifications in the training data. Notably, the weights of incorrectly classified instances are amplified, thereby increasing their influence in successive training iterations.

Table 2.1 provides a comparison for the different ensemble approaches.

## Design Methodology

The methodology for constructing a real-time LM predictor is illustrated in Fig. 2.2 via a flow diagram. Initially, a database containing input-output data pairs is generated, followed by cleaning and pre-processing procedures, including the handling of missing values, data format conversion, and data splitting into training and testing sets.

Subsequently, a feature selection process is carried out to eliminate redundant and irrelevant inputs using feature selection techniques like filter or embedded methods to select feature subsets for each model. The feature selection algorithms utilized in the feature ensemble scheme include the F-test, variance threshold (VAR), and Lasso. The F-test and VAR are classified as filter methods while Lasso belongs to the category of embedded methods.

Filter methods evaluate feature relevance using a ranking process and subsequently eliminate features with low scores. These methods are known for their speed, scalability, computational simplicity, and classifier independence. In contrast, Embedded methods perform feature selection as part of the modeling algorithm's execution. They are seamlessly integrated into the algorithm, either as its standard functionality or as an extended feature.

The chosen features are subsequently utilized as input for various regressors, including K-nearest neighbours (KNN), DT, and SVM. However, it's important to note that each regressor is employed individually with different feature selectors. Each regressor brings its own unique modelling capabilities and characteristics. KNN is a non-parametric algorithm that captures local patterns, DT is capable of representing complex decision boundaries, and SVM aims to find the optimal hyperplane for classification or regression. Different regressors may perform better or worse depending on the characteristics of the data. By using a variety of regressors, the ensemble approach becomes more robust to different data distributions, ensuring that at least one regressor is effective in capturing the underlying patterns and relationships. Additionally, by examining the performance of the same predictor with different feature selectors, we can gain insights into the importance and relevance of specific features. If certain features consistently appear in the selected subsets across different feature selectors, it suggests their significance for the predictor. This information can aid in feature interpretation and provide guidance for further analysis.

During training, cross-validation techniques are employed to prevent overfitting. The trained machine learning model, which captures the relationship between input voltage variables and loading margin, is used to predict values on the testing data. The final prediction is generated by combining the outputs of the trained models through a weighted average.

The final phase of the proposed approach involves online testing, where real-time data from PMUs is inputted into the trained machine learning model to predict voltage stability margin. To evaluate the model's performance, various contingencies are applied during this phase to simulate data drift. Additionally, the database is updated using the K-means clustering algorithm. This allows the database to adapt to changing data distributions over time. By incorporating the updated K-means clusters into the database, the model can continuously learn and enhance its predictions based on the most recent information.

For more details on these stages and validation examples, please refer to Chapter-3 and Chapter-

4. The proposed methodology utilizes real-time data from PyRAMSES [120], and the performance of the ensemble model is evaluated using metrics such as R-squared and mean squared error.

## 2.4 Preparation of Training and Testing Data

A comprehensive database of training examples is an essential requirement for the development of the proposed machine learning-based online load margin predicting system. It is not practical to obtain these data from historical measurements as voltage stability events which in reality changed a power system to an unstable state are very rare. Therefore, training data need to be invariably generated through simulations.

### 2.4.1 Overview of Data Generation Process

The process is initiated by generating a large set of initial operating points to obtain data corresponding to a diverse set of operating conditions. These operating points are obtained through a clustering of load profile data for electricity consumption. The clusters are validated with the Silhouette index method. The CPF is performed to trace the P-V curve to the voltage collapse point for each operating point. Subsequently, the VSI described in Section 2.2, and the corresponding LMs are calculated at each operating point. These calculations start from the initial operating point and end at the voltage collapse point. This process is repeated over different operating points under various contingencies to produce a learning database for ensemble training.

### 2.4.2 Operating point generation

The training dataset of P-V curves is generated based on different operating points and fault conditions (contingencies). The consumption patterns from the retailer set of hourly customer load readings over twelve months period are used to extract load operating points. Fig. 2.3 shows the hourly load profile for one year period. The yearlong load profile  $W$  is first separated in a continuous sequence of daily load profiles  $L_j$  built from hourly load data  $H_{d,x}$ .

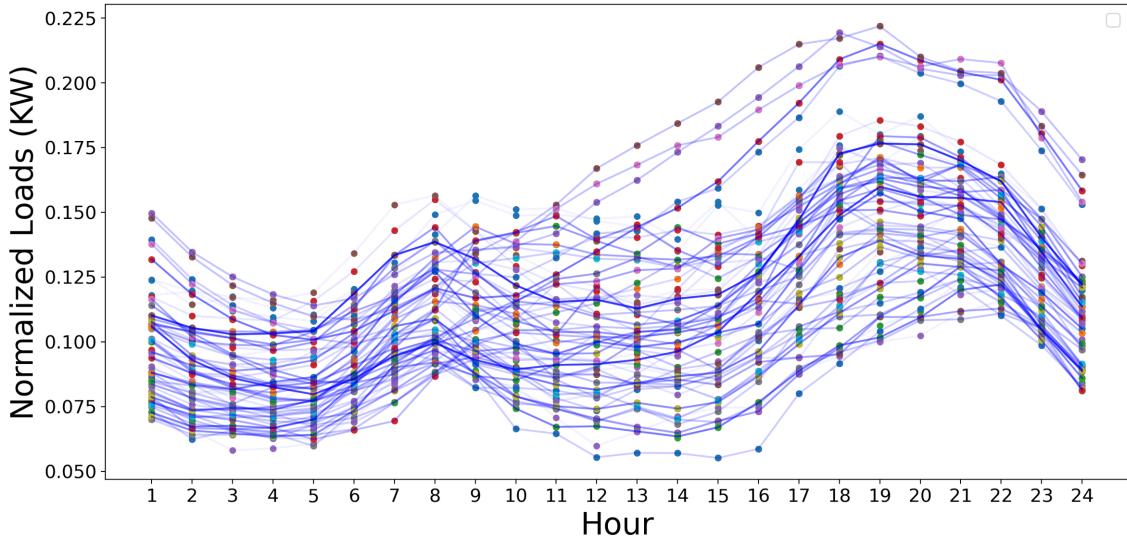


Figure 2.3: DBSCAN: Hourly load profile based on electricity consumption of end users.

$$W = [L_1, L_2, \dots, L_d] \quad (2.2)$$

$$L_j = [H_{j,1}, H_{j,2}, \dots, H_{j,24}]$$

where  $j = 1, 2, \dots, d$  and  $d$  is the number of days.

The feature vectors  $L_d$  can be input to the density-based spatial clustering of applications with noise (DBSCAN) algorithm [121] to group load profiles based on similarity. DBSCAN offers several advantages for load profiling. It can identify clusters of arbitrary shapes, which is beneficial for capturing complex and irregular energy consumption patterns. By considering density, DBSCAN adapts to different load scenarios without requiring prior knowledge of the number of clusters. It effectively handles noisy data by distinguishing outliers, ensuring robustness in load profiling. Additionally, DBSCAN can handle varying cluster sizes, accurately representing load variations across different customer segments.

### Density-based spatial clustering of applications

Let us denote a load profile dataset as  $\mathbf{L}$ , where point  $p \in \mathbf{L}$ , the epsilon parameter (a radius denoted as  $\varepsilon$ ) is usually determined by the user and it has a large influence on the right creation of clusters by this algorithm. The next parameter, i.e. the MinPts ( $\mu$ ) is the minimal number of neighboring points belonging to the so-called core point (see Fig. 2.4). The following definitions will be helpful in determining the DBSCAN parameters.

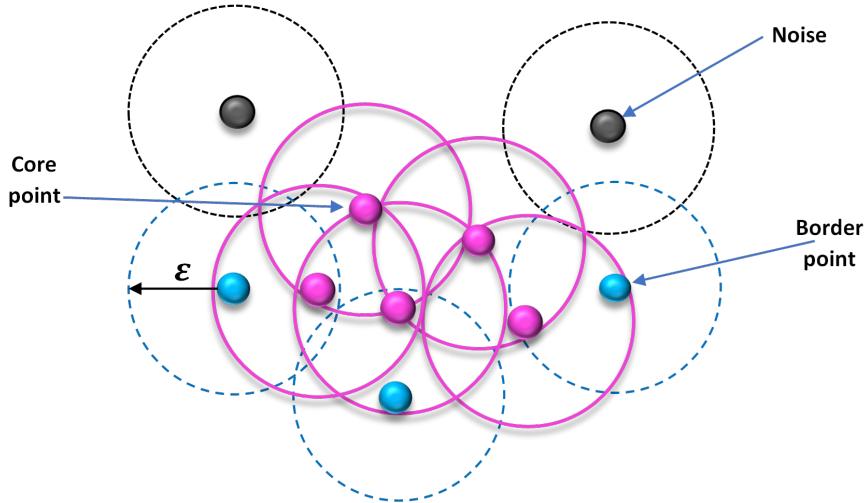


Figure 2.4: Cluster created by DBSCAN with  $\text{MinPts} = 3$ , black dots represent the noise, blue dots represent the border points, and pink dots represent the core points

- **Definition 1:** The  $\varepsilon$ -neighborhood of a point  $p$  in the load profile dataset  $\mathbf{L}$  is denoted as  $N_\varepsilon(p)$  and defined as:

$$N_\varepsilon(p) = \{q \in \mathbf{L} \mid \text{dist}(p, q) \leq \varepsilon\} \quad (2.3)$$

where  $\text{dist}(p, q)$  represents the distance between points  $p$  and  $q$ .

The  $N_\varepsilon$  refers to a region around a data point within a specified radius. It is a key parameter in the DBSCAN algorithm that determines the density of points in a cluster.

- **Definition 2:** A point  $p$  is considered a core point if the number of points belonging to  $N_\varepsilon(p)$  is greater than or equal to  $\mu$ .
- **Definition 3:** A point  $q$  is said to be directly density reachable from a point  $p$  (given  $\varepsilon$  and  $\mu$ ) if  $p$  is a core point and  $q$  belongs to  $N_\varepsilon(p)$ . In other words,  $q$  is directly connected to  $p$  without passing through any other core points.

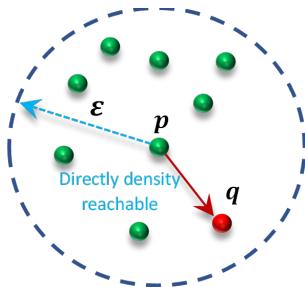


Figure 2.5: Directly reachable points: Point  $q$  is directly reachable from point  $p$ .

- **Definition 4:** If a point  $q$  is directly density reachable from a core point  $p$ , but the number of points belonging to  $N_\varepsilon(q)$  is less than  $\mu$ , then  $q$  is classified as a border point.
- **Definition 5:** Any point that is neither a core point nor a border point is considered noise.
- **Definition 6:** A point  $q$  is density reachable from a point  $p$  (given  $\varepsilon$  and  $\mu$ ) if there exists a chain of points  $q_1, q_2, \dots, q_n$  such that  $q_1 = p$ ,  $q_n = q$ , and each  $q_{i+1}$  is directly density reachable from  $q_i$ . It is worth clarifying that while "directly density reachable" focuses on the immediate connection between two points, "density reachable" allows for a chain of connections between them.

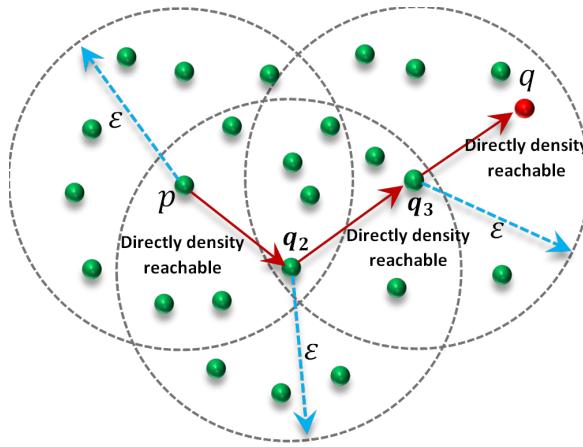


Figure 2.6: Directly density reachable: Point  $q$  is directly density reachable from point  $p$  through a chain of points  $q_1, \dots, q_n$ .

- **Definition 7:** A point  $q$  is density connected to a point  $p$  (given  $\varepsilon$  and  $\mu$ ) if there exists a point  $o$  such that both  $p$  and  $q$  are density reachable from  $o$ .
- **Definition 8:** Cluster  $C$  (given  $\varepsilon$  and  $\mu$ ) is a non-empty subset of  $\mathbf{L}$  and the following conditions are satisfied: first,  $\forall p, q$ : if  $p \in C$  and  $q$  is density-reachable from  $p$ , then  $q \in C$ , next  $\forall p, q \in C : p$  is density-connected to  $q$ .

The DBSCAN algorithm works as follows: Initially, a point  $p$  is randomly selected. If  $N_\varepsilon(p)$  contains at least  $\mu$  points,  $p$  is considered a core point and a new cluster is created. The cluster is then expanded by including all points that are density reachable from the core point. This process continues until no more points can be added to the cluster. If  $N_\varepsilon(p)$  contains fewer than  $\mu$  points,  $p$  is classified as noise. However, this noise point can still be included in another cluster if it is density reachable from a core point.

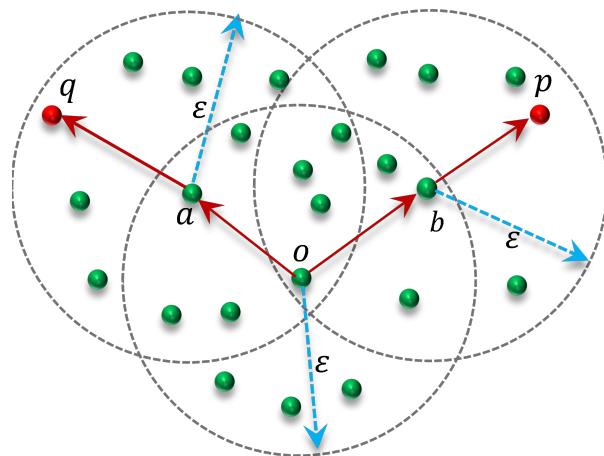


Figure 2.7: Density connected: Point  $q$  is density connected to point  $p$  through point  $o$  as  $p$  and  $q$  are density reachable from  $o$

### Cluster Validation

Assessing the results of a clustering algorithm is known as cluster validity which finds a set of clusters that best fits natural partitions (of given datasets) without any prior class information. A cluster validity index validates the outcome of the clustering process and can be categorized into three classes, namely internal, external and relative cluster validity index.

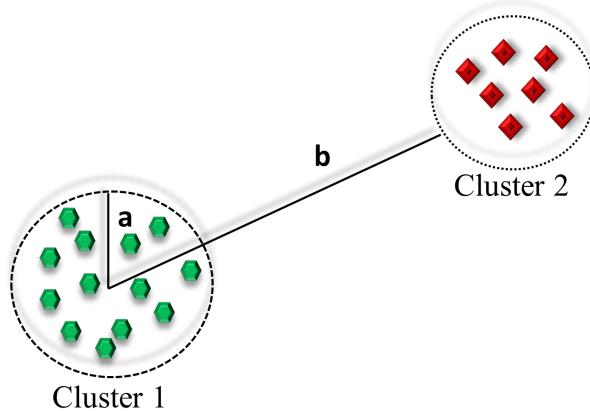


Figure 2.8: Silhouette coefficient example;  $a$  represents average intra-cluster distance i.e the average distance between each point within a cluster while  $b$  represents average inter-cluster distance i.e the average distance between all clusters

In this study, we have used the Silhouette Coefficient, an internal measure that utilizes the notions of intra-cluster similarity or compactness. The silhouette coefficient is a measure of both cohesion and separation of clusters and is based on the difference between the average distance to points in the closest cluster and to points in the same cluster.

The Silhouette coefficient of data object  $i \in C_k$  clusters is given by:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2.4)$$

where  $i = 1, \dots, N$  with  $N$  being total data objects in a given dataset,  $a_i$  denotes the average proximity of data object  $i$  to all other objects in its cluster, and  $b_i$  indicates the smallest average proximity of data object  $i$  to all objects in any other clusters. To measure the overall validity, we take the average over all data objects:

$$S = \frac{1}{N} \sum_i^N s_i \quad (2.5)$$

Where  $S$  is the Silhouette coefficient and ranges from -1 to 1. A value close to +1 indicates that data object  $i$  is much closer to points in its own cluster and is far from other clusters. A value close to zero indicates that  $i$  is close to the boundary between two clusters. Finally, a value close to -1 indicates that  $i$  is much closer to another cluster than its own cluster, and therefore, the point may be mis-clustered.

Furthermore, the optimal number of clusters ( $k$ ) is determined using the Silhouette coefficient. To achieve this, a range of potential cluster numbers ( $k$  values) is selected for evaluation. The DBSCAN algorithm is then applied to the data for each value of  $k$  within the chosen range. The average silhouette coefficient is computed across all data points for each  $k$  value. The optimal number of clusters is typically identified by the highest average silhouette coefficient, and the corresponding  $k$  value at the peak of the silhouette coefficient curve is selected as the optimal choice.

# Chapter 3

## Ensemble Feature Selection

### 3.1 Introduction

In Chapter 2, we discussed how we could obtain data from various operating points, as the model’s usefulness and performance depend on the data used. In this chapter, we want to generate the data based on simulation and train and validate the model offline.

This work proposes a machine learning (ML) algorithm able to map the real-time voltage measurements to the voltage stability margin (VSM) described in Section 2.2. A ML algorithm can find patterns or make decisions based on previous data. It is proposed to use an ensemble feature selection technique of ML models, each of which uses a different feature selection algorithm. Ensemble learning has been successfully applied to classification problems but is also a means for improving other ML functions, such as feature selection. Ensemble feature selection integrates feature selection and ensemble [122]. In a real-world scenario, the benefit is that the user does not have to decide which feature selection algorithm might be the most appropriate for a given problem. Such techniques improve the predictive performance of a single model by training multiple models and combining their predictions [123].

Table 3.1 provides a comparison of feature ensemble techniques with feature ensemble method used for the assessment of voltage stability. Feature ensemble offer several advantages compared to the techniques mentioned in Table 3.1. By leveraging the collective knowledge and diverse perspectives of individual models, feature ensembles can enhance performance. This is achieved by considering multiple subsets of features, which helps diminish the impact of

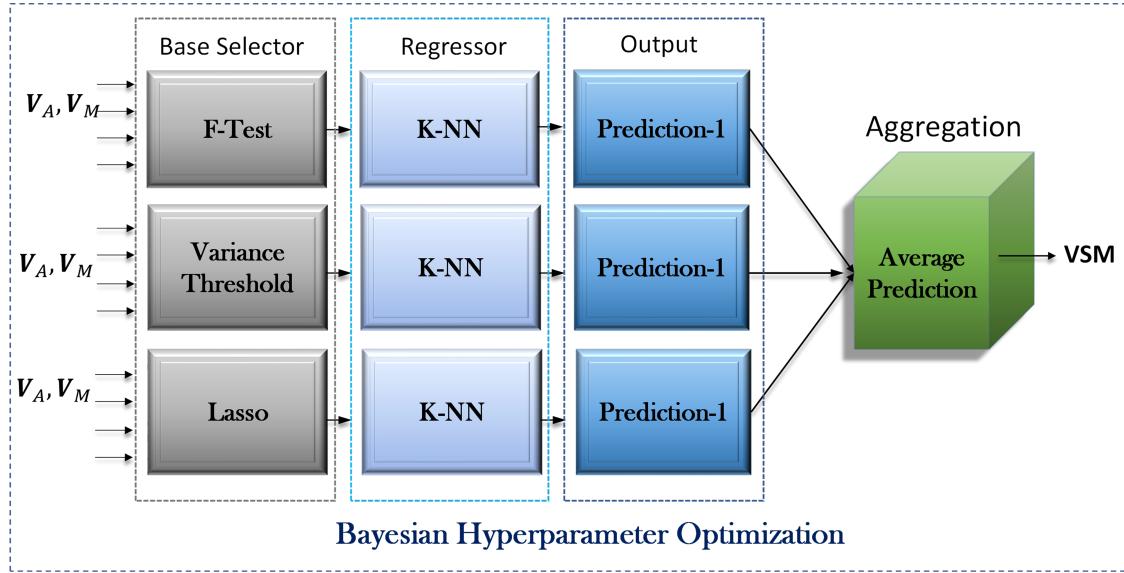


Figure 3.1: Feature-ensemble model with different base selector and same classifier.

noisy or irrelevant features when combining predictions from different models. Furthermore, feature ensemble methods tend to improve generalization performance by reducing overfitting and yielding more accurate predictions on unseen data. In addition, feature ensemble models exhibit greater stability and robustness when faced with changes in the training data, outperforming individual models. By aggregating predictions from multiple models, feature ensemble provide more reliable predictions while minimizing the influence of individual model biases or errors.

Fig. 3.1 shows the formation of feature ensemble models using different feature selection methods but the same training data. This approach is referred as heterogeneous feature ensemble. The objective is to generate N models using different feature selection methods but the same training data. The heterogeneous ensemble accounts for the individual techniques' strengths and weaknesses. The selected features are fed into the ML models to predict VSM. The final output is obtained by taking the average of all the predictions. The pseudo-code of this approach is given below:

We aim to achieve data diversity in our heterogeneous ensemble with three feature selection methods:

1. F-Test/F-Regression (F-Reg)
2. Variance threshold (VAR)
3. Least absolute shrinkage and selection operator (Lasso)

Approach	Methodology	Pros	Cons
<b><i>Correlation-based</i></b>	Calculate correlation between each feature and target variable	Easy to implement	Only captures linear relationships
<b><i>RELIEF algorithm</i></b>	Evaluate the importance of features by considering the difference between nearest neighbors	Considers local feature interactions and relevance to target	Assumes equal importance of all features
<b><i>Mutual Information</i></b>	Measure the mutual dependence between a feature and the target variable	Considers non-linear relationships	Assumes independence between features
<b><i>Genetic Algorithms</i></b>	Use genetic optimization to evolve feature subsets	Can handle large feature spaces	Computationally expensive and sensitive to parameters
<b><i>Recursive Feature Elimination</i></b>	Iteratively eliminate the least important features based on model performance	Captures feature importance	Can be time consuming for large feature sets
<b><i>Feature Ensemble</i></b>	Create multiple subsets of features, train separate models, and combine predictions	Captures diverse perspectives and feature interactions	Requires additional model training and complexity

Table 3.1: Comparison of feature selection algorithms with respect to existing feature selection methods

As mentioned in section 2.3.1, the first two feature selection algorithms can be classified as filter methods, while the latter is classified as embedded. Filter methods rank features by calculating a score for each feature using various statistical metrics. Embedded methods perform the feature selection during the learning process and then derive feature importance from this model, which measures the extent to which the feature contributes most to the prediction [124].

The feature ensemble method is validated using three different ML classifiers separately. The optimal number of features and fine-tuning of the ML models hyperparameter is conducted using the Bayesian optimization technique for the best validation set performance. Experimental results of the methodology show the suitability of the proposed ensembles. The following two sections provide a brief mathematical background of feature selection techniques and the classification algorithms used in the proposed study.

---

**Algorithm 1:** Pseudo-code of the heterogeneous feature ensemble

---

**Data:**  $N$  — number of feature selection methods  
**Data:**  $T$  — threshold of the number of features to be selected  
**Result:**  $P$  — classification prediction

```

for each  $n$  from 1 to  $N$  do
    | Obtaining feature ranking  $A_n$  using feature selection method  $n$ ;
end
 $A_t$  = Select  $T$  top attributes from  $A_n$ 
Build classifier SVM/KNN/DT with the selected attributes  $A_t$ 
 $P$  = Obtain prediction
Combining predictions  $P$  with average combinations

```

---

## 3.2 Feature Selection Techniques

### 3.2.1 F-Test

F-Reg is a uni-variate feature selection technique that calculates the interaction between a feature and a target variable for each feature separately. For each feature, a p-value is calculated, indicating whether that feature significantly impacts the target variable. The objective is to train simple linear regression models of each feature separately to predict the target variable and calculate the F-score. The F-score of a simple linear regression model becomes the F-score of the feature it is trained on. The F-score of each model explains how well that feature predicts the changes in the target variable. That score is used to determine the relevance of that feature to the target variable compared to other features.

To apply the F-test as a feature selection method let the training data set  $\mathbf{D}$  be  $(x_i, y_i)_{i=1}^n$ , where  $x_i \in \mathbf{X} \subseteq \mathbf{R}^m$  and LM be the output denoted as  $y_i \in \mathbf{Y} \subseteq \mathbf{R}$ . The input attribute vector  $x_i$  consists of voltage magnitude and voltage phase angle for  $i$ -th operating point operating point of the n samples.  $y_i$  is the associated target values of VSM. The column values in the data matrix  $\mathbf{D}$  would be treated as a group, and we find the F-ratio, which can be defined as between-group variance over within-group variance. The between-group variance is calculated as:

$$\sigma_{between}^2 = \sum_{j=1}^p \sum_{i=1}^{n_j} (\bar{x}_j - \bar{x})^2 = n_j \sum_{j=1}^p (\bar{x}_j - \bar{x})^2 \quad (3.1)$$

Where  $n$  is the sample size of group  $j$ ,  $\bar{x}_j$  is the mean of the group  $j$ , and  $\bar{x}$  is the overall mean.

The within group variance is calculated as:

$$\sigma_{within}^2 = \sum_{j=1}^p \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad (3.2)$$

Where  $x_{ij}$  is the  $i$ -th measurement of the  $j$ -th group. An F-ratio is then calculated as the ratio between the two variances:

$$F = \frac{\sigma_{between}^2}{\sigma_{within}^2} \quad (3.3)$$

Then, the  $p$ -value based on F-statistic is calculated as  $p\text{-value} = \text{Prob}\{F(j-1, n-j) > F\}$ , where  $j-1$  and  $n-j$  are the degrees of freedom. Features are ranked by sorting  $p$ -value in ascending order. The magnitude of the F-ratio shows the group separation. Features with an F-ratio greater than a specified threshold are retained, while those below the threshold are removed.

### 3.2.2 Variance threshold

In ML, variance is used to measure a set of data spread. VAR is a feature selection method that removes all the low variance features from the dataset. It is an unsupervised method that looks only at the feature values ( $\mathbf{X}$ ) and not the desired output ( $\mathbf{y}$ ).

VAR is a simple and effective method for feature selection that is motivated by the idea that low-variance features contain less relevant information and less value in predicting the response variable. It calculates the variance of each feature and removes those with a variance less than a given threshold.

In the case of VARing, the only hyper-parameter to be tuned is the threshold value of the variance.

### 3.2.3 Least absolute shrinkage and selection operator

Lasso [125] performs two main tasks: regularization and feature selection . Regularization helps in reducing errors and overfitting, while feature selection eliminates unimportant variables that are not associated with the response variables.

Consider the training dataset  $\mathbf{D}$  containing  $(x_i, y_i)_{i=1}^n$  pairs, where  $x_i \in \mathbf{X} \subseteq \mathbf{R}^m$  represents the input attributes for the  $i$ -th operating point among  $n$  samples, and  $y_i \in \mathbf{Y} \subseteq \mathbf{R}$  denotes

the corresponding target value of VSM. The input attribute vector  $x_i$  comprises the voltage magnitude and voltage phase angle, forming an  $m$ -dimensional vector.

In the context of multiple linear regression, the response variable  $y_i$  is often influenced by multiple explanatory variables in  $X$ . This relationship is captured by the equation:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \epsilon_i, \quad (3.4)$$

where  $\beta_0, \beta_1, \dots, \beta_m$  are regression parameters, and  $\epsilon_i$  represents the error term introducing random variation in  $y_i$  not explained by the  $X$  variables. The goal is to optimize  $\beta$  and  $\epsilon$  to minimize the cost function.

In Lasso regression, the cost function is modified by introducing a penalty term:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left( y_i - \sum_{j=1}^m \beta_j x_{ij} \right)^2 + \delta \sum_{j=1}^m |\beta_j|, \quad (3.5)$$

where  $\hat{y}_i$  represents the predicted value and  $\delta$  controls the level of shrinkage. When  $\delta = 0$ , all features are considered, and equation (3.5) reduces to standard linear regression. As  $\delta$  increases toward  $\infty$ , more features are gradually eliminated. The trade-off involves increased bias with higher  $\delta$  values, while variance increases as  $\delta$  decreases. The regularization process adjusts the coefficient values of the regression variables, shrinking some towards zero, thereby neutralizing the impact of irrelevant features in the data [126].

### 3.3 Classification Algorithms for VSM prediction

The feature ensemble produces a subset of features that are used as an input for three different regressors, namely:

- Support Vector Regressor (SVR)
- Classification and regression trees(CART)
- K-nearest neighbor (KNN)

These regressors utilize the selected features to predict VSM. In the context of feature ensemble, single models like SVR, KNN, and decision tree (DT) are preferred choices over ensemble

algorithms like Random Forest or AdaBoost. By using single models, the focus remains on creating diversity through feature selection, while avoiding unnecessary complexity associated with double ensemble. Additionally, using deep learning models within the feature ensemble can increase the timing complexity, making it less desirable in scenarios where computational efficiency is a priority. Table 3.2 presents a comparison of the computational complexity and advantages of using SVR, KNN, and DT algorithms for feature ensemble.

Algorithm	Advantage	Computational Complexity
<b>Support Vector Regressor (SVR)</b>	- Capable of capturing complex relationships	- Training: $O(n^2)$ or $O(n^3)$
	- Good generalization ability	- Prediction: Depends on the number of support vectors
	- Suitable for high-dimensional data	
<b>K-Nearest Neighbors (KNN)</b>	- Considers local dependencies and spatial patterns	- Training: Fast
	- Potentially reveals underlying trends	- Prediction: $O(n \log n)$ or $O(n)$
	- Simple and interpretable	
<b>Decision Tree (DT)</b>	- Captures non-linear relationships	- Training: $O(n * m * \log m)$
	- Interpretable and easy to visualize	- Prediction: $O(\log n)$
	- Efficient for large datasets	

Table 3.2: Comparison of Machine Learning Algorithms for Feature Ensembles

In comparison to other ML methods:

- Deep learning models (e.g., neural networks) have higher computational complexity and training times compared to the mentioned algorithms.
- Ensemble methods (e.g., Random Forest, XGBoost) introduce an additional layer of ensemble and may increase computational complexity.
- Linear models (e.g., Linear Regression) typically have faster training times compared to complex models like SVR or neural networks, they may not be the best choice for capturing the non-linear relationships typically present in power system data.

Please note that the computational complexity mentioned is a generalization and may vary depending on specific implementations, data size, and other factors. Below we provide brief mathematical background for the above mentioned algorithms.

### 3.3.1 Support Vector Machine

Let the training data set  $\mathbf{D}$  be  $(x_i, y_i)_{i=1}^n$ , where  $x_i$  input attribute vector consists of voltage magnitude and voltage phase angle for  $i$ -th operating point of the  $n$  samples.  $y_i$  is the associated target values of VSM that correspond to the  $m$  size of the training data. Using mathematical notation, the regression function takes the form:

$$f(x) = \langle \omega, x \rangle + b \quad (3.6)$$

where  $f(x)$  is the output function,  $\omega$  is the weight vector,  $x$  is the input and  $b$  is bias threshold,  $\langle \cdot, \cdot \rangle$  is the dot products in the feature space.

Where  $\omega \in \mathbf{R}^n, b \in \mathbf{R}$  denote the linear regression in a high-dimensional feature space, which is non-linearly mapped to the input space  $x$ . The objective is to find the values of the weight vector  $\omega$  and bias  $b$  such that the values of  $x$  can be determined by minimizing the following objective function with constraints:

$$R[f] = \sum_{i=1}^l \varepsilon(y_i, f(\bar{x}_i, \omega)) + \frac{1}{2} \|\omega\|^2 \quad (3.7)$$

where  $\omega$  denotes the Euclidean norm,  $l$  denotes the sample size and  $\varepsilon$  is the insensitive loss function which is given by:

$$\varepsilon(y_i, f(\bar{x}_i, \omega)) = 0 \quad |f(x) - y| < \varepsilon \quad (3.8)$$

otherwise:

$$\varepsilon(y_i, f(\bar{x}_i, \omega)) = |f(x) - y| - \varepsilon \quad (3.9)$$

The solution of this problem does not allow any errors. To allow some errors to deal with noise in the training data, the soft margin SVR uses slack variables  $\xi_i^-, \xi_i^+$

Then, the optimization problem can be revised as follows:

$$\min R[f] = U \sum_{i=1}^n (\xi_i^-, \xi_i^+) + \frac{1}{2} \|\omega\|^2 \quad (3.10)$$

subject to:

$$y_i - \omega \rho(x_i) + b \leq \epsilon + \xi_i^-, \quad i = 1 \dots l \quad (3.11)$$

$$\omega \rho(x_i) + b - y_i \leq \epsilon + \xi_i^+, \quad i = 1 \dots l \quad (3.12)$$

$$x_i^- \xi_i^+ \geq 0 \quad i = 1 \dots l \quad (3.13)$$

where  $U$  is a pre-specified value and  $\xi_i^-$ ,  $\xi_i^+$  are slack variables that measure the error.

The slack variables  $\xi_i^-$ ,  $\xi_i^+$  deal with infeasible constraints of the optimization problem by imposing the penalty to the excess deviations which are larger than  $\epsilon$ .

To solve the optimization problem of equations (3.10)-(3.13), we can construct a Lagrange function from the objective function with Lagrange multipliers as follows:

$$L = \frac{1}{2} \|\omega\|^2 + U \sum_i^l \xi_i^- \xi_i^+ - \sum_i^l \alpha_i^- (\epsilon + \xi_i^- - y_i + \langle \omega, x \rangle + b_i) \quad (3.14)$$

$$- \sum_i^l \alpha_i^+ (\epsilon + \xi_i^+ - y_i + \langle \omega, x \rangle + b_i) - \sum_i^l (\eta_i^- \xi_i^- + \eta_i^+ \xi_i^+) \quad (3.15)$$

where  $\eta_i^-$ ,  $\eta_i^+$ ,  $\alpha_i^-$ ,  $\alpha_i^+$  are the Lagrange multipliers which satisfy positive constraints.

Differentiating the Lagrangian function with respect to  $\omega$ ,  $b$ ,  $\xi^-$  and  $\xi^+$ , we can derive the dual problem of equation (3.7) as follows:

$$\max_{\alpha^+ \alpha^-} -\frac{1}{2} \sum_{i,j}^l (\alpha_i^- - \alpha_i^+) (\alpha_j^- - \alpha_j^+) k(x_i \cdot x_j) + \sum_i^l (\alpha_i^- - \alpha_i^+) y_i - \epsilon \sum_i^l (\alpha_i^- - \alpha_i^+) \quad (3.16)$$

subject to:

$$\sum_i^l (\alpha_i^- - \alpha_i^+) = 0 \quad (3.17)$$

$$0 \leq \alpha_i^- \xi_i^+ \leq U \quad i = 1 \dots l \quad (3.18)$$

The kernel function, denoted by  $k(x_i \cdot x_j)$ , is defined in equation (3.16). It calculates the product of two vectors,  $x_i$  and  $x_j$ , in the feature space  $\rho_i(x)$  and  $\rho_j(x)$ . By using the kernel function  $K$ , the inner product in the transformed feature space can be calculated as efficiently as the inner product  $(x_i \cdot x_j)$  in the original input space. Once we have determined the solutions of  $\alpha^-$  and  $\alpha^+$  in equation (3.16), we can use them to obtain the linear regression function:

$$f(x) = \sum_i^l (\alpha_I^- - \alpha_j^+) k(x_i \cdot x) + b \quad (3.19)$$

### 3.3.2 Decision Tree Algorithm

A DT is a widely used supervised learning algorithm. It has a hierarchical tree structure consisting of a root node, branches, internal nodes and leaf nodes.

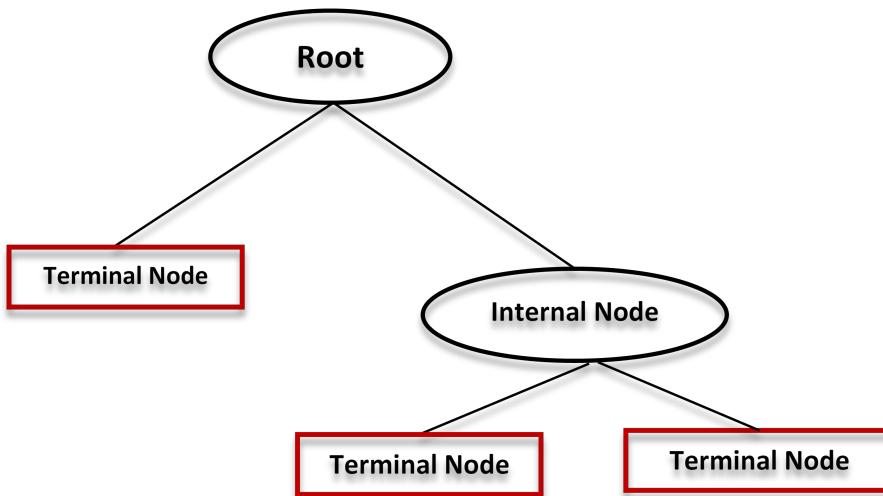


Figure 3.2: Structure of the Decision Tree

As shown in Fig. 3.2, the DT has a root node at the top, which does not have any incoming branches. The root node takes the entire training dataset and is split into two by considering the best attribute and threshold value. The outgoing branches from the root node join the internal nodes, also known as decision nodes. The decision nodes are further split using the same logic mentioned above. Both types of nodes perform evaluations to formulate homogeneous subsets represented by terminal nodes/leaf nodes. The leaf nodes characterise all the possible outcomes within the dataset.

In this study, we utilized the CART method introduced by Breiman et al. [127], which is one of several DT algorithms available, such as Iterative Dichotomies 3 (ID3), CHi-squared Automatic Interaction Detector (CHAID), and Multivariate Adaptive Regression Splines (MARS).

CART is a non-parametric technique that accommodates data with numerical or categorical values and handles missing attribute values. It employs a binary split criterion, meaning that a node in a decision tree can only be divided into two groups. In CART, the Gini index is utilized

as a purity measure for attribute selection. The Gini index quantifies how often a randomly chosen element from a set would be incorrectly labeled if assigned a label according to the distribution of labels in the subset. The attribute that leads to the most substantial reduction in impurity is selected for node splitting. The Gini Index is calculated by subtracting the sum of squared probabilities of each class from one:

$$G(\mathbf{D}) = 1 - \sum_{i=1}^c (P_i)^2 \quad (3.20)$$

Here,  $G(\mathbf{D})$  represents the Gini index of the dataset  $\mathbf{D}$ , and  $P_i$  denotes the probability of class  $i$  among all classes in the entire set  $\mathbf{D}$ . The decision tree continues to grow through successive subdivisions until a point is reached where further divisions do not significantly reduce impurity. At this stage, the node cannot be subdivided any further and becomes a terminal node automatically.

### 3.3.3 K-Nearest Neighbor

$K$ -nearest neighbor (KNN) regressor is based on learning by comparing the given test instances with the training set. Let the training data set  $\mathbf{D}$  be  $(x_i, y_i)_{i=1}^n$ , where  $x_i \in \mathbf{X} \subseteq \mathbf{R}^m$  and output  $y_i \in \mathbf{Y} \subseteq \mathbf{R}$ . Here  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$  is the  $i$ -th instance denoted by  $m$  attributes with its output  $y_i$ , and  $N$  is the number of instances. With the given a test instance  $x_{test}$ . The distances of the new observations  $x_{test}$  with the other observations of the dataset  $D$  is computed by using distance measure  $d_i$ :

$$d_i = dist(x_{test}, x_i) \forall i \in \{1, \dots, n\} \quad (3.21)$$

Using the distance calculation function  $d_i$ , the  $k$  observations from the dataset  $D$  close to  $x_{test}$  are retained. In regression problems, the output  $y$  of  $x_{test}$  is the mean of the outputs of its  $k$  nearest neighbors i.e.  $\hat{y} = \frac{1}{k} \sum_{i=1}^k y_i(x)$ .

## 3.4 Hyperparameter Optimization

The main aim of hyperparameter optimisation (HPO) is to automate the hyper-parameter tuning process and enable users to apply ML models to practical problems effectively. The optimal model architecture of a ML models is expected to be obtained after a HPO process.

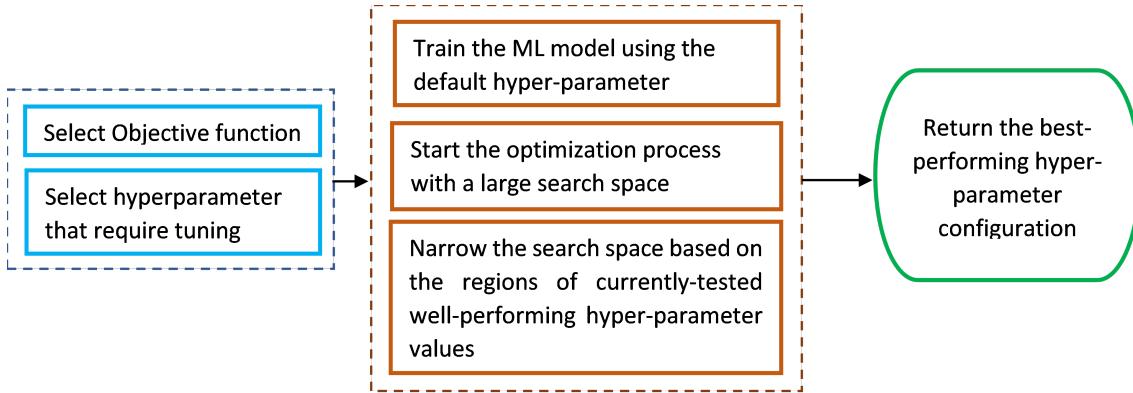


Figure 3.3: Steps for hyperparameter optimization

Some important reasons for applying HPO techniques to ML models are as follows [128].

- It reduces the human effort required since many ML developers spend considerable time tuning the hyper-parameters, especially for large datasets or complex ML algorithms with many hyperparameters.
- It improves the performance of ML models. Many ML hyper-parameters have different optimums to achieve best performance in different datasets.
- It makes the models and research more reproducible. Applying the same HPO process to different ML algorithms promotes reproducibility and facilitates fair comparisons. Fairness in comparing algorithms relies on implementing the same level of HPO, allowing for a meaningful evaluation of their performance. By using a consistent HPO methodology, biases arising from different optimization techniques are eliminated. This ensures that any differences in performance can be attributed to the inherent characteristics of the models rather than variations in the optimization process. The goal is to identify the best-performing model for a specific problem or task by evaluating them under similar conditions. Model complexity, problem-specific considerations, bench-marking, and domain knowledge should also be considered for comprehensive evaluations.

The main steps of HPO are provided in Fig.3.3. However, most traditional optimization techniques are unsuitable for HPO problems since many HPO problems are non-convex or non-differentiable optimization problems and may result in a local instead of a global optimum [128]. Compared with traditional optimization methods like gradient descent, many other optimization techniques are more suitable for HPO problems, including decision-theoretic approaches, Bayesian optimization models, multi-fidelity optimization techniques, and meta-heuristics algo-

rithms [129]. In this thesis, we use Bayesian hyperparameter optimization (BHO) to get the optimal hyperparameters.

### **Bayesian hyperparameter optimization**

BHO leverages Bayesian methods to guide the search for optimal hyperparameters, aiming to strike a balance between exploration (searching different hyperparameter configurations) and exploitation (exploiting promising regions). The motivation behind using BHO can be summarized as follows:

- Efficiency: BHO techniques aim to optimize hyperparameters with minimal model evaluations. Traditional methods like grid search or random search can be time-consuming, particularly when dealing with a large hyperparameter search space. In contrast, Bayesian optimization, utilize probabilistic models to capture the relationships between hyperparameters and model performance. This enables more efficient exploration of the hyperparameter space.
- Prior Knowledge Incorporation: BHO allows for the incorporation of prior knowledge or beliefs about the hyperparameters. This is particularly useful when there is prior information available about the expected ranges or relationships between the hyperparameters. By incorporating such knowledge, the search can be guided towards more promising regions of the hyperparameter space, leading to faster convergence and improved optimization results.
- Handling Noisy Evaluations: BHO can handle noisy evaluations by modeling the surrogate function and using uncertainty estimates to guide the search. It can intelligently balance exploration and exploitation even when evaluations are noisy or have some degree of uncertainty. In contrast, grid search and gradient-based methods do not explicitly account for noisy evaluations, potentially leading to sub-optimal results or requiring a larger number of evaluations.

BHO searches for the best hyperparameter on the domain space  $\Phi$  by using Bayesian optimization [130]. For a given phasor measurement unit (PMU) data of training set and validation set  $G = \{G_{train}, G_{val}\}$  we train a feature selection model involving hyperparameter vector  $\phi$ . The best hyperparameter vector is determined by minimizing the validation error  $E(\phi, G_{train}, G_{val})$ . Generally there are three inputs to BHO [131], (i) a target function  $E(\phi, G_{train}, G_{val})$  which

determines validation error or classification accuracy based on the hyperparameter vector and training/validation datasets, (ii)  $h$  different hyperparameter vectors  $\phi_s = (\phi_1^*, \dots, \phi_h^*)$  and (iii) a limit  $L$  which specifies the number of candidates of hyperparameter vectors to search the best configuration.

The BHO searches a minimum, gradually accumulating  $(\phi_s^*, E_1(\phi_s^*))$  with  $s$  increasing and returns the best configuration of hyperparameters  $\phi^b$ . Using a predictive distribution, the BHO guides the search to only focus on the areas of the input space that are expected to provide the most useful information about the solution to the optimization problem. Starting with a set of initial hyperparameter vectors  $\{(\phi_1^*, E_1), \dots, (\phi_h^*, E_h)\}$ , a surrogate function model  $F_{surrogate}$  is fitted to the data with the accumulated set of hyperparameter vector and its corresponding validation error.

In this thesis, the Gaussian process regression model  $F_G$  serves as a surrogate function that approximates the landscape of  $E$  over the space  $\Phi$ . BHO utilizes all the information in the history (reflected by the built surrogate model) to determine what will be sampled next. Thus, the next hyperparameter is sampled at the place optimizing an acquisition function  $A(\phi|F_G)$  at which the validation error  $E$  is evaluated.

## 3.5 Case Study

A comprehensive stability database is necessary to train MLMs. Such databases can be obtained in power systems by performing simulations on various scenarios and fault recording. In the present thesis, we generate an artificial database by inputting clustered load profiles derived from historical electricity consumption into PyRAMSES [132]. PyRAMSES is a Python module that facilitates the use of the dynamic simulator RAMSES and provides advanced integration like defining test cases, running a simulation, and extracting information. For this thesis, we employed Electromagnetic Transients simulation (EMT). EMT captures the dynamic and transient behaviour of power systems. It models the electrical quantities as time-varying waveforms, taking into account the detailed electromagnetic and electromechanical dynamics of the system. EMT simulations are used to study system responses to faults, switching events, and other transient disturbances.

Moreover, a test system is crucial for understanding the phenomena under investigation, as it

provides a controlled and simplified representation of a real power system. In this study, the Nordic test system [133] is employed, which is widely utilized for long-term voltage stability analysis, including scenarios that may lead to system collapse. By employing the Nordic test system, we can effectively analyze and comprehend the dynamics of voltage stability in a realistic and controlled environment.

The single-line diagram of the test system is provided in Fig. A in Appendix. A.1. The system is composed of four areas:

- 'North' with hydro generation and some load.
- 'Central' with much higher load and thermal power generation
- 'Equiv' connected to the North, which includes a simple equivalent of an external system
- 'South' with the thermal generation, which is loosely connected to the rest of the system

Altogether there are 74 buses in the network, of which 32 transmission buses, 22 distribution buses, and 20 generator buses. There are 102 branches, among which 22 distribution and 20 step-up transformers. The nominal frequency is 50 Hz. The system carries long transmission lines with 400 kV nominal voltage and is burdened with heavy inter-area transfers from North to Central.

### 3.5.1 Data generation

DBSCAN algorithm is used to extract the load operating points from the historical electricity consumption load profile data [134]. The load profile dataset contains 8760 operating points (365 days times 24 hourly consumption data) for 2018. Regarding the DBSCAN algorithm, the difficulty lies in choosing proper values for parameters  $\epsilon$  and  $MinPts$ .

We have used the nearest neighbors method to reach a fair estimation for  $\epsilon$ . The technique calculates the average distance between each point and its  $k^{th}$  nearest neighbor. Fig. 3.4-(a) depicts the result of a distance plot of the load profile data points sorted in ascending order to the 20-th nearest neighbor. The angle bent the most is selected as the  $\epsilon$  value and is found to be 0.005. Once the  $\epsilon$  value is found, the Silhouette index is applied to find the  $MinPts$ . Fig. 3.4-(b) shows the results of applying the Silhouette index, iterated for different sample sizes. The red line shows the maximized scores for the Silhouette index. Based on the score, we can choose a minimum sample size of 4 or 5. With  $\epsilon=0.005$  and  $MinPts=5$ , ten clusters were obtained. The

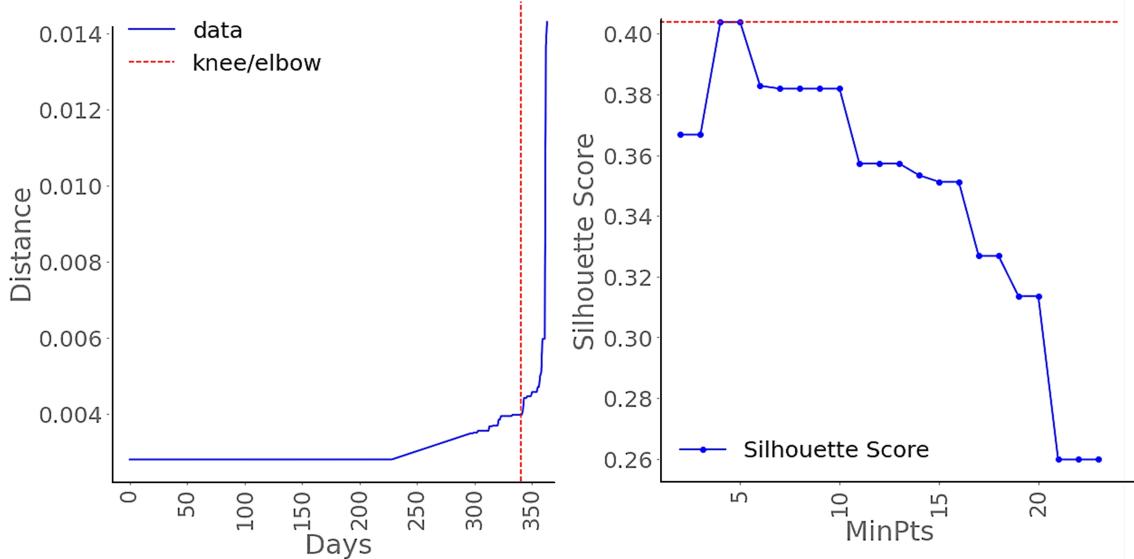


Figure 3.4: a) Elbow plot for K selection, b) Silhouette Score

clustering results of the method in this thesis are shown in Fig. 3.5. The representative members from the clusters were obtained using mean, minimum, and maximum. After clustering 8760 profiles with DBSCAN, 720 representative load operating points were extracted.

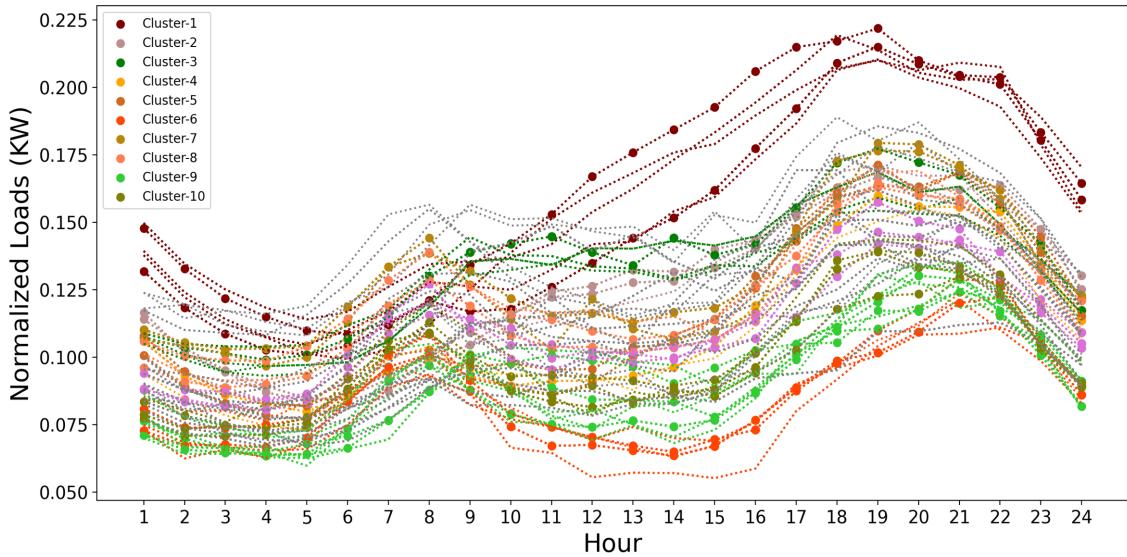


Figure 3.5: Cluster of operating points

## Dynamic Simulations

Starting from the 720 load operating points, dynamic simulations were carried out to generate the corresponding P-V curves. At the beginning of the simulation, the loads were set to the initial operating point  $P_0$ . The loads in the system were uniformly and slowly scaled up while the power factor was kept constant. This approach is used instead of a CPF analysis since

it allows us to capture better the dynamics of the generators and loads (OXL or OEL limits, saturation, etc.).

Dynamic simulations in RAMSES involve solving a set of time-domain differential equations that represent the dynamic behavior of various components in the power system. These simulations require smaller time steps compared to steady-state simulations, as they need to capture the system's response at different time intervals during transients. The computational implications of dynamic simulations are typically more demanding compared to steady-state analysis. The simulations may require higher computational resources and longer simulation times due to the increased complexity and time resolution needed to model the dynamic behavior accurately.

Additionally, conducting multiple dynamic simulations to generate P-V (Power-Voltage) curves for the 720 load operating points can be computationally intensive. However, dynamic simulations offer valuable insight into the system's stability and dynamic behavior, allowing for a better understanding of the system's response under various operating conditions and disturbances.

### **Conversion of clustered load profile into system loads**

Table 3.3 shows the original loads from the Nordic test system. Loads of the highlighted buses are iteratively scaled during the simulation process with the 720 load operating points from the DBSCAN load clusters. Scaling is performed to adjust the values of the clusters to a specific range that is suitable for multiplication with the load data. By scaling the clusters, their values are transformed to be on a similar scale as the load data. This ensures that the multiplication and combination of the scaled clusters and load data are meaningful. In this case, the scaling range is set to (0, 0.09). The scaled clusters are then multiplied with the corresponding load data to produce the desired results.

Table 3.4, shows an example where the highlighted loads are converted after scaling with the clustered load operating points. Starting from 720 load operating points, dynamic simulations were carried out for database generation.

### **Calculation of P-V margin**

At the beginning of the simulation, the loads were set to the initial operating point  $P_0$ . Subsequently, the loads were uniformly scaled up while keeping the power factor constant. This scaling of loads increases the power demand on the system, which can lead to a voltage drop

No.	Bus name	vnom	pload	qload	bshunt	qshunt
1	1	20	600	148.2	0	0
2	2	20	330	71.0	0	0
3	3	20	260	83.8	0	0
4	4	20	840	252.0	0	0
5	5	20	720	190.4	0	0
6	11	20	200	68.8	0	0
7	12	20	300	83.8	0	0
8	13	20	100	34.4	0	0
9	22	20	280	79.9	0	0
10	31	20	100	24.7	0	0
11	32	20	200	39.6	0	0
12	41	20	540	131.4	0	0
13	42	20	400	127.4	0	0
14	43	20	900	254.6	0	0
15	46	20	700	211.8	0	0
16	47	20	100	44.0	0	0
17	51	20	800	258.2	0	0

Table 3.3: Parameters of Nordic buses

and voltage variations. By capturing system measurements during this load scaling process, the P-V curve is generated to analyze voltage stability. This approach provides valuable insights into voltage stability margins and the behavior of voltage levels as the load increases. It allows for controlled analysis of the system's response to changing load conditions, enabling the evaluation of its capacity to handle load variations while maintaining voltage stability. Furthermore, uniform scaling of loads helps capture system limits and saturation effects, such as voltage collapse or equipment saturation near operating limits. This information is crucial for identifying system vulnerabilities and establishing appropriate operating limits. With this approach, we generate data that relate the system measurements to the VSM.

Furthermore, two worst-case contingencies [135], namely Fault-1 and Fault-2, were considered to investigate their impact on the P-V curves and generate more VSM data under faulted conditions. Fault-1 is a short-circuit of 100 ms on transmission lines 4031-4041. This transmission line pass between the 'North' and the 'Central' area and is close to bus 4022. The fault is cleared by tripping the line. Fault-2 takes place near bus 4012 in the 'North' region by tripping generator g10.

Given the 720 operating states, the two contingencies are simulated, yielding two more datasets. The generated datasets are stored in a database with input vectors comprising voltage magnitude and phase angle of all the buses. The output vector/target is the  $P_{Margin}$ .

Bus name	vnom	pload	qload	bshunt	qshunt
1	20	607.68	150.1	0	0
2	20	334.22	71.91	0	0
3	20	263.33	84.87	0	0
4	20	850.75	255.22	0	0
5	20	729.21	192.84	0	0
11	20	200	68.8	0	0
12	20	300	83.8	0	0
13	20	100	34.4	0	0
22	20	280	79.9	0	0
31	20	100	24.7	0	0
32	20	200	39.6	0	0
41	20	546.91	133.8	0	0
42	20	405.12	129.3	0	0
43	20	911.52	257.86	0	0
46	20	708.96	214.51	0	0
47	20	101.28	44.56	0	0
51	20	810.24	261.5	0	0

Table 3.4: Parameters of Nordic buses

Fig. 3.6 illustrates three P-V curves, both before and after contingencies. The base case is represented by the blue curve, while the red and green curves correspond to Fault-2 and Fault-1, respectively. In this scenario, the base case refers to node 1041, which is identified as the critical bus. The critical bus is determined by analysing the variation in voltage magnitudes across buses in the power system. The bus with the highest variance in voltage magnitude is considered more susceptible to voltage instability and is designated as the critical bus. Fig. 3.7 displays the variance of voltage magnitudes for transmission and distribution buses. It is evident that bus-1041 exhibits the highest variance.

In Fig. 3.6 , at the initial operating point  $P_0$ , the base load's active power is 29.58 pu, while at the critical point, the maximum active power  $P_{\max}$  reaches 32.40 pu. Consequently, for this initial operating point, the load's  $P_{margin}$  is calculated as 2.82 pu. The load margin for the post-contingency P-V curves is smaller than that of the pre-contingency curve (blue) due to changes in the system's topology and characteristics following the disturbance (contingency) in the network.

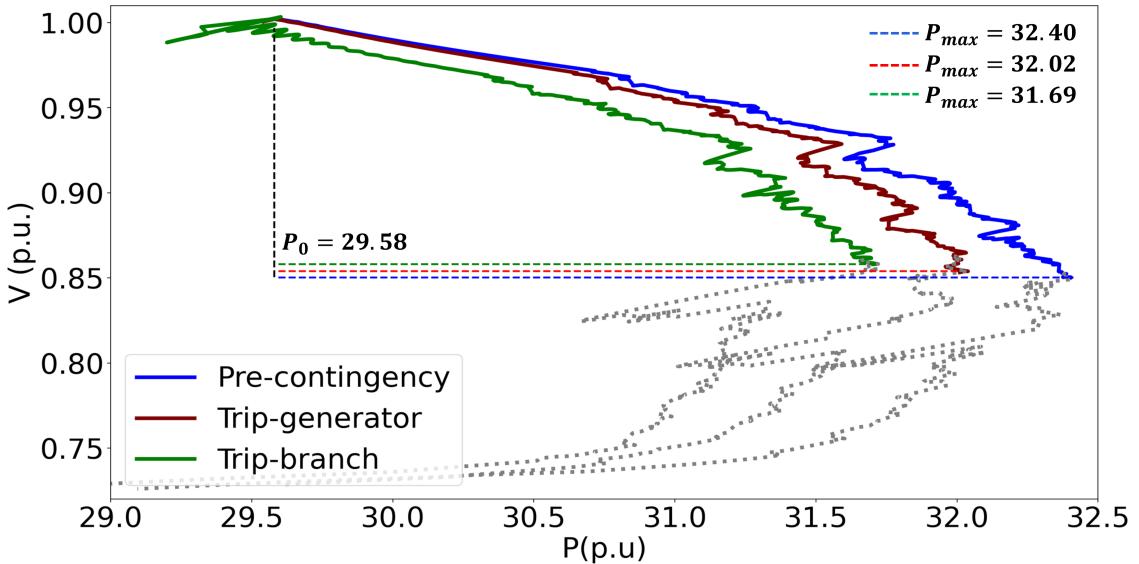


Figure 3.6: The P-V curves are obtained by gradually increasing all loads in the Central area. The blue curve shows when the system operates without a contingency. The margin is higher as the power is transferred without endangering the limits of transmission lines and the bus itself. After the contingency is applied (red curve and green curve), the reduction of VSM occurs. As noticed, in the contingency case, the critical level is reached more quickly, resulting in the loss of voltage stability.

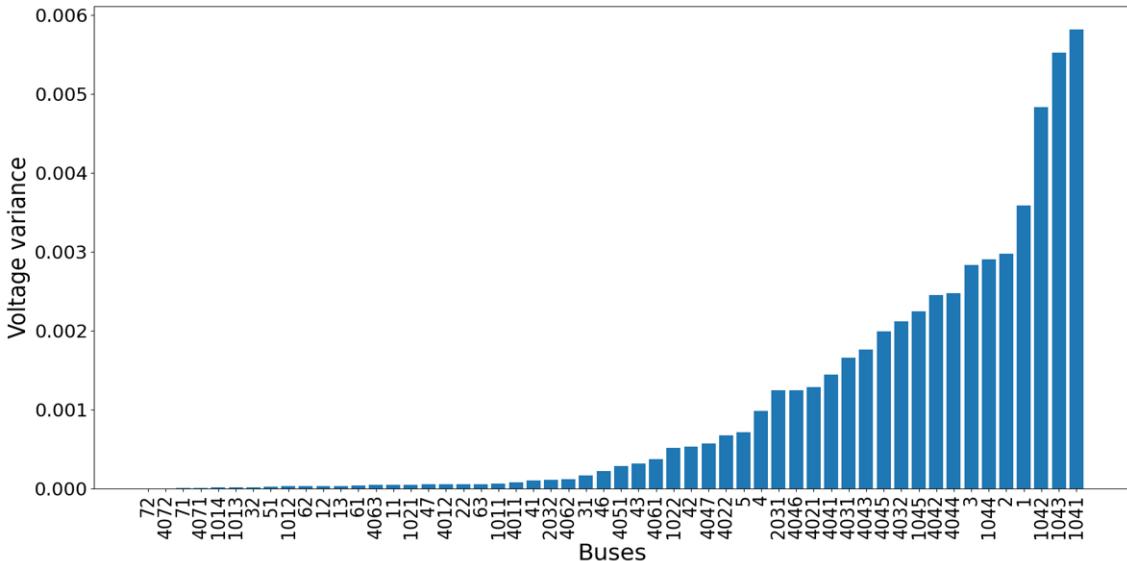


Figure 3.7: Variation of voltage magnitude across all buses

Fig. 3.8 shows the PV margin for the base case assuming a uniform percentage load increase for all the loads. In Fig. 3.8, the curve labelled 1041 is a load bus, and this bus reached its active power limit around  $\lambda = 0.0885$ . After this point, its voltage drops rapidly, leading to eventual voltage collapse. At initial point  $P_0$ , the base load active power is 29.45 MW. At this critical point ( $\lambda = 0.0885$ ), the maximum load active power is 32.31 MW. Therefore for this

initial operating point, the load P margin is 2.861 MW. Similarly the  $P_{Margin}$  is calculated for each point on the PV-curve i.e.  $P_{(Margin_i)} = P_{max} - P_i$ , where  $i = [0, \dots, N]$ .

Many such computations starting from randomly generated different operating points were carried out to generate the training and testing data for designing the feature ensemble.

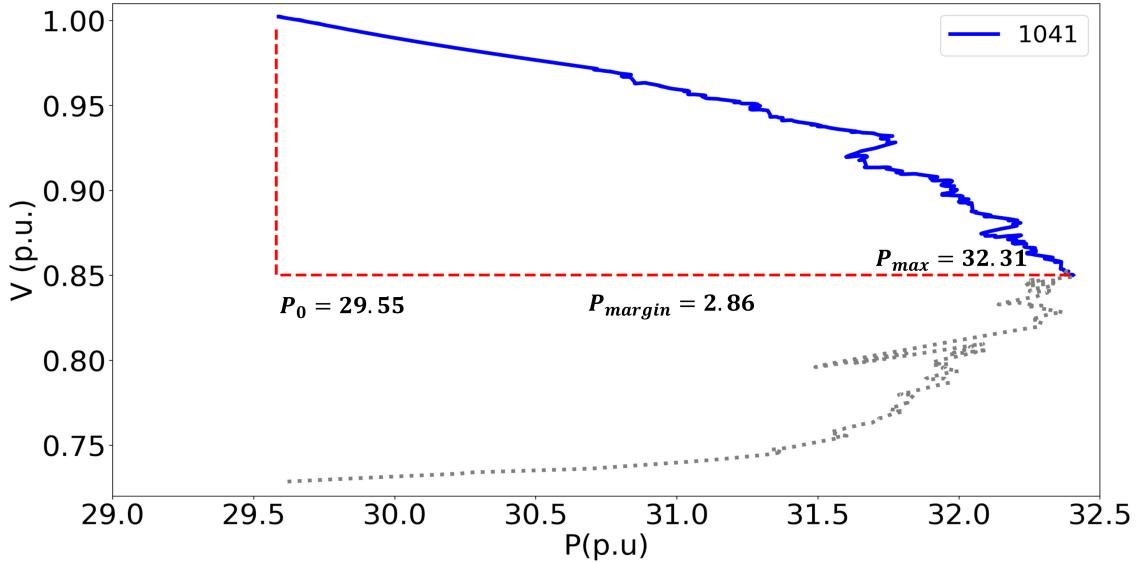


Figure 3.8: PV curve generated for initial load operating point

### 3.5.2 Structure of the Data

In the data generation process, cases with different initial operating points and contingencies are considered. Therefore rows of the database are generated at each operating point for each of these cases. The factors which define the created data base size such as; number of initial operational points considered, number of contingency cases used in addition to base case and the total number of data points generated through resampling of CPF are tabulated in Table 3.5.

Table 3.5: Dataset information

Test System	Number of initial operating points	Number of columns	Number of rows
Nordic Test System	720	154	2847124

The generated data comprised three dimensions, namely 1) time, 2) location, and 3) type of variables. Every value recorded by PMU has a timestamp, a location represented by buses, for instance, Generator Buses or Transmission Buses, and a measurement variable, in this case, Voltage Magnitude and Voltage Phase Angle. The variables measured by a PMU are a power

system's voltage and current phasors. If we denote the number of available PMU by  $p$ , each provides  $m$  measurements, then a total of  $N = p \times m$  measurements is collected at each time sample. This study  $p = 2$  corresponds to phase angle (A) and voltage magnitude (M). Table-3.6 shows a set of  $n$  such samples, each taken at different times.

Time	Generator Buses				Transmission			Distribution			Target <b>VSM</b>	
	<b>g1_m</b>	<b>g1_a</b>	...	<b>g20_m</b>	<b>g20_a</b>	<b>1011_a</b>	...	<b>4072_m</b>	<b>12_m</b>	...	<b>72(M)</b>	
<b>0</b>	1.05	9.298	...	1.02	-2.94	-0.079	...	1.058	0.98	...	0.996	4.55
<b>0.02</b>	1.05	9.316	...	1.02	-2.91	-0.0619	...	1.058	0.98	...	0.996	4.55
<b>0.04</b>	1.05	9.335	...	1.02	-2.87	-0.043	...	1.058	0.98	...	0.996	4.55
:	:	:	:	:	:	:	...	vdots	:	:	:	:
<b>20</b>	1.06	18.04	...	1.02	13.85	7.615	...	1.057	0.98	...	0.995	4.55

Table 3.6: Structure of Data

For the prediction of long term VSM we do not require timestamp therefore we drop the time variable. The above data can be represented as  $n \times d$  matrix, with  $n$  rows and  $d$  columns, where rows corresponds to observations in the dataset, and columns represent attributes or voltage measurements, including voltage magnitude and voltage phase angle. Each row in the data matrix records the observed attributes values for a voltage measurements. The  $n \times d$  matrix is given as:

$$\mathbf{D} = \left( \begin{array}{c|cccc} & X_1 & X_2 & \dots & X_d \\ \hline x_1 & x_{12} & x_{13} & \dots & x_{1d} \\ x_2 & x_{22} & x_{23} & \dots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n1} & x_{n2} & \dots & x_{nd} \end{array} \right) \quad (3.22)$$

where  $\mathbf{x}_i$  denotes the  $i$ -th rows. Each row may be considered as a  $d$ -dimensional column vector:

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T \in \mathbb{R}^d \quad (3.23)$$

and  $\mathbf{X}_j$  denotes  $j$ -th column which can also be treated as a vector in  $n$ -dimensional space  $\mathbb{R}^n$ :

$$\mathbf{X}_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

As we can see, we can consider the entire dataset as an  $n \times d$  matrix, or equivalently a set of  $n$  row vector  $\mathbf{x}_i^T$  or as a set of  $d$  column vector  $X_j \in \mathbb{R}^n$ .

The objective of ML models is to predict a target/output or response variable  $y_i$ , given an input

feature vector  $\mathbf{x}_i$ . The model's output is a function  $f$  that automatically produces a prediction. The predicted output is denoted with  $\hat{y}$ , for any set of predictors:  $\hat{y}_i = f(\mathbf{x}_i)$ .

The generated dataset is used for training and validation of the feature ensemble model. The data is split into training and testing sets and 3-fold cross validation was conducted. Cross-validation involves running multiple iterations of model training and evaluation. Higher fold values (e.g., 10-fold) would require more iterations, leading to increased computational resources and time. Using 3-fold cross-validation provides a good compromise for obtaining reliable estimates of model performance without excessive computational burden. In each 3-fold cross validation the data is partitioned into 3 subsets of equal size and the results are averaged over 3 runs. In each run, a distinct subset is used for testing, while the remaining instances are provided as training data.

## 3.6 Model Performance Evaluation

In order to evaluate the performance of the model, this study uses following indicators are used:

### **Mean squared error**

The mean squared error (MSE) is the average squared distance between the observed and predicted values. The MSE tells you how close a regression line is to a set of points.

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n} \quad (3.24)$$

Where  $y_i$  is the i-th observed value,  $\hat{y}_i$  is the corresponding predicted value and  $n$  is the number of observations in the dataset.

### **Root mean squared error**

While MSE represents the difference between the original and predicted values which are extracted by squaring the average difference over the data set. It is a measure of how close a fitted line is to actual data points. Root mean squared error (RMSE) on the other hand is the error rate by the square root of MSE. It is defined as the square root of the average of the squared errors.

Classifier	MSE	RMSE	R2	Training_time (S)	Testing_time (S)
<b>KNN</b>	0.000019	0.00432	0.999999	61.662174	2.216540
<b>SVR</b>	0.002552	0.050515	0.998261	448.202024	0.429013
<b>DT</b>	0.000387	0.019682	0.999736	1273.867204	0.008904

Table 3.7: Performance of Individual Classifiers

$$RMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} \quad (3.25)$$

### Coefficient of determination

R-squared or Coefficient of determination is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (3.26)$$

Where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

The testing and validation of the proposed feature ensemble is carried out in four phases.

1. **Phase 1:** In the first phase, the performance of each classifier is evaluated without using any feature selection method.
2. **Phase 2:** In the second phase performance of the proposed feature selection algorithm (F-Reg, Lasso, VAR) is compared with other similar feature selection methods.
3. **Phase 3:** In the third phase performance of the feature ensemble method is evaluated and compared with the performances of classifiers evaluated during the first phase.
4. **Phase 4:** In the final phase Robustness of the feature ensemble against noisy data is evaluated by introducing three different noise levels.

#### 3.6.1 Phase-1: performance of the classifier without feature selection

In this work, KNN, DT, and SVR are used to conduct experiments, compare and find the best algorithms suitable for estimating the VSM. Table- 3.7 shows the average evaluation metrics of each ML algorithm without feature selection. MSE represents the average squared difference between the original and predicted values in the data set. It measures the variance of the residuals, while RMSE is the standard deviation of residuals. The  $R^2$  represents the proportion

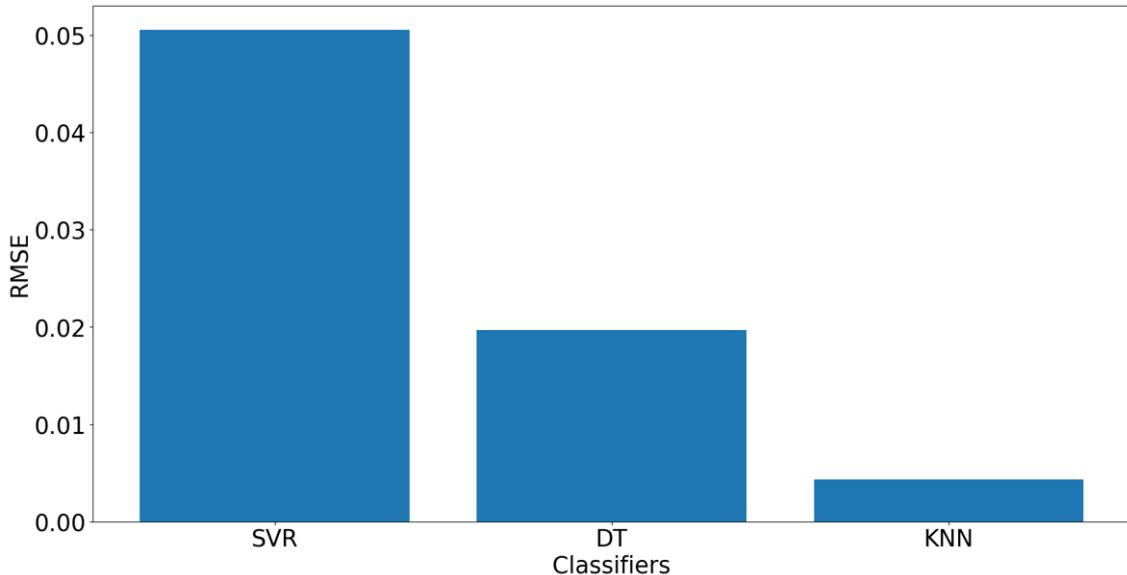


Figure 3.9: Comparison of classifier performance

of the variance in the dependent variable, which the linear regression model explains.  $R^2$  can have negative values, which mean that the regression performed terribly.  $R^2$  can have value 0 when the regression model explains none of the variability of the response data around its mean. Analysis of  $R^2$  score reveals the score closer to 1 meaning perfect prediction.

Performance results for the RMSE for the three models are plotted in Fig. 3.9 which indicates that KNN with an RMSE of 0.00432 gives the best results compared with the other two classifiers, Table- 3.7 also show the training time and testing time taken by these three algorithms. These values indicate the computational time taken by each classifier during the training and evaluation process.

The training time (S) refers to the amount of time, measured in seconds, that each classifier took to complete its computations and training on the dataset. It represents the total processing time consumed by the respective classifier to perform the required tasks, such as fitting the model while the testing time is the time required in making predictions.

Fig. 3.10 shows the actual vs predicted results for all three classifiers. All the model plots are very accurate, however the best result is achieved by KNN. There's a strong correlation between the model's predictions and actual results. Fig. 3.11 is the residual plot for the same classifiers where the prediction made by the model is on the x-axis, and the accuracy of the prediction is on the y-axis. The distance from the line at 0 is how bad the prediction was for that value. Since.

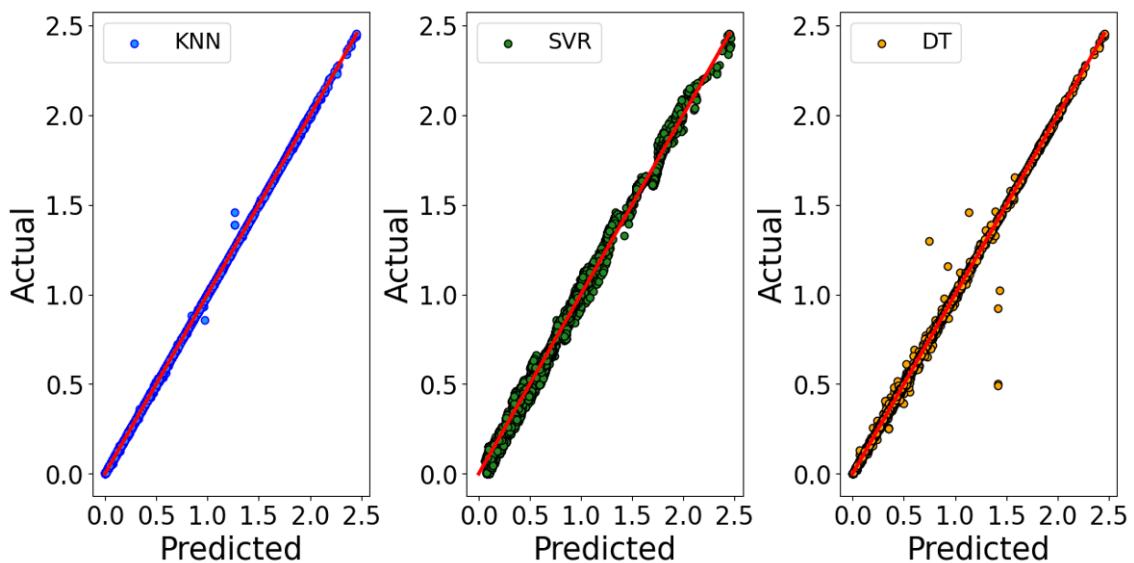


Figure 3.10: Comparison of classifier performance

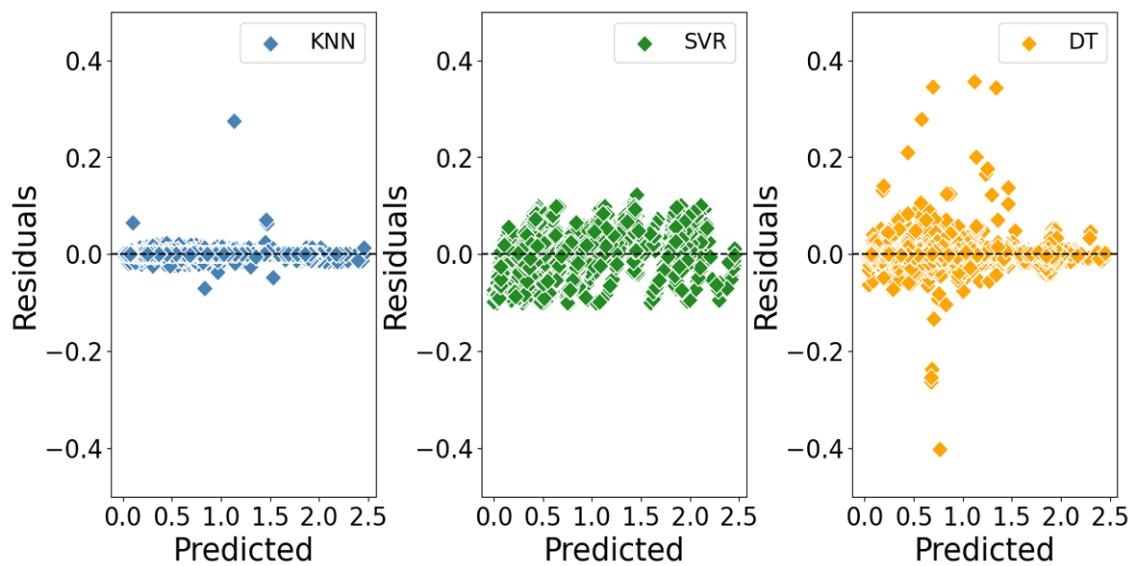


Figure 3.11: Residual plot for the classifier performance

$$\text{Residual} = \text{Observed} - \text{Predicted} \quad (3.27)$$

The positive values for the residual (on the y-axis) mean the prediction was too low, and negative values indicate the projection was too high; 0 means the estimation was correct. In general, the plots are symmetrically distributed, tending to cluster towards the middle of the plot.

### 3.6.2 Phase-2: performance comparison of the proposed feature selection algorithms with similar feature selectors

This section describes the of performance comparison of the proposed feature selection algorithm i.e. F-Reg, VAR and Lasso with other similar filter and embedded methods. We have compared the performances with MI, Ridge, FFE and RFE.

The target of feature selection is to select features highly correlated with the class label in the data set and delete irrelevant and redundant features. As mentioned above, the collected data is normalized and randomly divided into training and testing sets. The training set is employed for feature selection, and the testing set is utilized to check the quality of the selected feature subset.

This process of feature selection method for the VSA can be described in two stages. In the first stage, the feature score is calculated with the training set and used for measuring the relevance between features and classes. The features are ranked from large to small based on the score. In the second step, the ranked features' classification performance is calculated using the KNN, SVR, and DT models, and MSE, RMSE, and  $R^2$  is calculated.

Table 3.8 shows the performance of different feature selection algorithms and the number of features selected. The total number of features selected by each feature selection is based on the optimal hyperparameter configuration from the Bayesian optimization described in the section 3.4. The best scores for the hyperparameters are provided in Table B.1 in the Appedix B.

It can be seen from Table 3.8 that the proposed feature selection algorithms achieve significantly better results in terms of MSE and RMSE,  $R^2$ . Based on the RMSE values from the analysis in Table 3.8, it is evident that the combination of Lasso with KNN achieved the lowest RMSE value (0.00259). This result suggests that this specific combination exhibits the best performance among all the feature selection techniques and classifiers evaluated in the dataset. Conversely,

Feature selector	classifier	MSE	RMSE	R2	Training Time (S)	Testing Time (S)	No. of feature selected
<b>F-Reg</b>	<i>SVR</i>	0.00228	0.04776	0.99845	44.32969	0.43160	48
<b>F-Reg</b>	<i>KNN</i>	0.00051	0.02257	0.99997	10.21231	1.74670	43
<b>F-Reg</b>	<i>DT</i>	0.00053	0.02301	0.99964	4.14455	0.01562	47
<b>VAR</b>	<i>SVR</i>	0.00276	0.05257	0.99812	339.18407	1.80649	77
<b>VAR</b>	<i>KNN</i>	0.00066	0.02568	1.00000	11.64501	1.41669	21
<b>VAR</b>	<i>DT</i>	0.00185	0.04307	0.99874	1.84484	0.01504	23
<b>Lasso</b>	<i>SVR</i>	0.00242	0.04916	0.99835	59.40518	1.01620	20
<b>Lasso</b>	<i>KNN</i>	0.00001	0.00259	0.99999	32.81606	1.53322	20
<b>Lasso</b>	<i>DT</i>	0.00010	0.00978	0.99993	22.05807	0.00316	17
<b>MI</b>	<i>SVR</i>	4.45857	2.11153	0.80887	496.21278	28.09595	4
<b>MI</b>	<i>KNN</i>	0.00506	0.07113	0.89999	108.01582	0.15672	13
<b>MI</b>	<i>DT</i>	3.94625	1.98652	0.86159	108.77875	0.00156	11
<b>Ridge</b>	<i>SVR</i>	2.89817	1.70240	0.82595	949.53478	33.74993	40
<b>Ridge</b>	<i>KNN</i>	0.00340	0.05832	0.90000	4.26084	1.49275	40
<b>Ridge</b>	<i>DT</i>	2.94355	1.71568	0.80975	1.91896	0.00806	39
<b>RFE</b>	<i>DT</i>	0.00461	0.06792	0.89997	654.01146	0.01005	30
<b>SFE</b>	<i>DT</i>	0.02827	0.16814	0.89981	397.95250	0.01523	4

Table 3.8: Performance of Feature Selection methods.

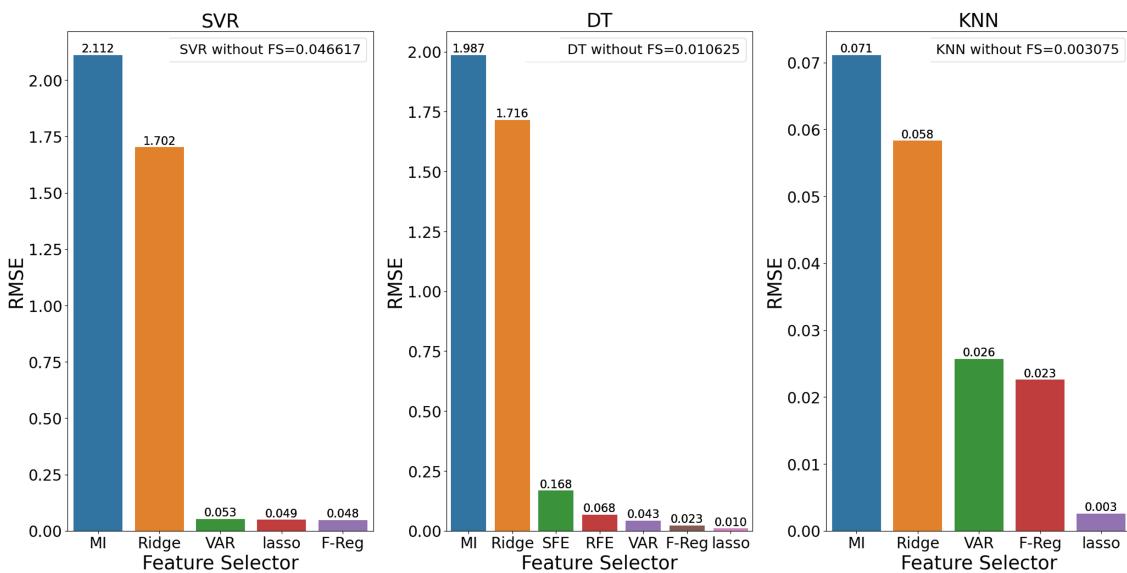


Figure 3.12: Comparison of RMSE of feature selection algorithms

the combination of MI with SVR has the highest RMSE value (2.11153), indicating that the model predictions using this combination might be less accurate.

Moving on to the  $R^2$  values from the analysis:

- The combination of Lasso with KNN achieved the highest R2 value (0.99999), demonstrating an exceptional fit of the model to the data.
- Several other combinations also display high R2 values, signifying their strong predictive capabilities. These combinations include F-Reg with DT, VAR with KNN, Lasso with DT, and Lasso with SVR.
- Conversely, combinations involving MI with SVR, MI with KNN, MI with DT, Ridge with SVR, Ridge with KNN, and Ridge with DT exhibit lower R2 values, indicating that these models might not explain the data as effectively as the top-performing combinations.

Regarding the training and testing times provided for each combination of feature selection techniques and classifiers:

- Training and testing times exhibit significant variation across different combinations.
- Combinations with SVR as the classifier (e.g., F-Reg with SVR, VAR with SVR, Lasso with SVR, MI with SVR, and Ridge with SVR) generally have longer training times compared to other classifiers.
- KNN as the classifier (e.g., F-Reg with KNN, VAR with KNN, Lasso with KNN, and MI with KNN) demonstrate moderate training times, but they tend to have longer testing times.
- DT as the classifier (e.g., F-Reg with DT, VAR with DT, Lasso with DT, MI with DT, Ridge with DT, RFE with DT, and SFE with DT) generally exhibits shorter training and testing times compared to SVR and KNN.

These observations highlight the trade-offs between different combinations in terms of predictive performance and computational efficiency, providing insights into the strengths and limitations of various feature selection techniques and classifiers. Figures Fig. 3.12 present visual representations of RMSE (Root Mean Squared Error) based on the data provided in the above Table 3.8.

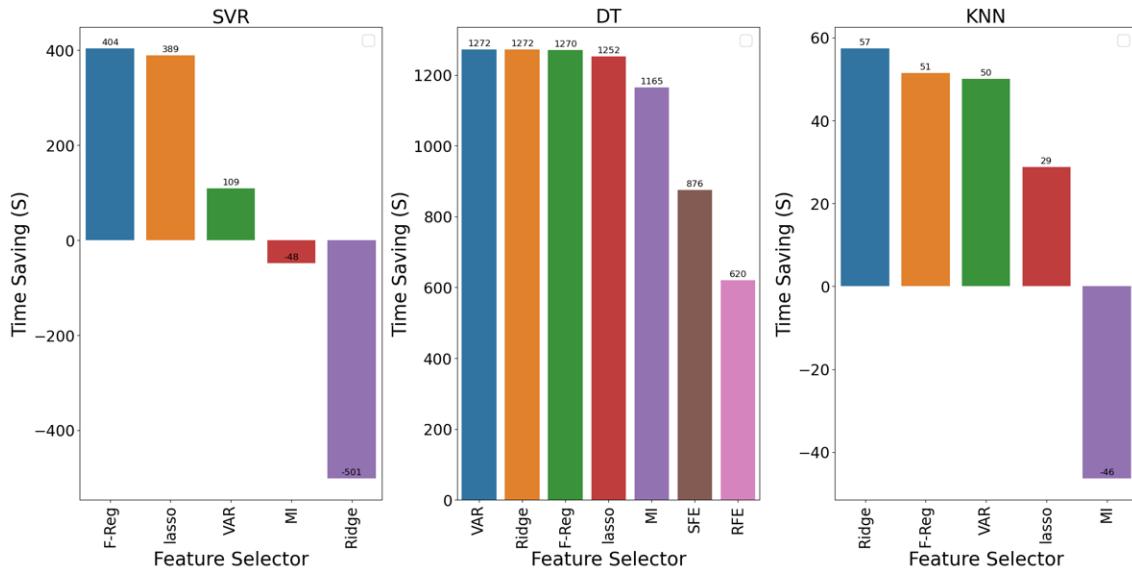


Figure 3.13: Time consumption by classifiers

Fig. 3.13 shows the change in computational training time when the proposed classifiers are used with feature selection algorithms. Positive bars represent the time saved when classifiers are used with feature selection methods, while negative bars indicate that specific feature selection methods have taken longer computational time than the classifier itself. As observed in Fig. 3.13, the proposed feature selection algorithms have achieved better computational efficiency compared to other feature selection methods. This is attributed to the fact that VAR and F-Reg belong to the filter feature selection method, known for their computational efficiency. In contrast, RFE and FFE fall under the wrapper class and necessitate a longer computational time compared to filter methods.

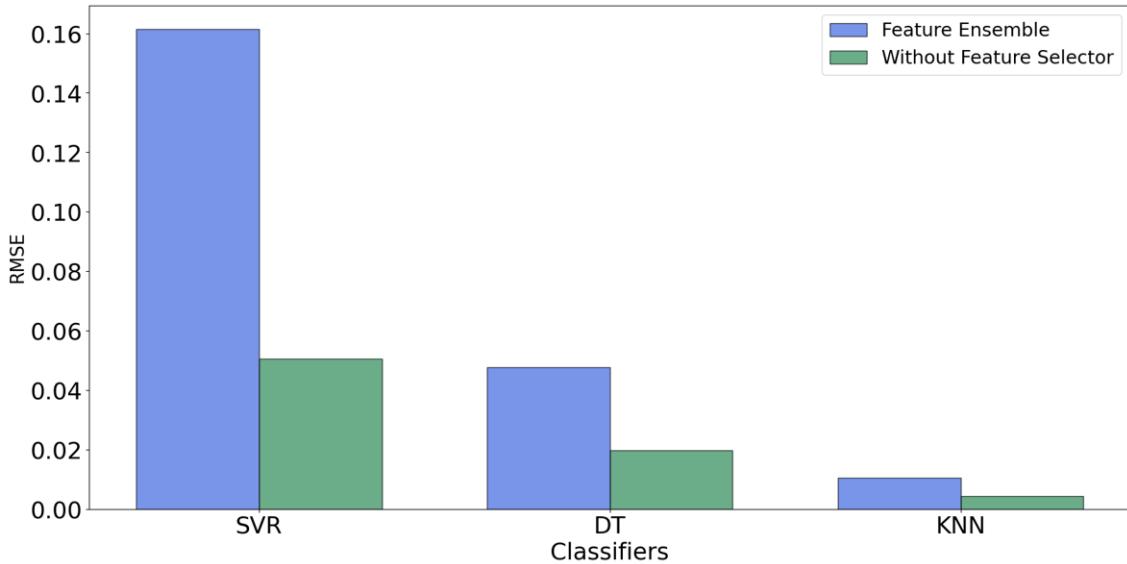


Figure 3.14: Comparison of root mean squared error for the classifier performance

Classifier	MSE	RMSE	R2	Training time (S)	Testing time (S)	No.of Features
<b>Ensemble-SVR</b>	0.026033	0.161347	0.982264	501.11298	35.94650	29
<b>Ensemble-DT</b>	0.002270	0.047645	0.999845	110.72919	0.05010	20
<b>Ensemble-KNN</b>	0.000110	0.010488	0.999993	56.78106	4.18316	17

Table 3.9: Performance of feature ensemble

### 3.6.3 Phase-3: performance of the feature-ensemble

This section describes and compares the performance of ML classifiers with and without feature selection algorithm. Fig. 3.14 shows the RMSE results for each classifier used in the proposed scheme, It can be noticed that without feature selection the error is low as the features. However considerable amount of time can be saved when feature selection algorithms are employed as can be seen from Fig. 3.15.

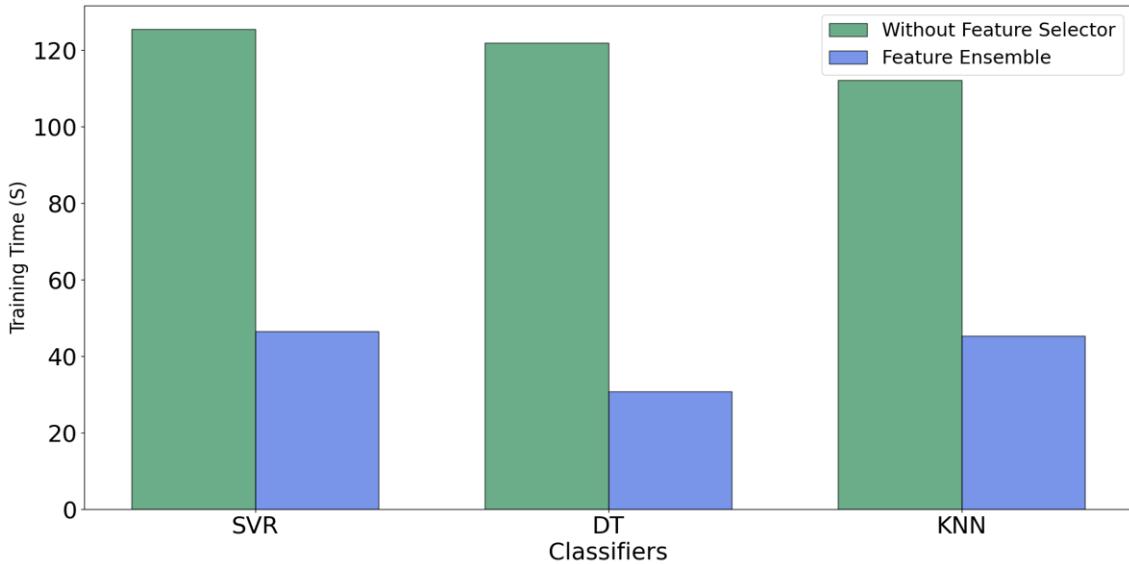


Figure 3.15: Training time with and without feature selection

### 3.6.4 Phase-4: Robustness of the feature ensemble against noisy data

This section describes the effect of noise on the proposed ensemble. The first and foremost consideration for the implementation of the algorithm for a real system would be measurement noise. To check the proficiency of the proposed algorithm under noisy measurements, simulations have been completed with noise. The ensemble is retrained with white noise, which is added after the simulation of the P-V curve. Simulations showed that the proposed method works well within the measurement noise limit for the PMU device. The performance of each of the learners was evaluated at three noise levels 80dB, 60dB and 40dB. We compute error ratios to capture the degree to which algorithms outperform each other in noise outcomes.

Comparing the results of different classifiers used in feature ensemble, we can see that all three models perform better in the absence of noise. When the noise level is increased to 60dB, the performance is not much affected. When the noise level is further increased to 40db, DT performs significantly worse than the other base learner because it tends to place a lot of weight on the noisy examples.

Fig. 3.16, Fig. 3.17, and Fig. 3.18 show the scatter plots of Actual vs Predicted which tells us how well the model performs under the influence of noise. For the Ideal model, the points should be closer to a diagonal line. From the plots, we can see that at noise level of 40dB, KNN and SVR perform similarly, while the data points in DT are more widespread from the diagonal line.  $R^2$  has a closer relationship to the actual vs predicted plots. If the model has a higher  $R^2$

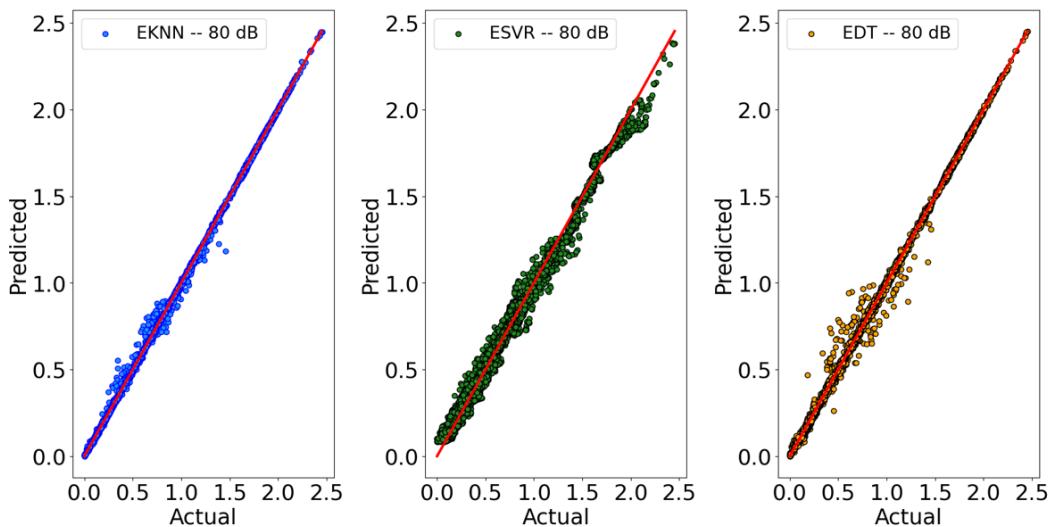


Figure 3.16: Noise with 80dB

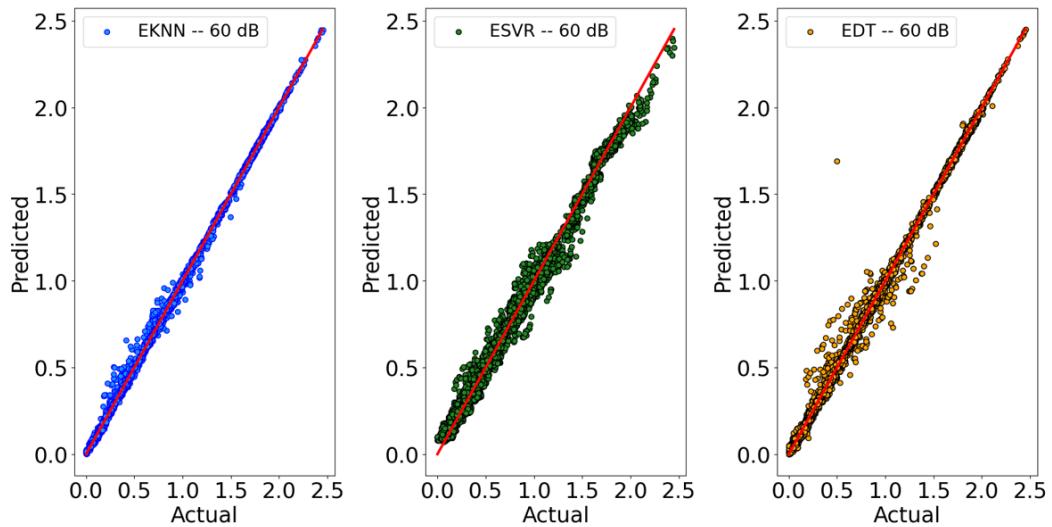


Figure 3.17: Noise with 60dB

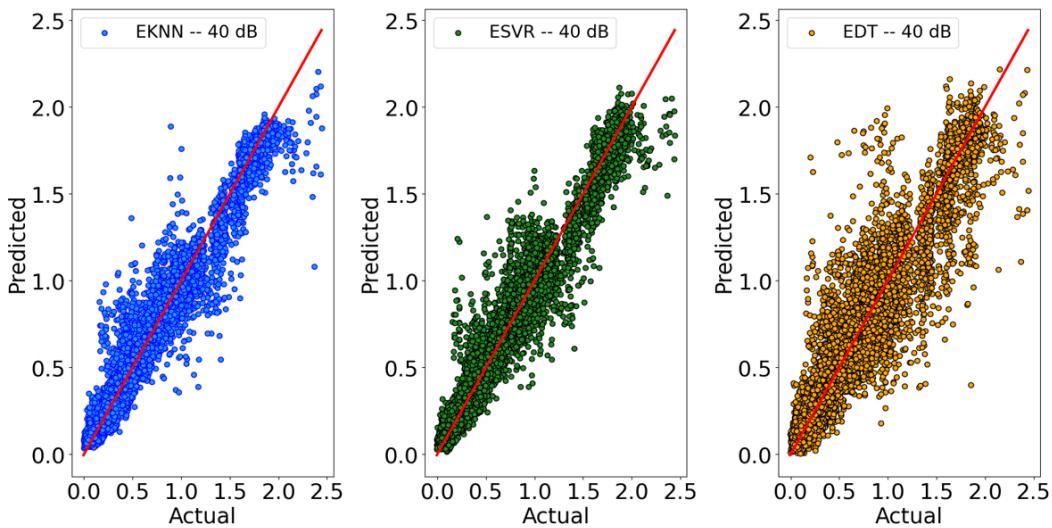


Figure 3.18: Noise with 40dB

Classifier	Noise Level	MSE	RMSE	$R^2$	Training Time (S)	Testing Time (S)	No. of Features
<b>Ensemble-SVR</b>	80	0.0024	0.0488	0.9919	87.6468	3.1846	31
<b>Ensemble-KNN</b>	80	0.0002	0.0126	0.9995	74.7639	7.9444	32
<b>Ensemble-DT</b>	80	0.0003	0.0182	0.9989	40.7827	0.0204	31
<b>Ensemble-SVR</b>	60	0.0029	0.0536	0.9901	119.2287	5.1489	34
<b>Ensemble-KNN</b>	60	0.0002	0.0153	0.9992	74.8683	7.9336	34
<b>Ensemble-DT</b>	60	0.0008	0.0284	0.9972	44.281	0.0204	34
<b>Ensemble-SVR</b>	40	0.0159	0.1262	0.9453	661.8249	27.298	40
<b>Ensemble-KNN</b>	40	0.0139	0.1181	0.9521	77.48	9.1753	40
<b>Ensemble-DT</b>	40	0.0347	0.1862	0.8809	46.3534	0.0184	40

Table 3.10: Performance of feature ensemble under the influence of noise

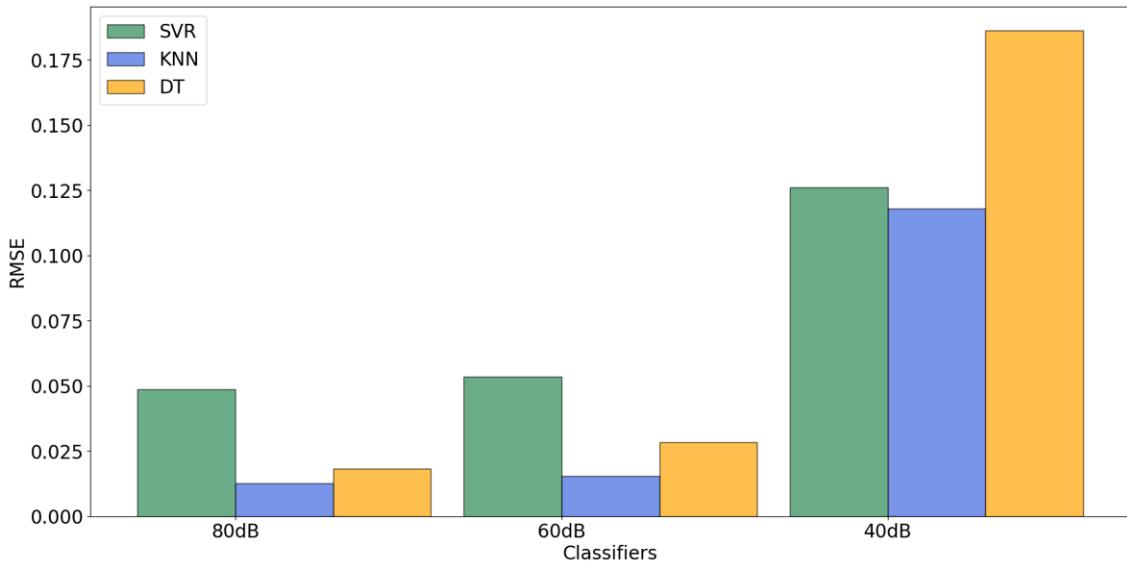


Figure 3.19: Root Mean Squared Error (RMSE) after introducing noise to the data.

value, all the points would be closer to the diagonal line. However, lower  $R^2$  indicates the data points are far away from this diagonal line. We can see from table 3.10 that the  $R^2$  score for DT when the noise level is 40db is 0.8809, which is lowest among all the three classifiers.

Fig. 3.19 provides the comparative analysis of the model accuracy based on RMSE for three levels of noise. It can be seen that the RMSE increases with increase with noise level. However, despite the RMSE is high at noise level of 40dB it is still significantly low in general which proves robustness of the proposed feature ensemble against the noise.

For commercially-available PMUs, SNR is at least 100 (total vector error is less than 1%) [22]. This means the magnitude of the noise is always less than 1% of the measured signal magnitude. But, it was found that the algorithm can predict stability even with noise to SNR 40.

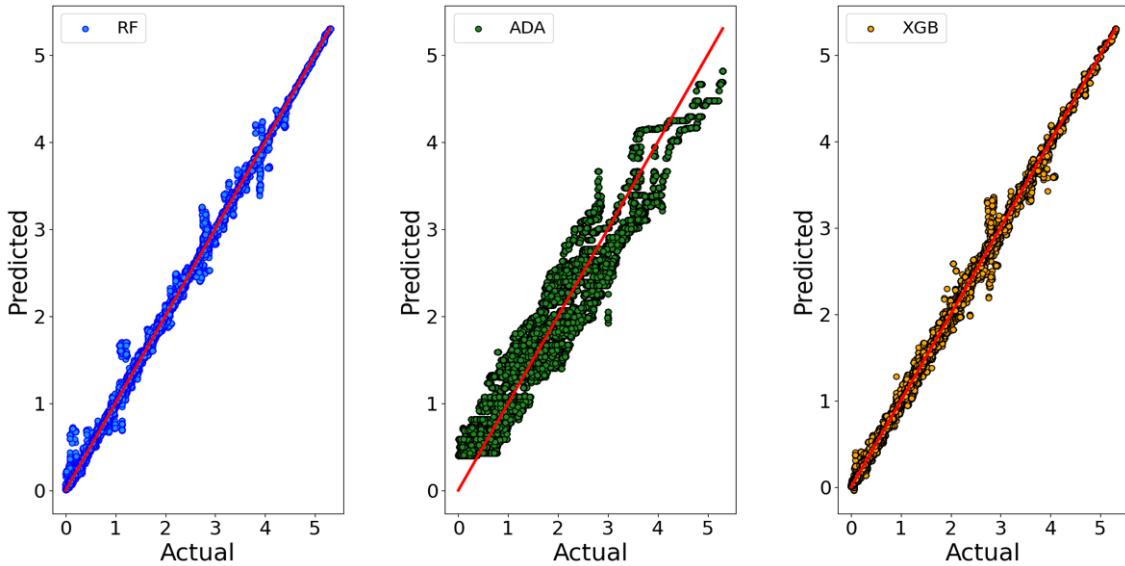


Figure 3.20: Scatter Plot: Actual vs Predicted for Random Forest, AdaBoost, and XGBoost Models.

Model	MSE	RMSE	R2	Training Time (S)	Testing Time (S)
<i>Ensemble_KNN</i>	0.00011	0.010488	0.999993	56.78106	4.18316
<i>Ensemble_DT</i>	0.00227	0.047645	0.999845	110.7292	0.0501
<i>Random_Forest</i>	0.005127	0.071603	0.994533	676.5657	1.799916
<i>XGBoost</i>	0.005411	0.073559	0.99432	566.3191	0.047421
<i>Ensemble_SVR</i>	0.026033	0.161347	0.982264	501.113	35.9465
<i>AdaBoost</i>	0.098608	0.314019	0.932921	300.1995	1.000281

Table 3.11: Comparison of Feature Ensemble with Random Forest, Adaboost, and XGBoost

### 3.6.5 Comparative Analysis: Feature Ensemble Method vs. State-of-the-Art Machine Learning Techniques

This section provides an in-depth comparison between the feature ensemble method and other prominent machine learning techniques. The detailed analysis is presented in Table 3.11, where the performance of the feature ensemble is juxtaposed against Random Forest, AdaBoost, and XGBoost.

Considering the provided metrics, the Ensemble-KNN and Ensemble-DT models outshine other ensemble approaches in terms of MSE, RMSE, and R2 values. These two ensembles consistently deliver accurate predictions and robust fits to the data. Notably, Ensemble-KNN achieves the lowest MSE (0.00011), RMSE (0.010488), and highest R2 (0.999993), positioning it as an exceptional performer. Fig. 3.21 further illustrates the comparative performance of Random Forest, AdaBoost, XGBoost, and the feature ensemble.

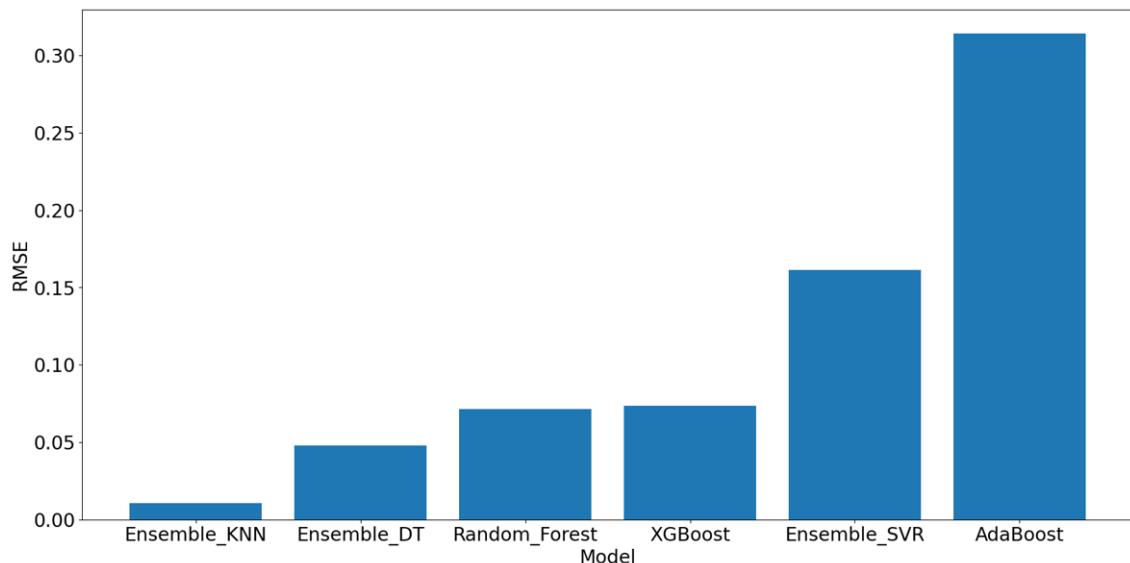


Figure 3.21: Comparison of the RMSE.

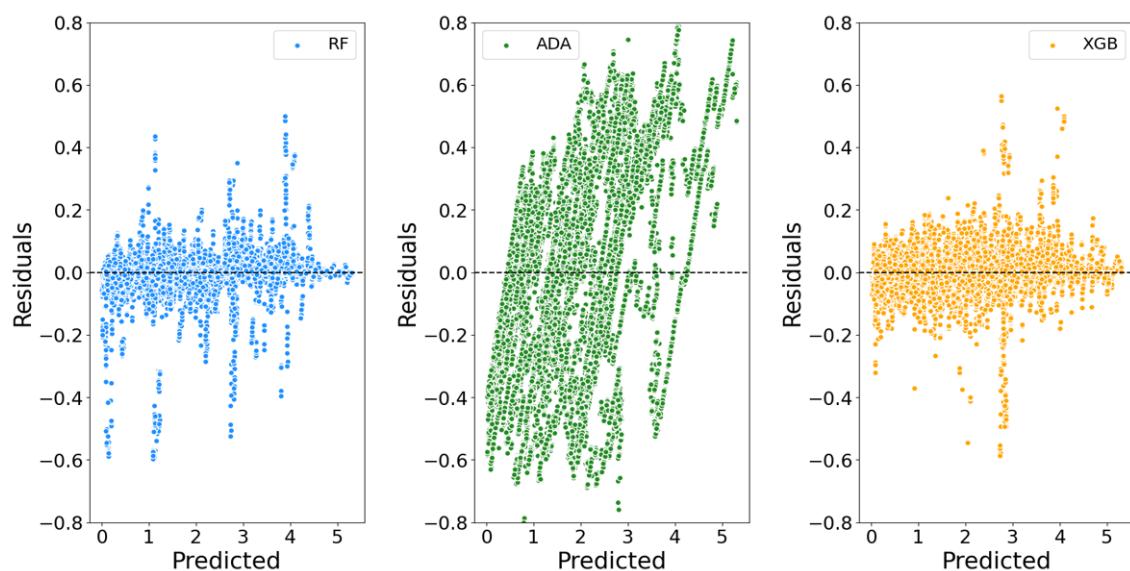


Figure 3.22: Residual Plot: Actual vs Predicted for Random Forest, AdaBoost, and XGBoost Models.

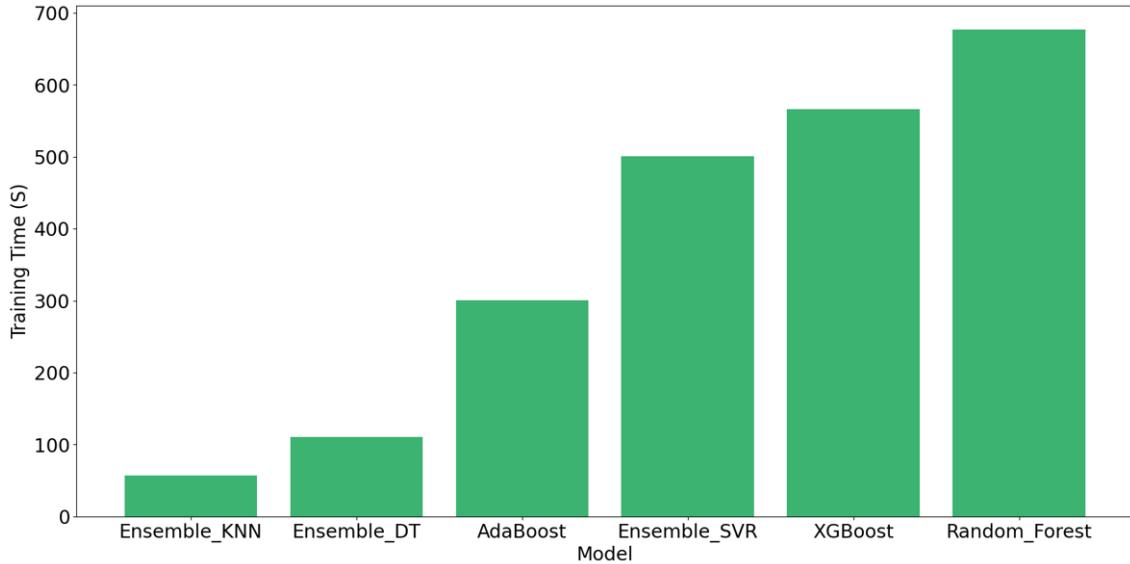


Figure 3.23: Comparison of the training time.

The Random Forest and XGBoost models exhibit commendable performance, although they have slightly higher MSE and RMSE values compared to the top-performing ensembles. Conversely, the Ensemble-SVR and AdaBoost models manifest higher MSE and RMSE values, indicating comparatively less accurate predictions and weaker fits the data.

Fig. 3.20 and Fig. 3.22 depicts scatter plots for Random Forest, AdaBoost, and XGBoost. Notably, the scatter plot for AdaBoost displays a wide dispersion, while those of Random Forest and XGBoost tend to align more closely with the diagonal. Their respective residual plots showcase the extent of the spread.

When considering data requirement, the ensemble methods generally benefit from larger datasets due to their complex nature. Computation time varies across models, with Ensemble-KNN and Ensemble-DT being more computationally efficient, while Random Forest and XGBoost may require more resources.

Ultimately, the choice of model depends on the specific requirements of the application. If high accuracy and training efficiency are essential, Ensemble-KNN and Ensemble-DT are strong contenders. For versatile options with reasonable accuracy, Random Forest, XGBoost, and Ensemble-SVR may be suitable. It's important to strike a balance between accuracy, efficiency, and computational resources when selecting the most appropriate model for a given task.

It's important to note that Ensemble-KNN and Ensemble-DT boast the quickest training times, while Random Forest and XGBoost exhibit longer training durations due to their ensemble

nature. Ensemble-SVR and AdaBoost fall in between these extremes. The selection of an appropriate model should consider the trade-off between training time and performance metrics such as MSE, RMSE, and R2. Fig. 3.23 provides a visual comparison of the training times across all methods.

## Chapter 4

# Online application of Machine Learning Model

Chapters 2 and 3 involve the creation of a voltage stability database during the offline stage, achieved by conducting simulations across various operating points and fault scenarios. The database is then subject to an ensemble feature selection process that eliminates irrelevant features, leaving behind only the relevant ones. Subsequently, a machine learning (ML) model is trained on the database, and the voltage stability margins (VSMs) estimated by the model are combined to generate the final prediction. This chapter discusses the practical implementation of the trained ML model in real-time for VSM estimation. The ML model is designed to quickly assess long-term voltage stability, and its findings are employed to facilitate control decision-making. Additionally, in the online stage, the stability database can be updated with current system information, such as network configuration and load composition, to enhance the accuracy of the existing database. Importantly, this online update ensures that the model is capable of adapting to unexpected system changes. Fig. 4.1 provides a conceptual overview of the proposed approach for real-time long-term voltage stability assessment.

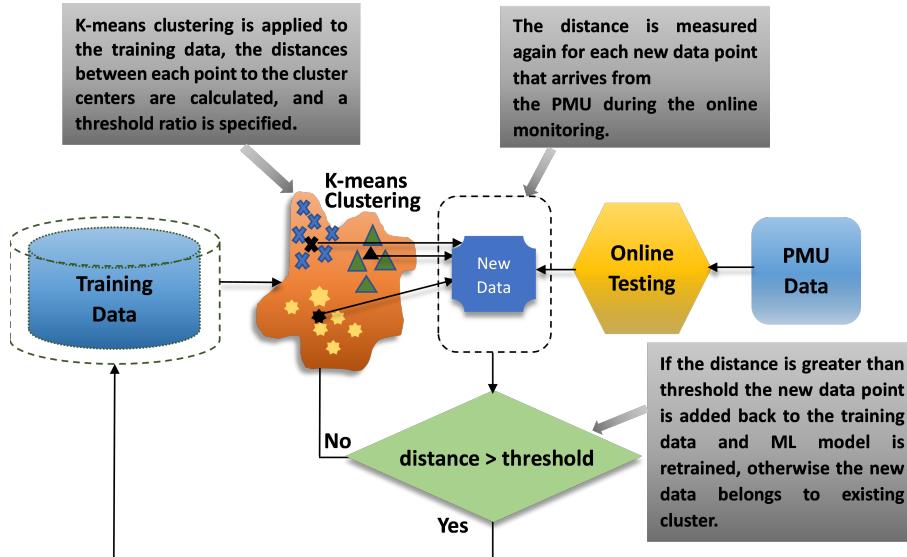


Figure 4.1: Online VSM evaluation scheme

The Scheme comprised of the following steps:

1. K-means clustering is applied to the training dataset. The objective is to divide the clusters into groups so that members within each group share comparatively more similarities than those from other groups.
2. Once we get the data inputs from the phasor measurement unit (PMU), the trained feature ensemble is applied, and VSM is estimated. At the same time, the distance of the input data to the cluster centroids is calculated.
3. The estimated VSM and distance of data points are compared against the threshold values.

The threshold values are the 99% confidence interval calculated for the following values:

- VSM
  - Root mean squared error (RMSE) and
  - Centroid Distance
4. If the estimated VSM is below the threshold value, the operator should be informed to take necessary action. Else no action is required, and the algorithm continues to estimate. However, it is important to acknowledge that the initial database used for training the algorithm offline may not accurately reflect the current operating conditions and dynamics of the power system. System dynamics in the context of power systems refer to the variations and changes that occur over time due to factors such as load fluctuations, network

re-configurations, and equipment failures. To account for these dynamic changes, the algorithm needs to be adaptive and capable of adjusting to the evolving system conditions. This adaptability is crucial for maintaining reliable and accurate predictions of the VSM. As new situations arise, such as changes in the network structure or load composition, the algorithm should be able to learn from and adapt to these scenarios. In this study, we propose a method to address these system dynamics and contingencies in the power system. We compare the RMSE and distance of the newly arrived data points to the threshold values to detect online drifts or significant deviations. We update the database used by the algorithm to incorporate the new observations and ensure it can adapt to the evolving system dynamics effectively.

## 4.1 K-means Clustering

The k-means is an unsupervised learning algorithm that solves the well-known clustering problem. Let the training data set  $\mathbf{D}$  be  $(x_i)_{i=1}^n$ , where  $x_i$  is the input attribute vector consists of voltage magnitude and voltage phase angle for  $i$ -th operating point of the  $n$  samples in a  $d$ -dimensional space. Given the number of desired clusters  $k$ , the goal of k-means clustering is to partition the dataset into  $k$  groups or clusters, denoted as  $C = C_1, C_2, \dots, C_k$ . The cluster means are initialized by randomly generating  $k$  points in the data space. Each iteration comprised of two steps a) cluster assignment and b) centroid update. In the cluster assignment step, each observation  $x_i$  is assigned to the cluster  $C_i$  with the closest center to that observation. In the centroid update step, the centers are redefined using the observations in each cluster: then the column means  $\mu_i$  are used to define the centroid of all points in the cluster, that is

$$\mu_i = \frac{1}{n_i} \sum_{x_j \in C_i} \mathbf{x}_j \quad (4.1)$$

To evaluate its quality or goodness of the algorithm, the sum of squared errors (SSE) scoring function is used which is defined as:

$$SSE = \sum_{i=1}^k \sum_{\mathbf{x}_j \in C_i} \|\mathbf{x}_j - \mu_i\|^2 \quad (4.2)$$

The aim is to find the clustering that minimizes the sum of squared errors scoring function:

$$i^* = \arg \min_{i=1}^k \{\|\mathbf{x}_j + \mu_j\|^2\} \quad (4.3)$$

The ArgMin function is commonly employed to identify the minimum values under certain constraints. To achieve a fixed point or local minimum, the k-means algorithm performs iterative cluster assignment and centroid update steps. The process of updating the centroids involves computing the mean of the data points in each cluster and potentially transferring data points to different clusters. This assignment and centroid update process is repeated until convergence is achieved, which is characterized by either no points changing clusters or unchanged centroids. Typically, the Euclidean distance is utilized to compute the distance between the data points and centroids in this algorithm. The k-means clustering algorithm's pseudo-code is presented in the following Algorithm.

---

**Algorithm 2:** The k-means clustering algorithm

---

**Require:**  $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  // Set of  $n$  data points  
 $k$  // number of desired cluster.

**Ensure:** a set of  $k$  clusters

**Steps:**

1. Arbitrarily choose  $k$  data points from  $D$  as initial centroids
2. **Repeat:**

    Assign each point  $x_i$  to the cluster which has the closest centroid;

    Calculate the new mean for each cluster;

**Until:**

    Convergence criteria is met.

---

#### 4.1.1 Database Update

During the online assessment, it is pretty standard that the distribution of the collected data changes over time due to the dynamic nature of the underlying phenomena, e.g., maintenance of equipment, topological changes, adjustments in configurations, environmental factors, etc. Consequently, when estimating VSM in real-time, new data belonging to unseen class labels, i.e., labels that the ML model did not know at the time of training, are likely to be present and may result in the wrong estimation.

Updating the prediction model frequently by appending the new data to the existing training set is computationally exorbitant. It may require the intervention of domain experts to absorb

the changes for VSM prediction. Due to these reasons, it is often impractical to retrain the ML model frequently. To this aim, we propose a self-evaluation step to automatically identify the change in the data distribution and degradation of the prediction quality over time. The main challenge in this approach is the absence of ground-truth predictions for the newly arrived PMU samples. Therefore we base our solution on detecting the changes in the underlying distribution of the data by applying a k-means clustering algorithm. The process is accomplished by segregating the training data  $\mathbf{D}$  into  $k$  clusters and calculating the cluster centroids. During the online phase, each subsequently received value trigger the application of the k-means algorithm. The distances between each newly arrived PMU data point to the cluster centroids, obtained from the training data  $\mathbf{D}$ , are calculated. The distance values are compared against the specified threshold ratio. If the distance is within the confidence limit, the latest data belongs to one of the existing clusters, and the prediction accuracy is not much affected. On the other hand, if the distance of new data is above the threshold, the operating point is added back to the dataset, and the training process is restarted offline.

It is important to acknowledge that in the online mode, the actions are constrained, and the system's ability to handle unforeseen or significant changes may be limited. The primary focus of the online phase is to make predictions based on the available information and adapt the model as best as possible without complete retraining. While the online mode may not allow for extensive actions or model modifications, it still serves a crucial purpose of providing real-time predictions and adapting to the data within the limitations. The self-evaluation step using clustering helps identify when the model's predictions may be affected by changes in the underlying data distribution and prompts the offline retraining process.

## 4.2 Case Study

The feature ensemble is applied online to monitor and inspect new situations during the online evaluation stage. Four time-domain simulations of online VSM estimations are presented in Fig. 4.2. The plot shows the result of three feature ensemble models by substituting the base learners with K-nearest-neighbor (KNN), classification and regression tree (CART), and support vector regressor (SVR). The simulations were carried out without increasing the load factor.

In the first simulation, the system initially operated under normal conditions. The estimated VSM remains constant, around 2.45 pu to 2.825 pu for all three ensemble models. Next, fault-1

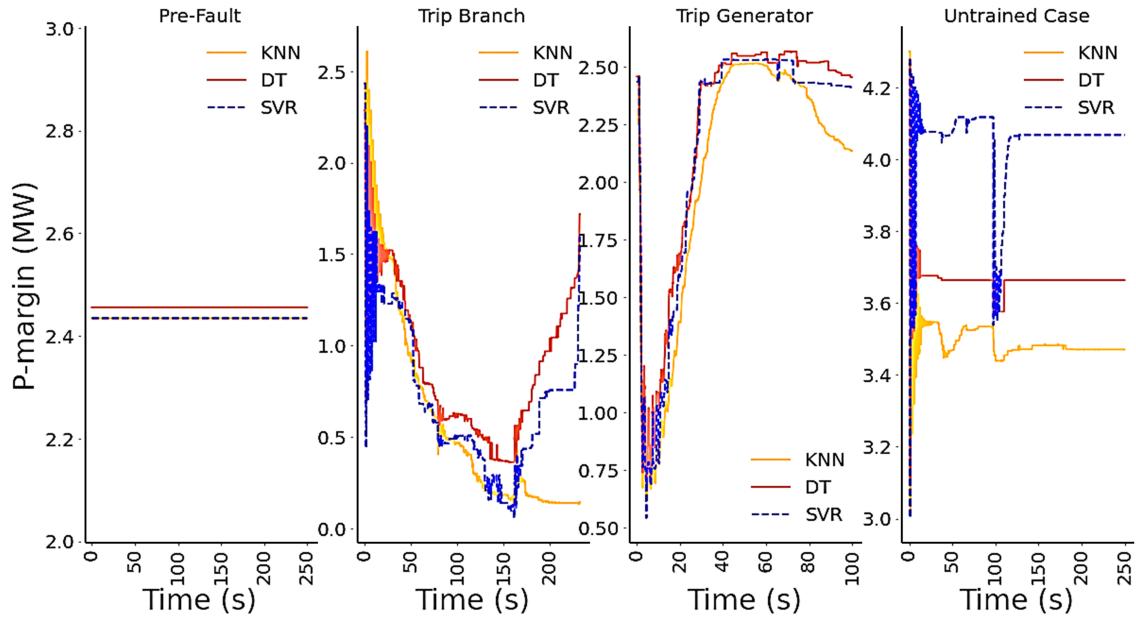


Figure 4.2: Online monitoring

is simulated. For this case, all three feature ensemble models show a declining trend for the estimated VSM. A warning will be given when the VSM crosses a threshold set by the system operator. In the subsequent simulation, fault-2 is applied. We can notice a significant dip in the estimated VSM that recovers after 40 seconds. In the last case, a new contingency is introduced by tripping branch 4012-4022 close to bus 4031 at  $t = 1s$ . The feature ensemble model was not trained previously on this fault condition. The database update phase is comprised of applying the k-means clustering algorithm and runs parallel with the online assessment to detect the changes in the distance of the new data points to the cluster centroids.

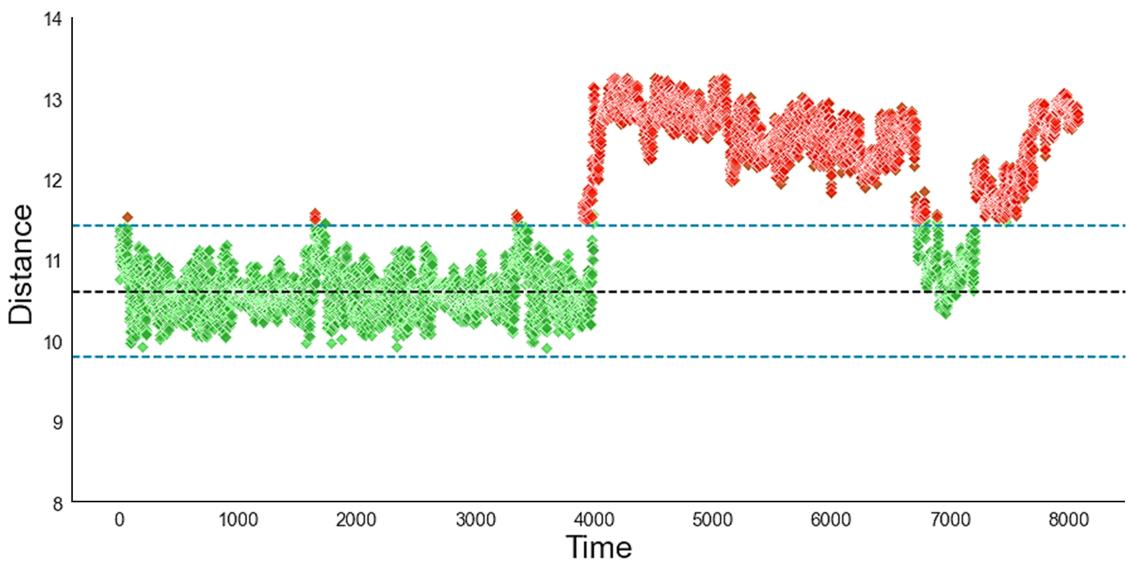


Figure 4.3: change detection using k-means clustering

Fig. 4.3 depicts this scenario. The green dots in Fig. 4.3 represents the distance of data points from the cluster centroids on which the ML model has been trained. At time = 4000 s, new contingencies are introduced. The distribution of the data changes, and the distance of the unknown sample data points. The newly measured data point distance is measured from the existing cluster centroids. The average distance and the confidence interval of each cluster member from the cluster centroid are used as a benchmark. If the distance of new data falls between the two confidence intervals, the point is considered similar to the existing clusters. Otherwise, the operating point is put back in the database for retraining. The existing dataset is updated, and the newest incremental samples are injected into the initial dataset. Subsequently, the updated datasets would, in turn, serve as the initial database for the next update period.

#### 4.2.1 Identifying Secure and Insecure Operating Points

For a Nordic test system, generally two base operating points are considered.

- Operating point A, which is very insecure , i.e. the system cannot stand some N-1 contingencies' several single contingencies cause instability or even some transient angle may result in instability cases.
- Operating point B : which is secure i.e. the system can stand a 5-cycle (0.1 s) fault on any line, cleared by tripping the line . The system can stand the outage of any single generator.

The 700 operating points analyzed for long-term voltage stability used variations of operating point B. This choice enables a focused assessment of how the system's voltage stability responds to different contingencies, starting from a stable initial state. This approach comprehensively evaluates the system's ability to manage disturbances and maintain stable operation over time.

Operating point A is highly insecure, vulnerable to N-1 contingencies causing instability and transient angle issues. Conversely, operating point B offers greater stability. Generating P-V curves using operating point B, rather than A, is aimed at evaluating the system's long-term voltage stability under realistic and stable conditions. Starting with a more stable operating point like B is essential in stability analysis, as it yields meaningful insights into the system's behavior and response to contingencies, whereas using an insecure point like A could obscure results due to inherent instability and transients.

The following contingencies have been considered:

- a 5-cycle (0.1 s) fault on the following lines
  1. 4032-4044
  2. 4031-4041
  3. 4041-4061
  4. 4045-4062
  5. 4043-4046
  6. 4032-4042

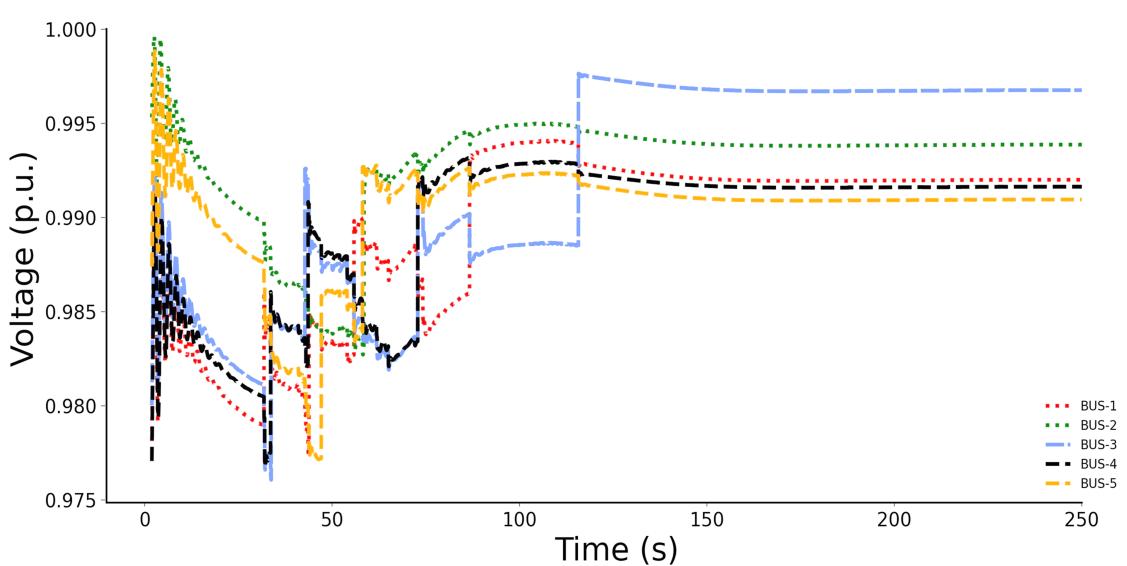


Figure 4.4: operating point B, fault on line 4032-4044 cleared by opening line: distribution voltages

The fault is cleared by tripping the lines.

- the outage of following single generators:
  1. g6
  2. g7
  3. g16
  4. g14

#### Criteria used in long-term dynamic simulation

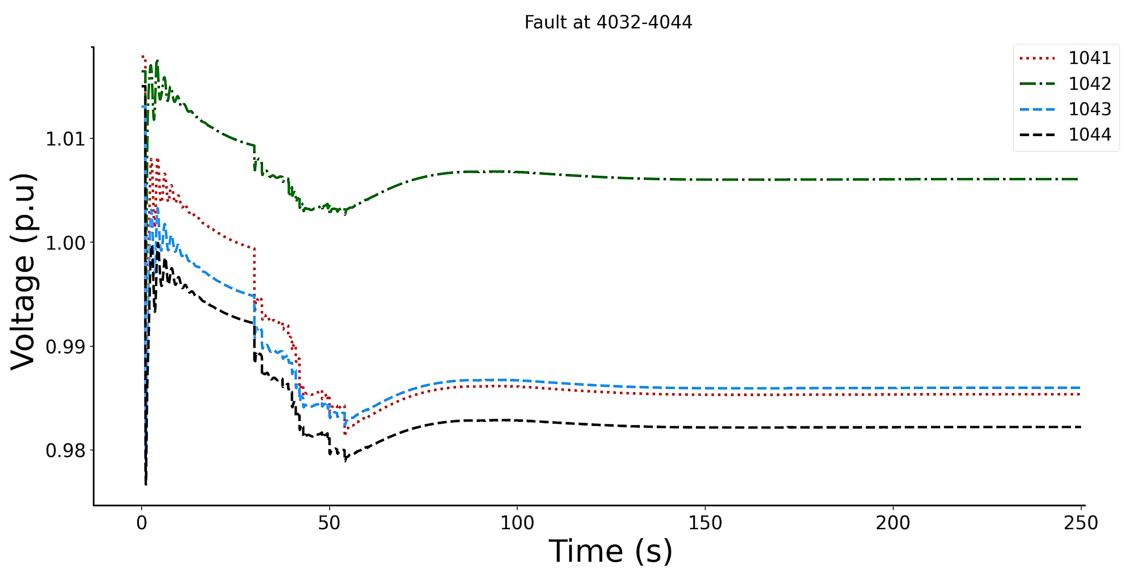


Figure 4.5: operating point B, fault on line 4032-4044 cleared by opening line: distribution voltages

The post-contingency evolution is considered acceptable if, over a simulation interval of 600 seconds:

1. All distribution voltages are restored in their dead bands ([0.99 1.01] pu).
2. No generator voltage falls below 0.85 pu; and
3. No loss of synchronism takes place.

### Branch tripping

Fig. 4.4 and Fig. 4.5 shows the evolution of distribution and transmission system voltages due to the three-phase solid fault on line 4032-4044, near bus 4032, lasting 5 cycles (0.1 s). The fault is cleared by opening the line, which remains opened. The stable evolution of the voltage at buses 1, 2, 3 and 4 can be noticed in Fig. 4.4. Fig. 4.5, relative to the same disturbance, shows four distribution voltages 1041, 1042, 1043 and 1044 that are successfully restored in their dead-bands by load tap changers (LTCs).

### Generator

A dynamic model of the power system can be integrated over time to determine if there will be a stable state or a collapse. Fig. 4.6(a) depicts the time evolution of voltages at the 130-kV buses of the Central region for three generators g-6, g-7 and g-16 respectively. The corresponding distribution voltages are shown in Fig. 4.7. For generator g-6 it is clear that the impact is

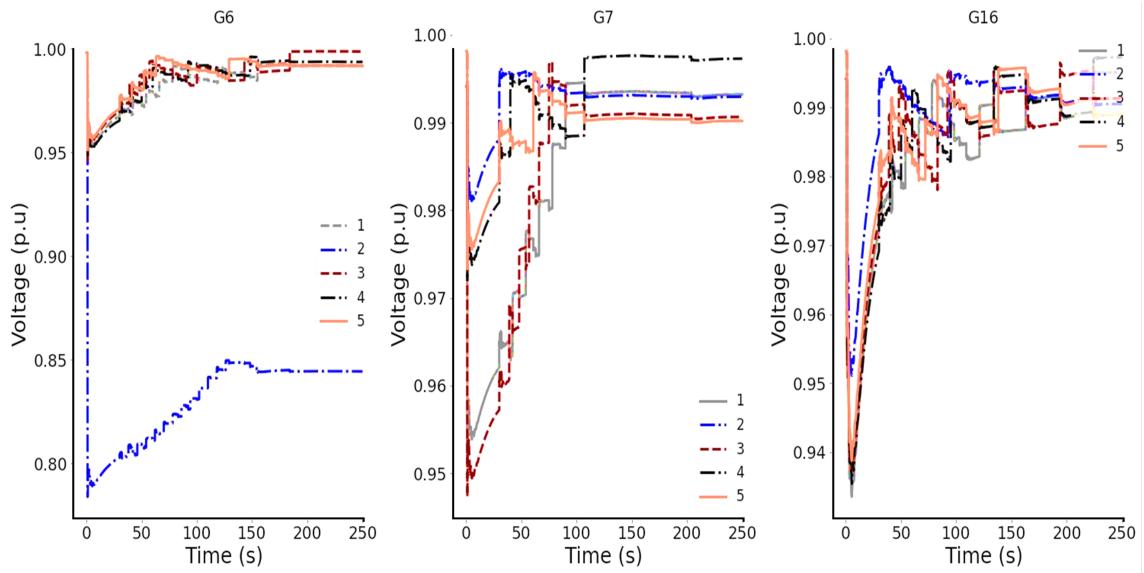


Figure 4.6: Generator contingencies

limited to bus 1042. It can be seen that the LTCs succeed in restoring the distribution voltages in their dead bands, except for the one controlling bus 2 (Fig. 4.7(a)). After some unsuccessful steps, the ratio of transformer 1042-2 hits its lower limit, which explains the pseudo-stabilization of the system, though with unacceptably low voltages at buses 1042 and 2. For generator g7 and generator g16, it can be noticed that the voltage drops after the initial disturbance but gets restored after 100 s and 50 s respectively.

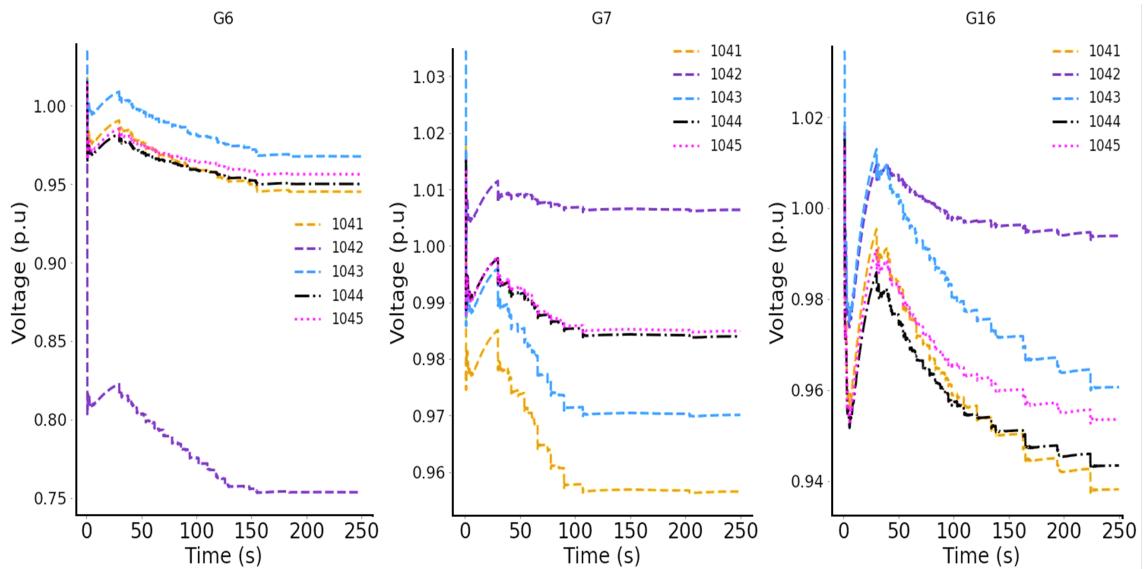


Figure 4.7: Online monitoring

Fig. 4.8 shows the VSM for the above-mentioned generator buses. For comparison, the VSM is plotted for the three ML algorithms i.e. KNN, SVR and CART. It can be noticed that for

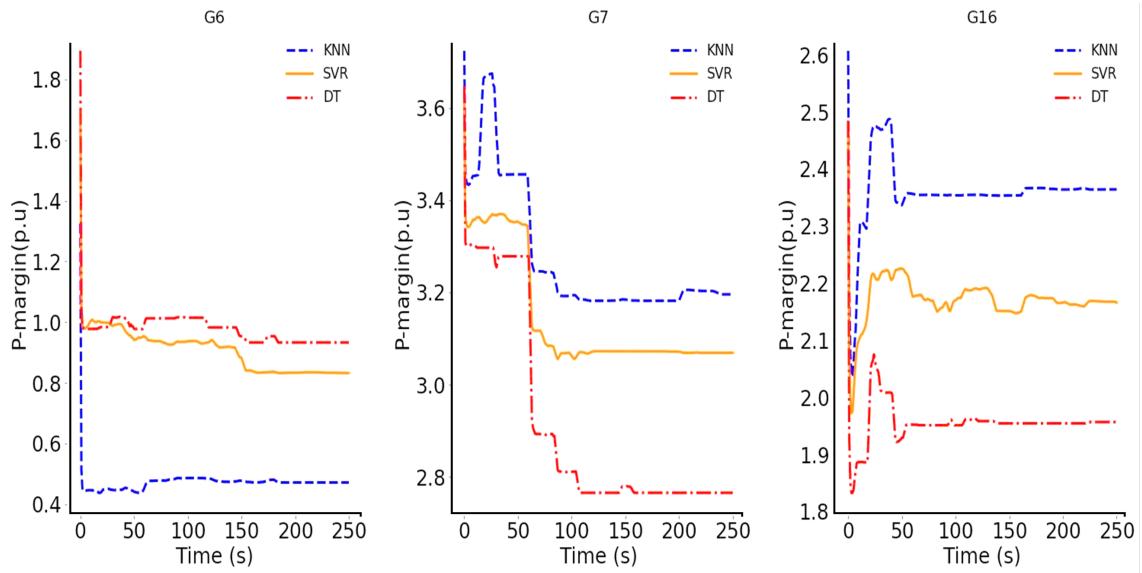


Figure 4.8: Online monitoring

generator g6 there is very little stability margin left. Simulation results given in Fig. 4.8 show that all three proposed ensemble feature selection can estimate the stability margins, however, the performance of KNN shows much more sensitivity to the voltage fluctuations. As can be observed, the behaviour is quite similar to the distribution buses, indicating that with the outage of generator g6, Central Area demand cannot be restored due to the weakened system and corresponding reduced power transfer capability after the generator outage. Also, Fig. 4.8 shows that following the increase in load at generator g7 and g16, the voltage varies suddenly which causes the algorithm to detect the initial disturbance occurrence by a drop in the VSM, which gets stabilised after  $t=110s$  for g7 and  $t=50 s$  for g16.

## Chapter 5

# Power system oscillations

In this chapter, an introduction is given to the electromechanical oscillation phenomena that occur in power systems with low frequencies. The definition of low-frequency oscillation is provided, and its classification is explained in detail. Some examples of power system blackout incidents caused by low-frequency oscillation are also provided. Recent research and developments in monitoring and estimating modal parameters for inter-area oscillations are discussed. Furthermore, the chapter emphasizes the importance of continued research in power system stability by motivating further investigation related to inter-area oscillation.

This chapter serves as a precursor to subsequent chapters, introducing the concept of inter-area oscillations and underscoring their pivotal role. These oscillations share similarities with voltage stability, exhibiting gradual changes over prolonged periods. Left unattended, these oscillations can accumulate, resulting in heightened amplitudes and potential instability. Much like the analysis of voltage stability, utilizing machine learning (ML) models for inter-area oscillations presents an effective approach to capturing the dynamic behavior of the power system. By examining historical oscillation patterns, system parameters, and operational conditions, these models can learn from data and forecast the occurrence and magnitude of inter-area oscillations. Consequently, they contribute to the formulation of efficient damping control strategies and provide guidance to operators in proactively mitigating these oscillations. Although machine learning models necessitate offline training to unveil underlying patterns, they can be swiftly employed in real-time to analyze current measurements and make prompt decisions for damping control.

## 5.1 Power System Electromechanical Oscillations

Elecromechanical oscillations are inherently present in large interconnected power systems. Electromechanical oscillations can occur at different frequencies but are not harmful if they decay rapidly. The parameters of such oscillations (e.g. damping, magnitudes, and frequencies) depend on the system parameters and the contingencies that cause them. These parameters provide crucial information about the modes of the power systems [40] and help determine power system stability in real-time. For a large stability margin, the damping ratio of all the system modes must be greater than some value, typically 3–5% [136]. An event such as tripping of generator or a branch can result in a declining damping ratio of a system mode, suggesting a system shift towards a less stable ratio, as observed in the 1996 western grid outage[137]. Thus, continuous monitoring of these modes can provide critical information on the system stability margin and can help prevent system outage by allowing a timely action.

### 5.1.1 Classification of Power System Oscillations

Oscillations were observed in the power system as soon as synchronous generators were interconnected to provide more power capacity and reliability. Originally interconnected generators were close to each other, and oscillations were in a frequency range of 1 to 2 Hz [138]. As more utilities interconnected, oscillations in different frequency ranges were observed. Nowadays, oscillations can be classified by specific types, with each type exhibiting a certain degree of oscillation frequency when the phenomenon occurs [139]. In general, oscillations can be grouped into two broad classes [140]:

- Forced Oscillations and
- Natural Oscillations

Forced oscillations refer to oscillations that are driven or forced by an external source or input [141]. Forced oscillations can occur due to various factors, such as disturbances, faults, or control actions. These external influences can introduce periodic or cyclic variations in the system's parameters, leading to oscillatory behavior. These oscillations occur in the range of 0.2 Hz to 2 Hz [142].

Natural oscillation refers to the inherent oscillatory behaviour of a system without any external forcing or input. Natural oscillations are commonly observed as the system responds to distur-

bances or changes in operating conditions [143]. These oscillations arise due to the dynamics and characteristics of the system's components, such as inertia, stiffness, and damping. The natural oscillations are further classified as: 1) ambient response, and 2) transient response. Fig.5.1 provides information about the classification of power system oscillations [144]. The transient response characterized by the electromechanical oscillation modes can be sub-classified as 1) Ultra-low frequency oscillations, 2) Low-frequency oscillations, and 3) sub/super-synchronous oscillations/resonances.

Here the ultra-low frequencies have a range of +0.1 Hz to -0.1 Hz. It occurs due to the negative damping of generators.

Out of these sub-classifications, Low frequency oscillations can be further categorized as:

- Global frequency oscillations,: is a low-frequency mode of 0.01–0.2 Hz (approx. 0.06– 1.2 rad/s) in which all generating units move in unison.
- Inter area mode: occurs when a group of generators in one area swing against group of generators located in another area of a wide area power system. The frequency range of inter area oscillation mode is in the range of 0.1 to 0.7 Hz.
- Local plant mode oscillations: occurs when one or more synchronous generators, in a specific power station, swing together against the whole large power system. This type of oscillation occurs within a frequency range of 0.7 to 3 Hz.
- Local machine system oscillations: is an oscillatory electro-mechanical mode and is usually associated with the rotors of synchronous generating units in a station swinging against the rest of the power system or against electrically-close generating stations. The range of its frequency is normally from 6–12 rad/s (1– 2 Hz).

Sub-synchronous oscillation (SSO) are characterized by equivalent oscillations between different components of the power system, such as turbine-generators, series capacitors, power electronic converters, and HVDC controllers.

The causes of SSO are varied and complex. [145],[146],[144] mentioned several causes, including:

1. Induction generator/machine effect, torque amplification, and torsional interaction between rotating components and a series compensated grid.
2. Control-device-dependent SSO, which occurs when steam-/hydro-turbines interact with

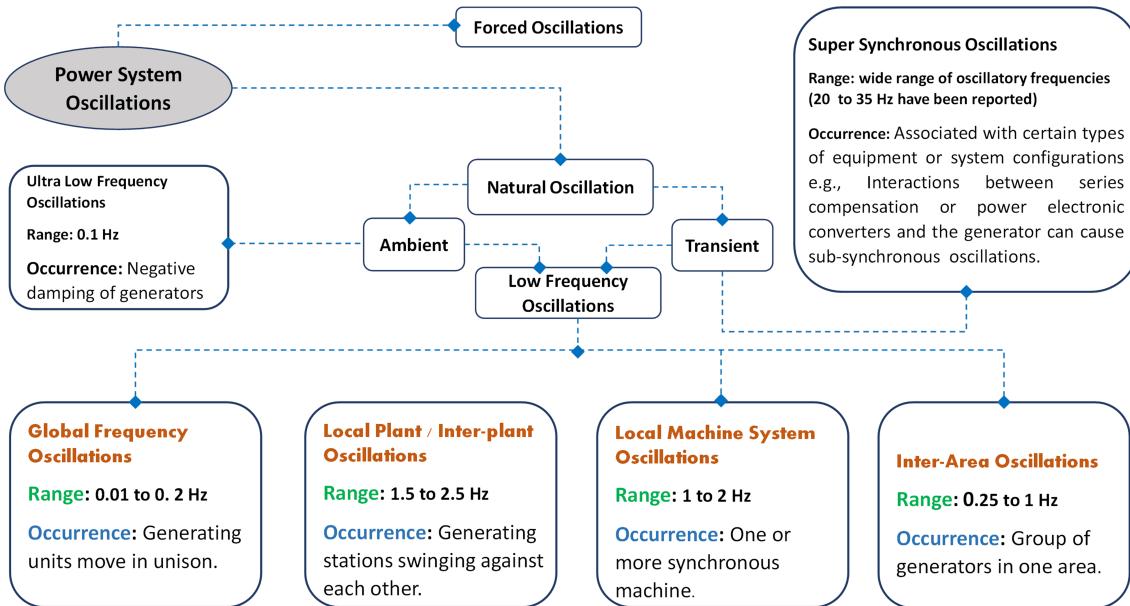


Figure 5.1: Classifications of power system oscillations

fast response controllers, resulting in sub-synchronous torsional interaction.

3. Sub-synchronous control interaction among power electronic converters and series compensated grids.

These causes reflect the impact of different equipment and control systems on the stability of the power grid. SSO can occur within a wide range of oscillatory frequencies, unlike the low-frequency oscillations that have a narrower frequency range. In the Hami power system, wide-area measurement systems (WAMs) have detected SSOs with frequencies ranging from 20 Hz to 35 Hz since June 2014 [145].

## 5.2 Estimation of power system modes

Estimation of power system modes can be performed through two groups of methods: model-based methods and measurement-based methods. Model-based methods linearize the governing equations of the system about an operating point, and measurement-based methods fit a linear model to the system measurements.

### 5.2.1 Model based method

There are many different methods for mode damping estimation and most of them require a mathematical model of the system. A state variable is one of the set of variables that are

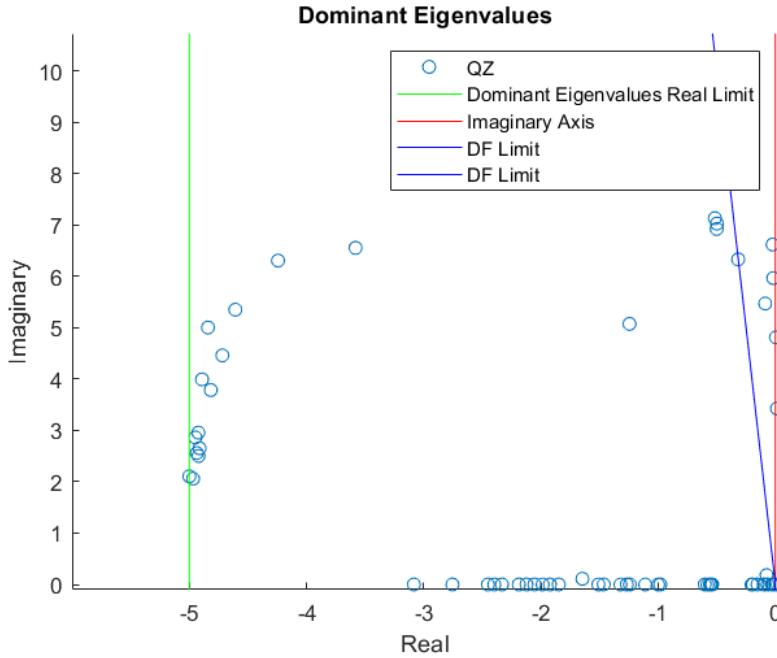


Figure 5.2: Plot of the eigenvalues of the Nordic test system. The system is small-signal stable as most of the eigenvalues have negative real parts, but it is poorly damped due to many of its eigenvalues being outside the 10% damping line shown in blue.

used to describe the mathematical 'state' of a dynamic system. Intuitively, the state of a system describes enough about the state to determine its future behaviour in the absence of any external forces affecting the system. The system represented by a state space model is described by the following equation

$$x_{k+1} = Ax_k + Bu_k + w_k \quad (5.1)$$

$$y_k = Cx_k + Du_k + v_k \quad (5.2)$$

where:  $x$  is the state vector;  $u$  is the inputs vector;  $y$  is the outputs vector;  $w$  and  $v$  is the random vectors; and  $A$ ,  $B$ ,  $C$ , and  $D$  are the state space matrices. Whereas  $k$  represents the discrete time interval.

For the purpose of damping estimation, only the  $A$  matrix is needed. The primary task in modal identification is to determine the system poles of the system transfer function or equivalently, the eigenvalues of  $A$ . The eigenvalues of the state matrix  $A$  have an important influence on the system response, and understanding them is key to understanding the general solution of

the state-space system. The eigenvalues are the mathematical roots of the system; they tell about the system's stability. The eigenvalues of the system matrix  $A$  (poles) can be plotted in a complex coordinate system. Furthermore, each mode's natural frequency and damping ratio can be defined using the real and imaginary components of the poles  $\lambda_i = \sigma_i \pm j\omega_i$  and the damping ratio of the corresponding  $i$ -th mode can be calculated as:

$$\zeta_i = \frac{-\sigma_i}{\sqrt{\sigma_i^2 + \omega_i^2}} \quad (5.3)$$

Fig. 5.2 shows the locations of the eigenvalues for a damped mode and an undamped mode. An eigenvalue with a negative real part results in damped oscillations, and an eigenvalue with a positive real part results in growing oscillations. In case of Fig. 5.2 the circles on red lines represents the modes with growing oscillations. If the solution contain oscillations, then the imaginary part of the eigenvalue defines the frequency of these oscillations.

Due to the large-scale nature of interconnected power system, modal analysis by utilizing mathematical models has become a tedious task. With wide area implementations of phasor measurement units (PMUs), most modern modal analysis techniques nowadays utilize measurements based modal analysis.

### 5.2.2 Measurement based method

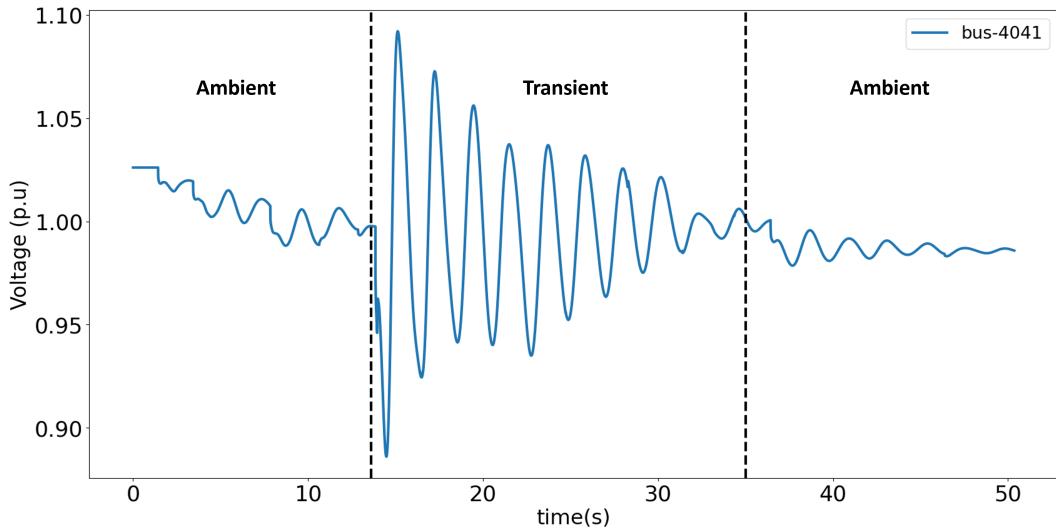


Figure 5.3: Ambient vs ringdown oscillation for the bus 4041 in Nordic test system [2]

The measurement data obtained from PMU can be classified into two types:

1. Ambient data
2. Transient /ringdown data

Fig. 5.3 depicts both types of data. An Ambient Response is an example of a Natural Response. Ambient data is obtained when a system is working under an equilibrium condition, and the major disturbance is from small-amplitude random load changes [147].

A Transient Response is the system's response immediately after a sudden disturbance, such as a fault, line tripping, generator trip, or load tripping. A Natural Response typically characterizes small-scale transient responses. This thesis addresses ringdown detection in the presence of noise using an Empirical Wavelet Transform (EWT). The motivation behind using EWT is to provide a flexible and adaptive time-frequency analysis tool that can effectively capture non-stationary and multi-scale features in signals. EWT allows for the decomposition of a signal into different frequency components based on the local characteristics of the signal, enabling enhanced analysis and understanding of complex and dynamic signals in various fields such as signal processing, image analysis, and pattern recognition.

In a ringdown approach the damping is estimated by fitting the transient (or ringdown) response  $y(t)$  to a simple function  $\hat{y}(t)$ . Suppose that a linear, time-invariant dynamic system is brought to an "initial" state  $x(t_0) = x_0$ , at time  $t_0$ . by means of some test input or disturbance. Then, if the input is removed and there are no subsequent inputs or disturbances to the system, it will "ring down" according to a differential equation of form:

$$\dot{x} = Ax \quad (5.4)$$

where  $x$  is the state of the system. Here,  $A$  is an  $(n \times n)$  matrix representing the system dynamics, and the dot above  $A$  represents the derivative with respect to time. This equation describes how the state of the system evolves over time during the ringdown phase when there are no subsequent inputs or disturbances to the system. Let  $\lambda_i, p_i, q_i$  be respectively the eigenvalues, right eigenvectors, and left eigenvectors of  $(n \times n)$  matrix  $A$ , then the solution to (5.4) can be expressed as [148]

$$x(t) = \sum_{i=1}^n (q_i^T x_o) p_i \exp(\lambda_i t) \quad (5.5)$$

$$= \sum_{i=1}^n R_i x_o \exp(\lambda_i t) \quad (5.6)$$

where  $R_i = p_i q_i^T$  is an  $(n \times n)$  residue matrix. Note that  $q_i^T x_o$  in (5.6) is a scalar (that is a simple constant). This implies that, though  $x_o$  determines the stimulus to the mode associated with eigenvalue  $\lambda_i$ , the distribution of modal response among the components of  $x$  is entirely determined by the corresponding right eigenvector  $p_i$ . Consequently information about  $p_i$  can be extracted by an appropriate modal decomposition of  $x(t)$ . For simplicity, suppose that there is just one output from the system and that it is of form

$$y(t) = \beta x(t) \quad (5.7)$$

Where  $\beta$  is a constant. Introducing  $y(t) = \beta x(t)$  as the output simplifies the estimation process because it provides a direct relationship between the observed record  $y(t)$  and the state of the system  $x(t)$  without the need for complex transformations or additional equations. The output from the system,  $y(t)$  is linearly related to the state of the system  $x(t)$  through a constant  $\beta$ . This linear relationship, allows us to directly estimate the parameters for the exponential terms in the observed record for  $y(t)$  by fitting a function  $\hat{y}(t)$  to the data. This simplification allows for more straightforward parameter estimation and facilitates the analysis and characterization of the system's transient response during the ringdown phase.

The ring down methods are generally designed to directly estimate the parameters for the exponential terms in (5.6) and/or (5.7) by fitting a function

$$\hat{y}(t) = \sum_{i=1}^Q A_i \exp(\sigma_i t) \cos(2\pi f_i t + \phi_i) \quad (5.8)$$

to an observed record for  $y(t)$ , the ring down methods aim to estimate the mode amplitude  $A_i$ , damping factor  $\sigma_i$ , oscillation frequency  $f_i$ , and phase angle  $\phi_i$  for each component in the system's response, whereas damping is calculated using (5.3). The  $Q$  represents the number of components or modes being considered in the model. It signifies the total number of exponential

terms being summed up to approximate the observed record  $y(t)$ .

### 5.3 Motivation for Damping Estimation

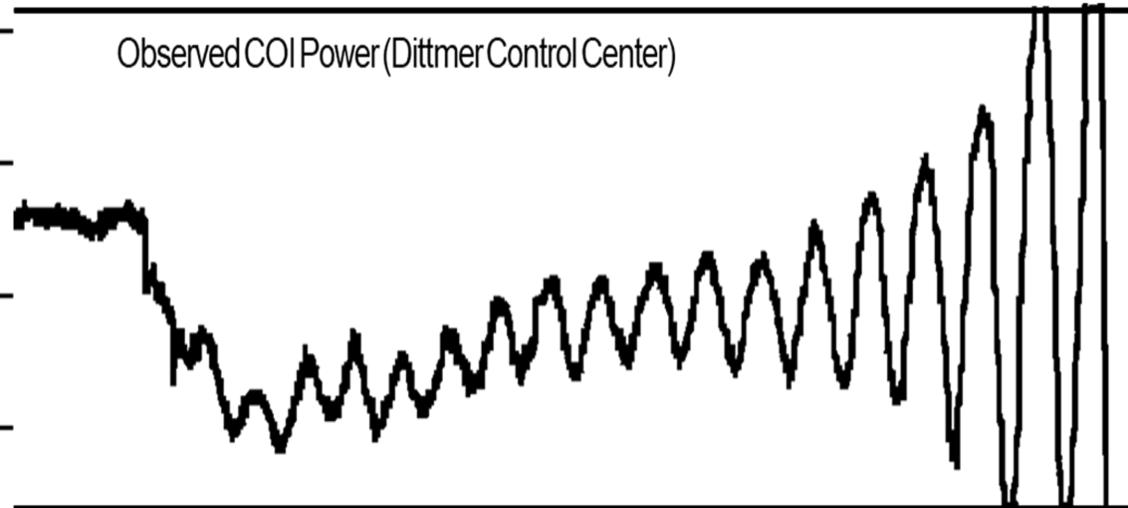


Figure 5.4: Growing oscillations at California Oregon border at Malin substation on August 10, 1996; *source [2]*

Fig. 5.4 shows power flow oscillations recorded at Malin substation by Bonneville Power Administration, leading to a break-up of the WECC system [149]. Usually, these power flow oscillations are damped by the system impedance, but if the interconnecting system is relatively weak or stressed, these oscillations can grow. These growing oscillations led to the August 10, 1996, blackout. There is no guarantee that another unforeseen problem cannot occur in the future, significantly when tie lines are continually upgraded, and power transfer is growing. These disturbances tend to become more severe and evolve into critical inter-area modes [150]. If not appropriately damped, such oscillations can result in increased losses and undue stress on the mechanical components of generators and, in extreme cases, may lead to instability [151]. Detecting these oscillations in a timely manner can prevent the system from major catastrophic events.

Due to time-dependent control actions and nonlinear dynamics, these oscillations are inherently non-stationary and may exhibit nonlinear trends and sudden variations[152]. Furthermore, complex observational data may contain many electromechanical modes of oscillation close in frequency and stochastic in nature. These non-stationary, nonlinear and stochastic characteristics of the data makes the analysis and interpretation of the system's behaviour difficult and

can lead to biased or incorrect results. While some existing techniques can still be useful in cases where linearity and stationarity assumptions hold, certain limitations exist when dealing with nonlinear and dynamic oscillations. Specifically, commonly used methods like Prony and matrix pencil algorithms for estimating oscillation modes have their drawbacks.

Prony and matrix pencil algorithms can be sensitive to noise in the data, compromising the accuracy of mode estimation, particularly in the presence of significant noise. Additionally, these algorithms assume a stationary signal model and may struggle to capture the time-varying nature of non-stationary oscillations accurately. Furthermore, as Prony and matrix pencil algorithms are based on linear signal models, they are less suitable for capturing the nonlinear dynamics exhibited by oscillations. If the oscillations demonstrate strong nonlinear behavior, such as amplitude modulation or frequency modulation, these algorithms may not provide accurate mode estimation. Lastly, when multiple oscillation modes are closely spaced in frequency, Prony and matrix pencil algorithms may face challenges in distinguishing between them, leading to mode aliasing. This can result in the inaccurate identification and tracking of individual modes.

With the advent of PMU, various measurement-based methods have been proposed for early diagnostics of inter-area oscillations. These methods can be loosely divided into adaptive and non-adaptive techniques. Adaptive methods have the ability to adapt and adjust their parameters based on the changing characteristics of the data. They can effectively handle non-stationary and time-varying signals, making them suitable for analysing complex oscillatory behaviour in power systems. On the other hand Non-adaptive signal processing techniques are commonly used in applications where the signals have known and consistent properties, and there is no need to dynamically adapt to changing conditions or learn from the input data. A discussion on the commonly used non-adaptive methods is first provided before focusing on the adaptive methods that the present study concerns.

Many modal analysis methods use pre-determined basis to process data and are therefore considered as non-adaptive or rigid [153]. Existing methods for inter-area oscillation analysis include Fourier transform [154], CWT [155, 156], Prony's method [157, 158, 159] Matrix Pencil method [160, 161], Kalman filter [162, 163], total least square [164] and singular value decomposition [165].

Most of the above methods start by first defining the basis function. Next, the signal is con-

volved with these basis function to obtain amplitude and frequency either for distribution or for filtering. Such an approach has the advantage of a solid mathematical foundation. Unfortunately, these methods are not adaptive at all [166]. Besides, most of these methods are based on a linear approximation of the system and suitable for analysing stationary oscillation signals only.

Another downside of these methods is the traditionalistic apriori basis, where the analysis is based on convolution of the data with the already established basis. Any approach with apriori basis could not fit well to the variety of the data from different underlying mechanism [166]. The term "traditionalistic apriori basis" refers to an established or predefined set of basis functions or components used in certain signal processing or analysis methods. For example, Wavelet analysis involves decomposing a signal into different frequency components using a set of wavelet functions. These functions can be chosen based on prior knowledge or assumptions about the signal's characteristics. The downside highlighted is that using a fixed, predetermined set of wavelet functions may not effectively capture the diverse and complex nature of different signals arising from various underlying mechanisms. Each signal might require a different set of wavelet functions to accurately represent its unique features. This limitation is further compounded by the convolution process and its interaction with the uncertainty principle, making the estimation of power system modal behavior more challenging.

To overcome these limitations, adaptive signal processing techniques may be more suitable. These techniques can better handle non-stationary and nonlinear behaviour and provide more robust and accurate estimation of oscillation modes in dynamic and nonlinear systems. Empirical mode decomposition (EMD) proposed in [167], is an adaptive signal processing method. Adaptive decomposition methods adapt to the transient feature and emphasise the local characteristics of the signals without requiring any prior basis to match the signal characteristics. They can adaptively extract the constituent oscillation modes of mono-components nature. At the same time, they can replicate the underlying properties from a random signal to represent it as a superposition of several mono components [168]. This permits accurate estimation of both the instantaneous frequency and instantaneous amplitude of each constituent component and provides an in-depth analysis of the time variability of signals.

Despite the above advantages, EMD suffers from the major drawback of mode mixing, instability under noise inferences, over/under fitting due to cubic spline interpolation and lack of

mathematical formulation. EMD suffers from a lack of a precise mathematical formulation primarily because it is an empirical or data-driven method. Unlike some other signal processing techniques that are based on well-defined mathematical principles and equations, EMD is derived from an iterative process that is guided by the intrinsic properties of the data itself. This empirical nature makes it challenging to establish a comprehensive mathematical framework that governs all aspects of EMD.

EMD is essentially a self-adaptive approach that aims to decompose a signal into components called Intrinsic Mode Functions (IMFs) that represent oscillatory patterns at different scales. These IMFs are obtained by identifying local extrema and constructing upper and lower envelopes using cubic spline interpolation. The process continues by subtracting the identified IMF from the original signal and repeating the decomposition on the residual signal until certain criteria are met. The lack of a formal mathematical formulation means that the behaviour of EMD and the properties of the resulting IMFs are not rigorously defined or proven. This can be seen as a drawback because it makes it challenging to precisely analyse the theoretical properties of EMD and establish mathematical guarantees regarding its performance and limitations.

Many scholars have carried out comprehensive researches on EMD and have developed related methods, such as Ensemble Empirical Mode Decomposition (EEMD) [169], local mean decomposition (LMD) [170] and more recently variational mode decomposition (VMD) for solving mode mixing and over/under fitting problems. Nevertheless, due to the characteristics of intricate signals encountered during the transient phenomenon in the power system, how to exploit the merits of these methods and effectively extract the meaningful features still deserves further investigation.

A recent addition to the adaptive mode decomposition family is EWT [171]. Unlike EMD, EWT works in the frequency domain and adapts according to the information contained in the analyzed signal [172]. The EWT process detects all the local maxima of the spectrum, then gets the boundaries which are the midpoint of two consecutive maxima to segment the Fourier spectrum. However, as mentioned in [153], it is a big challenge to employ the Fourier spectrum for determining the boundaries in noisy and non-stationary signals. False modes prevent proper segmentation of the spectrum. Another problem associated with the EWT boundary detection method is selecting the number of bounds in advance. In power systems there is a large number

of closely spaced low-frequency oscillations, thus it is difficult to guess the number of modes in a signal apriori. [173] proposed a solution to detect the boundaries of the Fourier spectrum using the sliding window approach; however, using just the sliding window results in unnecessary segmentation. This means that the entire spectrum is divided into smaller segments based solely on the window size, irrespective of the presence or significance of frequency peaks.

The decomposed EWT modes satisfy the requirements of mono-components signal and can be used with the HT based instantaneous attributes to get the time-frequency representation. However, HT also suffers from the energy leaks at the two ends, and the negative frequency will emerge when the signal maxima are lower than the minima. The sampling signal cycle must be ensured to be integral to avoid the influence of the end effect, which is not an easy option due to the randomness of the signal. Furthermore, the most common way of calculating the instantaneous damping ratio depends on first and second derivatives, which introduces discontinuities.

Motivated by these existing issues, this thesis proposes to automatically determine the boundaries of the Fourier segments by using a sliding window approach and then limit the number of segments by using a threshold value. With sliding windows, local maxima are determined automatically leading to a clear and concise separation of the Fourier spectrum. By thresholding the frequency amplitude, the number of modes can be controlled. To avoid the end effect in HT, in this paper instead of extending the end of the signal, we propose to discard the distorted portion of the instantaneous amplitude by segmentation, and the viability and effectiveness are verified by the case studies. An alternative damping estimation method by combining logarithmic decrements and instantaneous amplitude is also being proposed to improve the precision and accuracy of the damping ratio estimates.

The highlighted contributions of this work can be summarized as follows:

- A framework to automatically determine the boundaries of Fourier segments in EWT by detecting local maxima using a sliding window and appropriate threshold.
- Avoid the end effect in HT by discarding the distorted portion of the instantaneous amplitude through segmentation.
- A technique to improve the accuracy in instantaneous damping ratio estimation by applying logarithmic decrements on the instantaneous amplitude without introducing or

spreading discontinuities.

## 5.4 Interarea Oscillations

The second part of this thesis centers its attention on interarea oscillations, and a concise background is presented below to contextualize this focus.

All power systems have electromechanical modes. Systems with power transfer over long distances with long radial lines are the most likely to have inter-area electromechanical modes with lower damping. Several cases have been recorded in which interarea oscillations have resulted in blackouts of power systems [174, 175, 176].

Interarea oscillations appear as a result of supply and demand imbalance between the group of generators in two areas. These oscillations have frequencies in the range of 0.1 to 0.8 Hz [33]. Small disturbances such as load changes can induce interarea oscillations. The phenomenon can be visualized as two large generators trying to desynchronize each other in the event of supply and demand imbalance in each area [177]. As a result of desynchronization, a damping torque diminishes and the amplitude of rotor oscillation increases. If no remedial action is taken these oscillations can cause power system separation or major blackouts.

### 5.4.1 Modal Parameter Estimation of Interarea Oscillations

Modal analysis is a crucial technique employed in signal processing to extract essential information from measured oscillation data. It is aimed at identifying and characterizing the inherent oscillation modes present in a system. The fundamental parameters that define a mode of oscillation are as follows:

- **Modal Frequency:** This parameter signifies the inherent frequency at which a system or structure tends to oscillate when subjected to a disturbance or perturbation. It provides insights into the natural dynamic behavior of the system.
- **Modal Damping:** Modal damping reflects the rate at which the oscillations of a mode decay over time. It is a crucial indicator of the stability of the system. High damping implies rapid dissipation of energy and stable behavior, while low damping indicates persistent oscillations and potential instability.
- **Mode Shape:** The mode shape outlines the spatial distribution of the oscillations within

the system. It helps in understanding how different parts of the system contribute to the overall oscillatory behavior.

Modal parameter estimation is the process of deducing these essential parameters from the recorded oscillation waveform data [178]. Accurate estimation of modal parameters is of paramount importance for various reasons. Firstly, precise mode estimation can aid in stabilizing poorly damped modes, contributing to overall system stability. Secondly, real-time modal parameter estimation can provide situational awareness, enabling operators to make informed decisions and take timely corrective actions to maintain power system stability [179].

It's important to note that this thesis focuses specifically on two key aspects of modal analysis: modal frequency estimation and modal damping estimation. These parameters are integral for understanding the dynamic behavior and stability of power systems. The subsequent chapters will delve into a detailed methodology for effectively estimating these two crucial parameters, contributing to an enhanced understanding of interarea oscillations and their impact on power system stability.

# Chapter 6

## Estimation of low frequency electromechanical oscillation

### 6.1 Introduction

This section proposes a real-time automatic modal analysis of multi-dimensional PMU measurement data. The proposed algorithm aims to identify the natural frequency (NF) and damping ratio (DR) of the interarea oscillatory modes present in the power system's signal. In this study, an enhanced Empirical Wavelet Transform (EWT) approach based on a moving window segmentation of the Fourier spectrum is introduced for accurate time-frequency representation of noisy, non-stationary, and nonlinear signals. EWT is an adaptive signal decomposition method, but its main shortcoming is that Fourier segmentation strongly depends on the local maxima of the amplitudes of the Fourier spectrum. This shortcoming can be overcome by developing a feasible and efficient sliding window-based empirical wavelet transform (SEWT). The SEWT approach decomposes the low-frequency electromechanical oscillation signal into a series of intrinsic mode functions (IMFs). The Hilbert transform (HT) is applied to each IMF to obtain the instantaneous parameters of the signal. Then, the NF and DR of each mode are estimated as the average of the instantaneous parameters. The proposed method achieves perfect segmentation in noisy and non-stationary signals. Furthermore, simulated and experimental signals are used to verify the effectiveness of the proposed method.

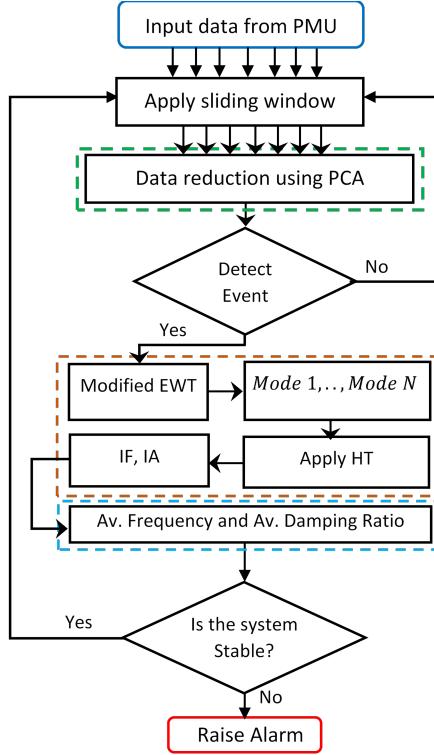


Figure 6.1: Flow diagram of the proposed methodology.

## 6.2 Proposed Scheme

Fig. 6.1 presents an overview of the proposed scheme for detecting oscillatory stability. Noisy PMU data is collected, and the moving window technique is applied to emulate the online detection method for the estimation of modal frequency. The moving window strategy assumes that recent data carries more significant implications than historical data. As new samples arrive, the old samples are discarded from the window, and the parameters of the models are updated iteratively. Due to a large amount of PMU streaming data, online analysis becomes computationally challenging. Therefore, principal component analysis (PCA) is used as a dimensionality reduction tool. PCA aims to find appropriate directions within data that maximize variance and sometimes reduce noise effects. As a result, the high dimensions of the data are reduced by using only those principal components that contribute to the covariance.

In the proposed approach, the oscillation levels are continuously monitored, thus making it possible to recognize the usual level of the ambient oscillations. Consequently, detecting events and other inconsistencies that deviate from the ambient electromechanical changes is possible. If the anomalies in the form of oscillations are detected, then SEWT is applied. The algorithm estimates the signal's oscillation frequencies and builds appropriate boundaries to create the

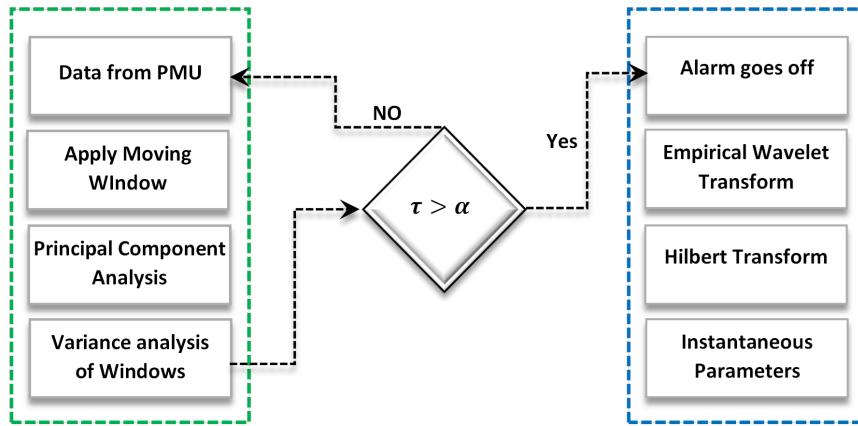


Figure 6.2: Detection of an event in the proposed methodology.

wavelet filter bank. The filter banks decompose the time-series signal into a set of frequency bands according to the estimated boundaries. This decomposition results in mono-component signals containing only a single mode known as IMFs. When the IMFs are accurately obtained, the signal's instantaneous frequencies (IF) and instantaneous amplitude (IA) are identified using HT, and the average of the instantaneous parameters is calculated. Finally, the instantaneous damping ratio (IDR) is estimated from the instantaneous parameters. In case of unstable or badly damped modes, an alarm is raised.

### 6.3 Theoretical Background

This section provides information about the theoretical background of the algorithms used in the proposed scheme. These include:

1. Event Detection
2. Principal Component Analysis
3. Empirical Wavelet Transform
4. Hilbert Transform

#### 6.3.1 Event Detection

A moving window technique is applied for the online detection of interarea oscillation. As shown in Fig. 6.2, the complete procedure is used only if a significant excitation (event) is identified. To detect an event, the signal's variance  $\sigma$  is measured every time the window moves and compared

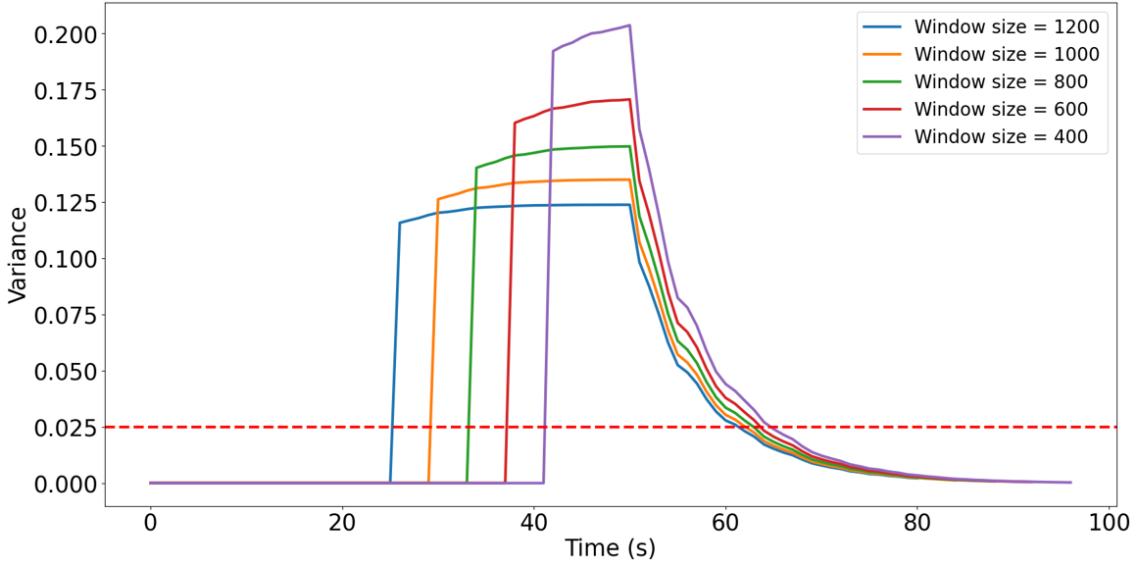


Figure 6.3: Variance analysis for different window sizes

with a threshold  $\tau$ . The complete identification method is carried out if  $\sigma$  is greater than the  $\tau$ . Fig. 6.3 shows the  $\sigma$  of a signal based on different window sizes. As can be noticed,  $\sigma$  is higher for smaller window sizes and lower for larger ones. The  $\tau$  has to be set up based on the size of the time window considered.

### 6.3.2 Principal Component Analysis

Due to the large number of operating variables and the presence of multiple modes in each variable, the estimation of NF and DR from voltage variables one at a time will lead to a heavy computational burden. Moreover, voltage PMU measurements from different buses are often strongly correlated as a result of redundant sensors and mechanism relationships [180]. Hence, it is necessary to carry out dimension reduction to eliminate redundant data information and capture the primary data structure. PCA can reduce the dimension effectively and maintain data information in the first few principal components. The description of PCA may be found in many sources[181][182]. PCA extracts the dominant pattern of the data and provides essential reliability for the detection of modal frequency components which could not be achieved at any given scale separately.

The voltage PMU data matrix to be analyzed by PCA is represented as  $\mathbf{X}_{(m \times s)}$ , consisting of  $s$  observations obtained from the online measurements of the voltage variables  $m$  ( $m \ll s$ ) measured over time  $t$ . The matrix  $\mathbf{X}$  can be decomposed via singular value decomposition (SVD) into  $\mathbf{X} = \mathbf{UDV}^T$ . Where  $\mathbf{U} \in R^{(s \times s)}$  and  $\mathbf{V} \in R^{(m \times m)}$  are unitary orthogonal matrices.

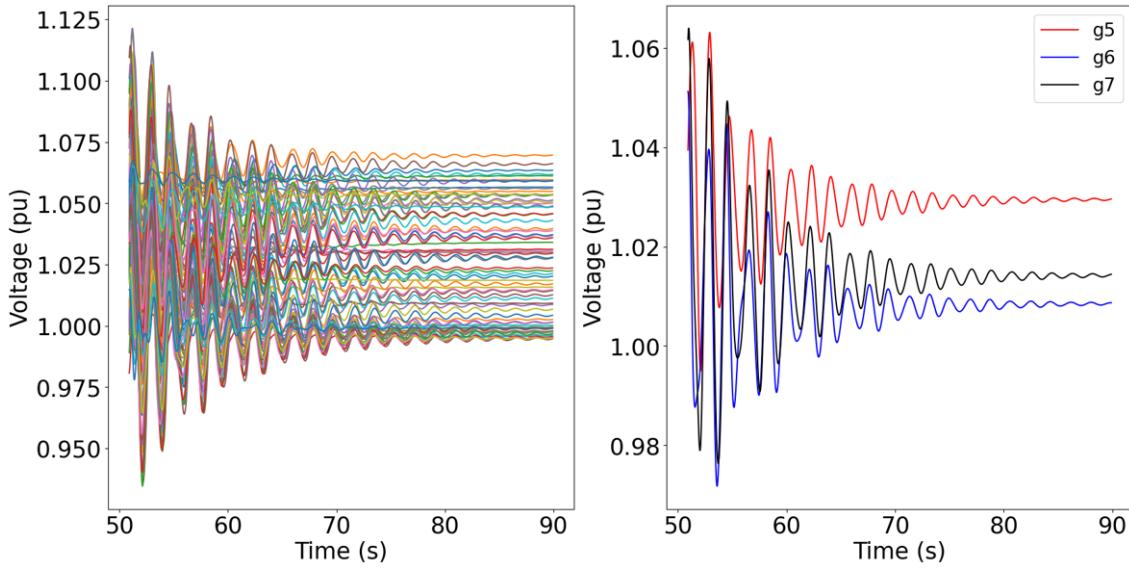


Figure 6.4: a) Voltage signal for all buses b) Voltage signal for generators g5,g6 and g7

The diagonal matrix  $\mathbf{D} \in R^{(s \times m)}$ , contains singular values of decreasing magnitude ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ ). If the elements of matrix  $\mathbf{X}$  are standardised, a correlation matrix  $\mathbf{C}(\mathbf{X}) = \mathbf{V}\Sigma\mathbf{V}^T$  can be obtained by squaring  $\mathbf{D}$  and dividing by  $(n - 1)$ . Similar to  $\mathbf{D}$ , matrix  $\Sigma$  is also a diagonal matrix and provides variance of  $\mathbf{X}$ , such that  $\sigma_1^2 \geq \sigma_2^2, \dots, \sigma_m^2 \geq 0$ .

The relationship between the PCs,  $\mathbf{Y}(y_1, y_2, \dots, y_m)$  and the original dataset  $\mathbf{X}(s \times m)$  is expressed as  $\mathbf{Y} = \mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{D}$ . The quotient  $(\sum_{q=1}^k \lambda_q) / (\sum_{q=1}^m \lambda_q) = 0.90$  is used to determine the dimensionality of the system under consideration as it describes the contribution of the  $k$ -th PC on the variance of the data. By keeping only  $k$  components, the original dataset can be reduced to lower dimensions  $y_k(t)$ , where  $(k \leq m)$ .

### 6.3.3 Empirical Wavelet Transform

EWT is a method for representing a signal or a time series as a combination of wavelets. A wavelet is a particular type of mathematical function that can be used to analyze signals and extract information about the signal's frequency content. The EWT is a generalization of the wavelet transform, a widely used method for analyzing signals in a time-frequency domain. One of the main advantages of the EWT is that it can be used to analyze signals with complex structures and non-stationary behavior, which cannot be easily analyzed with traditional methods such as the Fourier transform. The EWT is also well-suited for analyzing signals with sharp transitions or discontinuities, as it can adapt to local features in the signal. The following three steps can describe the EWT method [183]:

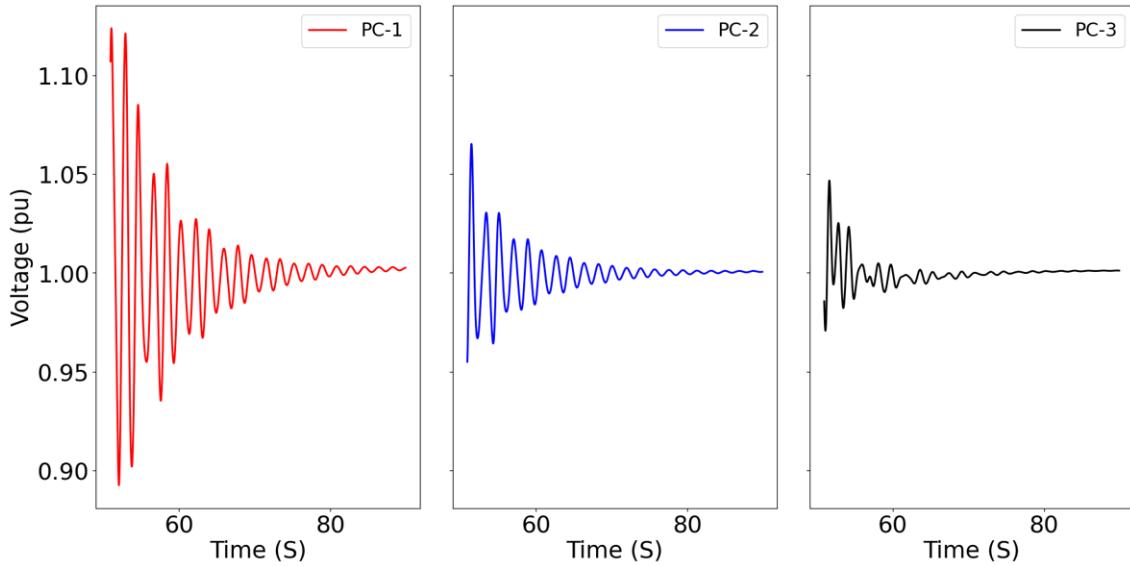


Figure 6.5: Reconstructed signal from first three principal components

- adaptively segment the Fourier spectrum;
- construct a suitable empirical wavelet filter bank according to the boundaries and filter the signal;
- reconstruct empirical modes.

The classic Wavelet transform applies fixed partitions over the Fourier spectrum of the signal for segmentation. In the EWT, the Fourier spectrum of the signal is adaptively segmented into multiple frequency bands, and a wavelet transform is applied to each band separately. This allows the EWT to capture both the local and global features of the signal and to adapt to the changing characteristics of the signal over time.

An empirical wavelet filter bank is a collection of filters that can be used to perform the EWT on a signal. The filters in the filter bank are designed to respond to different frequency components of the signal. They can decompose the signal into a series of wavelet coefficients as part of the EWT process. The wavelet coefficients produced by the filter bank can be analyzed in order to extract meaningful features from the signal, such as patterns or trends. The filter bank can also be used for denoising, compression, or classification tasks.

Reconstructing the signal from the wavelet coefficients can be considered a reverse convolution, in which the wavelet coefficients are convoluted with the inverse wavelet function to obtain the original signal. It's important to note that the reconstructed signal may not be exactly the same as the original signal due to the loss of information during the EWT process. However,

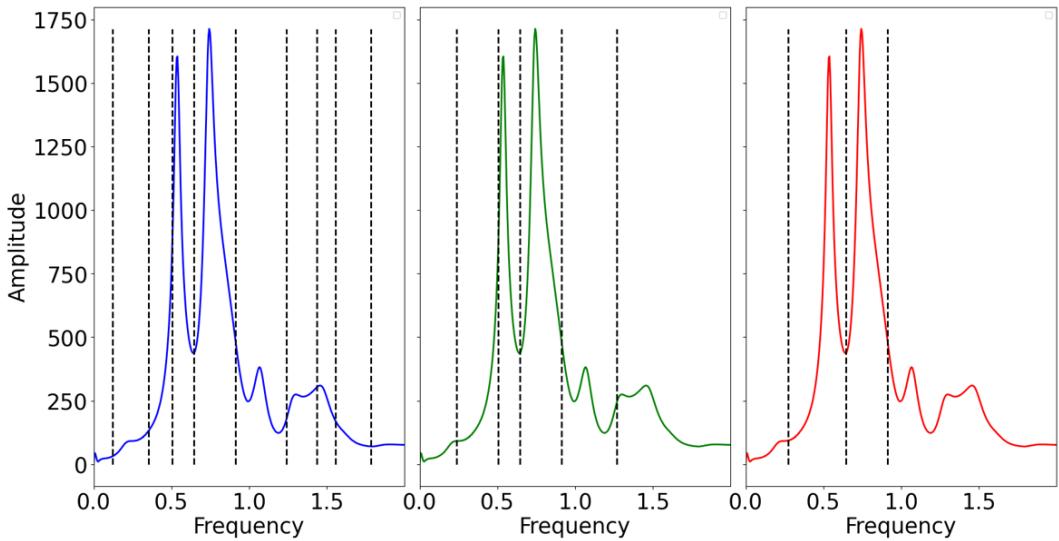


Figure 6.6: Boundaries detected using the EWT in the Fourier spectrum (the black dashed lines represent the position of boundaries). Segmentation of the Fourier spectrum of the signal for model-C2

the reconstructed signal should be a good approximation of the original signal and can be used for estimating the IF and IDR of the signal.

To calculate instantaneous parameters (IF and IDR), the individual oscillatory modes present in the signal should be separated from each other (mono-component). The steps required to analyse the signal are as follows:

*Step 1)* Apply fast Fourier transform to the discrete signal  $y(t)$  of the sliding window to obtain the frequency spectrum in the range of  $[0, \pi]$ .

*Step 2)* Let Fourier spectrum support be divided into  $N$  contiguous segments, then  $N - 1$  boundaries need to be extracted excluding 0 and  $\pi$ . To find the boundaries, we detect the local maxima  $L_n$  in the Fourier spectrum and obtain their corresponding frequency  $\omega_n$ , where  $n = 1, 2, \dots, N$ .

*Step 3)* The boundary  $\Omega_n$  between two segments is then defined as the centre of two consecutive maxima:

$$\Omega_n = \frac{\omega_n + \omega_{n+1}}{2} \quad (6.1)$$

where  $\omega_n$  and  $\omega_{n+1}$  are frequencies with  $n = 1, 2, \dots, N - 1$ . It should be noted that first ( $\Omega_0$ ) and last ( $\Omega_N$ ) boundary frequencies are 0 and  $\pi$ , respectively. This is because EWT has some constraints on the frequency range it operates on. It restricts the Fourier spectrum to the range  $[0, \pi]$ . This means that only positive frequencies of the normalized Fourier axis are considered

in the EWT. Positive frequencies are those that lie in the range  $[0, \pi]$  in the frequency domain. In many cases, when working with real-valued signals, the negative frequency components are redundant due to the symmetry properties of the Fourier transform.

*Step 4)* The empirical wavelet can be defined as a band pass filters on each segments of the frequency spectrum. The empirical scaling function  $\hat{\phi}_n(\omega)$  and the empirical wavelets  $\hat{\psi}_n(\omega)$  are given by:

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & \text{if } |\omega| \leq (1 - \gamma)\omega_n \\ \cos\left(\frac{\pi}{2}\alpha(\gamma, \omega_n)\right), & \text{if } (1 - \gamma)\omega_n \leq |\omega| \leq (1 + \gamma)\omega_n \\ 0, & \text{otherwise} \end{cases} \quad (6.2)$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1, & \text{if } (1 + \gamma)\omega_n \leq |\omega| \leq (1 - \gamma)\omega_{n+1} \\ \cos\left(\frac{\pi}{2}\alpha(\gamma, \omega_{n+1})\right), & \text{if } (1 - \gamma)\omega_{n+1} \leq |\omega| \leq (1 + \gamma)\omega_{n+1} \\ \sin\left(\frac{\pi}{2}\alpha(\gamma, \omega_n)\right), & \text{if } (1 - \gamma)\omega_n \leq |\omega| \leq (1 - \gamma)\omega_n \\ 0, & \text{otherwise} \end{cases} \quad (6.3)$$

where  $\alpha(\gamma, \omega_n) = \beta\left(\frac{1}{2\gamma\omega_n}\right)(|\omega| - (1 - \gamma)\omega_n)$ . The parameter  $\gamma$  ensures that no overlap between two consecutive transitions occur and can be selected as  $\gamma < \min_n \left(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n}\right)$ .

$\beta(x)$  is an arbitrary function defined as

$$\beta(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ 1, & \text{if } x \geq 1 \\ \beta(x) + \beta(1 - x) = 1, & \forall x \in [0, 1] \end{cases} \quad (6.4)$$

*Step 5)* Having defined the empirical wavelet and scaling function, the EWTs  $W_y^\epsilon(n, t)$  of the signal is defined in a way similar to the classic wavelet transform. The approximate coefficients can be expressed as the inner product of analysed signal  $y(t)$  with scaling function:

$$W_y^\epsilon(0, t) = \langle y, \phi_1 \rangle = \int y(\tau) \overline{\phi_1(\tau - t)} dt = (\hat{y}(\omega) \overline{\hat{\phi}_1(\omega)})^\vee$$

In the same way, the detailed coefficients are obtained by the inner product of analysed signal  $y(t)$  with empirical wavelets:

$$W_y^\epsilon(n, t) = \langle y, \psi_n \rangle = \int y(\tau) \overline{\hat{\psi}_n(\tau - t)} dt = (\hat{y}(\omega) \overline{\hat{\psi}(\omega)})^\vee$$

where  $W_y^\epsilon(n, t)$  denotes the detailed coefficients at time  $t$  for the  $n$  filter bank.  $\phi_1(\omega)$  and  $\psi(\omega)$  are empirical wavelet function and empirical scaling function respectively.  $\hat{\phi}_1(\omega)$  and  $\hat{\psi}(\omega)$  are Fourier transform of  $\phi_1(\omega)$  and  $\psi(\omega)$  which are defined by (6.2) and (6.3). The reconstructed signal can be obtained by:

$$y(t) = W_y^\epsilon(0, t)^* \phi_1(t) + \sum_{n=1}^N W_y^\epsilon(n, t)^* \phi_n(t) \quad (6.5)$$

The empirical mode  $y_n(t)$  can be given by:

$$\begin{aligned} y_0(t) &= W_y^\epsilon(0, t)^* \phi_1(t) \\ y_n(t) &= W_y^\epsilon(n, t)^* \psi_n(t) \end{aligned} \quad (6.6)$$

#### 6.3.4 Proposed Enhanced Empirical Wavelet Transform

To enable the automatic detection of boundaries for the adaptive decomposition of EWT, we propose SEWT for mode separation. A drawback of EWT is that when the analysed signal comprised of noise and contain frequencies close to each other some local maxima might appear in the detected peak sequence, which lead to improper segmentation. SEWT divides the Fourier spectrum without being stuck in the local maxima. However, as the window moves along the spectrum, some unnecessary segmentation is produced. Therefore, a limit on a spectrum scale is added, which retain peaks as a percentage of the maximum amplitude in the spectrum range. The implementation steps are given below:

1. **Input Signal:** Consider a real voltage signal  $y(t)$ , sampled at a frequency of  $F_s$ . To analyze this signal, we first subject the discrete signal  $y(t_d)$  to Fast Fourier Transform (FFT) to obtain its frequency spectrum, denoted as  $X_\omega$ .
2. **Sliding Window:** Determine the step size for the sliding window to slide over the spectrum  $X_\omega$  as the reciprocal of the Nyquist frequency ( $\mathcal{N}$ ) of the signal. The Nyquist frequency represents half of the sampling frequency, i.e.,  $\mathcal{N} = \frac{F_s}{2}$ . Set the step size ( $S$ )

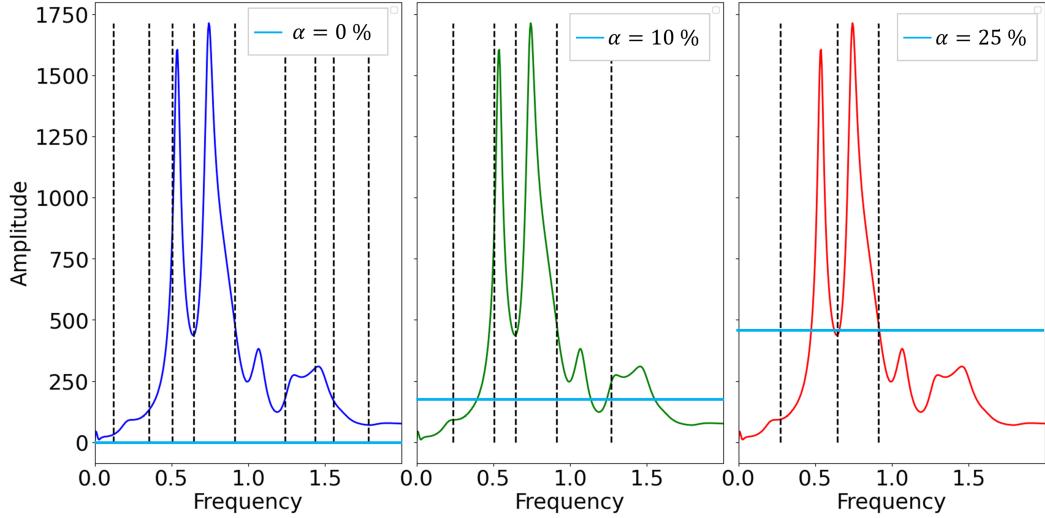


Figure 6.7: Segmentation of Fourier spectrum using SEWT, with the  $\alpha = 0\%$  (blue),  $\alpha = 10\%$  (green) and  $\alpha = 25\%$  (red) of the maximum amplitude.

as the reciprocal of the Nyquist frequency:  $S = \frac{1}{N}$ . This choice ensures that the sliding window has sufficient resolution to cover the low-frequency range relevant to the inter-area oscillations.

3. **Sliding Window:** Determine the step size for the sliding window as the reciprocal of the Nyquist frequency ( $N$ ) of the signal. The Nyquist frequency represents half of the sampling frequency, i.e.,  $N = \frac{F_s}{2}$ . Set the step size ( $S$ ) as the reciprocal of the Nyquist frequency:  $S = \frac{1}{N}$ . This choice ensures that the sliding window has sufficient resolution to cover the low-frequency range relevant to the inter-area oscillations.
4. **Peak Detection:** For each segment  $\Lambda_n$ , identify the maximum value and store these maxima in a sequence, denoted as  $L_{\max} = L_1, L_2, \dots$ . To prioritize dominant peaks, sort the sequence  $L_{\max}$  in decreasing order.
5. **Filtering Peaks:** Apply a filtering step to retain only the maxima  $L_{\max}$  that exceed a selected threshold  $\alpha$ . This threshold determines which peaks are retained as a percentage of the maximum amplitude. The filtered maxima are stored in the sequence  $L_{\max}^Z = L_n \geq \alpha | n = 1, 2, \dots, N$ . By doing this, the algorithm avoids being trapped by local maxima and focuses on significant peaks in the spectrum.
6. **Boundary Definition:** Define the boundaries between two consecutive maxima as the midpoints, similar to the process described in equation (6.1). These boundaries mark the frequency ranges that will be used for mode separation in the EWT.

Fig. 6.7 illustrates the segmentation of the Fourier spectrum using the SEWT algorithm. The dashed vertical lines represent the detected Fourier boundaries. By adjusting the threshold  $\alpha$ , we can retain the most dominant frequencies in the signal, which provides valuable information for analysis or further processing. However, selecting the appropriate value for  $\alpha$  involves a trade-off between preserving essential frequency components and suppressing noise or less significant components. There is no universally optimal threshold value, as it depends on the signal's characteristics and the specific requirements of the application. The choice of  $\alpha$  should be based on a balance between preserving important information and suppressing noise or less relevant components to achieve optimal mode separation.

### 6.3.5 Selecting the signal with highest power

After performing the EWT on the signal, the resulting modes are typically ranked based on their power or energy content. Each mode represents a specific frequency range or characteristic of the signal. These modes are sorted based on their power, which represents the amount of energy carried by each mode. The mode that exhibits the highest power or energy is then retained for further analysis and interpretation. By identifying the mode with the highest power, we can focus on the most dominant and significant component of the signal.

Retaining the mode with the most power allows us to prioritize the analysis of the strongest and most influential oscillatory behavior or pattern present in the signal. This mode often represents the dominant mode of interest or the major underlying phenomenon contributing to the observed signal dynamics.

The signal obtained from the EWT is typically a discrete signal. This decomposition process results in a set of discrete signals corresponding to each mode.

For a discrete signal  $y(t)$  defined over a finite sequence of samples  $N$ , the power is calculated as the average of the squared magnitude of each sample:

$$P = \frac{1}{N} \sum_{t=0}^{N-1} |y(t)|^2 \quad (6.7)$$

Here,  $y(t)$  represents the discrete signal obtained from the EWT. The absolute value squared,  $|y(t)|^2$ , represents the power of each sample. Summing up the squared magnitudes of all the samples and dividing by the total number of samples  $N$  yields the average power of the signal.

By calculating the power of the EWT output signal, we can assess the energy or strength of the signal, which can provide insights into its characteristics and behavior.

### 6.3.6 Hilbert Transform

The empirical modes are narrow band components, therefore HT can be applied to study the signal's time variability in detail. For a given real signal  $y(t)$ , the analytical signal  $z(t)$  can be given as:

$$z(t) = y(t) + iH[y(t)] = A(t)e^{i\theta(t)} \quad (6.8)$$

where  $A(t)$  represents instantaneous amplitude (IA),  $\theta(t)$  is the instantaneous phase (IP) and  $H[y(t)]$  is the HT of  $y(t)$  and is defined using Cauchy principal value (*p.v.*) as:

$$H[y(t)] = \frac{1}{\pi} p.v. \int_{-\infty}^{+\infty} \frac{y(\tau)}{t - \tau} d\tau \quad (6.9)$$

In terms of  $y(t)$  and its HT, the IA, IP and instantaneous frequency (IF) are defined as:

$$IA = A(t) = \sqrt{y^2(t) + H[y(t)]^2} \quad (6.10)$$

$$IP = \theta(t) = \tan^{-1} \left( \frac{H[y(t)]}{y(t)} \right) \quad (6.11)$$

$$IF = f(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} \quad (6.12)$$

The ID function  $\zeta(t)$  is calculated using the technique provided in [184]:

$$\zeta(t) = \sqrt{\frac{\rho(t)^2}{1 + \rho(t)^2}} \quad (6.13)$$

where

$$\rho(t) = \left| \frac{\ln \frac{IA_0}{IA(t)}}{2\pi \cdot IF(t) \cdot t} \right| \quad (6.14)$$

and  $IA_0$  is the initial amplitude.

The integral in the Hilbert Transform (HT) equation, given by equation (6.9), is used to compute the analytic signal, which is a complex representation of the original real signal  $y(t)$ . The HT is a mathematical operation that provides a way to extend a real signal into the complex domain,

enabling the analysis of the signal's instantaneous attributes such as amplitude, phase, and frequency.

When the variable of integration,  $\tau$ , equals  $t$  in the integral, the denominator of the integrand becomes zero, which results in a singularity. This leads to an undefined value for the HT at that specific point. To address this issue and provide a well-defined result for the HT, the Cauchy principal value (*p.v.*) is used.

The Cauchy principal value is a mathematical concept used to handle integrals with singularities. For the HT integral, the Cauchy principal value ensures that the integral is evaluated as a limit around the singularity, taking into account both sides of the singularity. This process allows for the proper computation of the HT even when  $\tau$  approaches  $t$ .

### 6.3.7 Average parameters

While the average frequency  $F_{avg}$  is calculated by taking the average of (6.12), to obtain the average damping coefficient, a damping equation (mentioned in [185]) is commonly used. However, it relies on computing the derivative, which introduces discontinuities. This work estimates the average DR  $\zeta_{avg}$  with a combination of a logarithmic decrement and IA envelope obtained from the (6.10). The most common way to calculate the DR of a free decaying oscillation is given below:

$$\zeta = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}} \quad (6.15)$$

where  $\delta$  is the logarithmic decrement of peak amplitudes of two points  $o_1$  and  $o_2$  in oscillatory signal, exactly  $v$  cycles apart

$$\delta = \frac{1}{v} \ln \frac{o_1}{o_2} \quad (6.16)$$

By using (6.15) and (6.16),  $\zeta_{avg}$  can be estimated much more precisely from successively discrete decaying points of the IA envelope (Fig. 6.11). First, the logarithmic decrement is calculated in (6.15) for each successive sample ( $v = 1$ ) in the envelope and then dividing it by the total number of samples  $V$  in the envelop:

$$\delta_i = \frac{\left( \frac{1}{v} \ln \frac{IA_i}{IA_{i+1}} \right)}{V} \quad (6.17)$$

where  $IA_i$  is the value of the  $i$ -th instantaneous envelop sample. Next  $\zeta_{avg}$  is calculated by substituting (6.17) in (6.15) and summing up all the values:

$$\zeta_{avg} = \sum_{i=1}^V \frac{\delta_i}{\sqrt{(2\pi)^2 + \delta_i^2}} \quad (6.18)$$

## 6.4 Online consideration

In order to meet the criteria for real-time analysis, a procedure must be capable of estimating the stability of the system before to the arrival of the next set of measurement data. This highlights the importance of employing fast and accurate techniques for analyzing online low-frequency inter-area oscillations. These techniques are essential for providing the control system with sufficient time to evaluate the situation and make appropriate decisions in a timely manner [186].

This work proposes the utilization of a moving window technique to closely align with the online detection method to estimate the modal frequency and DR. The moving window strategy assumes that recent data carries more significant implications than historical data. As new samples arrive, the old samples are discarded from the window, and the parameters of the models are updated iteratively. There is a significant trade-off between memory to keep data and performance. The larger the memory results in higher accuracy and low performance—lower memory results in low accuracy but faster performance [187].

The main steps of the proposed methodology are summarised in the flowchart of Fig-6.1.

## 6.5 Numerical results

### 6.5.1 System model and case study description

The proposed scheme is tested using the Nordic test system [188]. Time-domain simulations were conducted using the dynamic simulation software PyRAMSES [132], and the results were saved in a database, emulating the PMU measurements. Only voltage magnitude measurements

Table 6.1: Small-signal stability analysis.

	C1		C2		
	Mode2	Mode3	Mode2	Mode3	Mode4
Frequency(Hz)	0.538	0.742	0.533	0.766	0.949
DR (%)	3.7	3.9	-0.4	-0.1	0.3

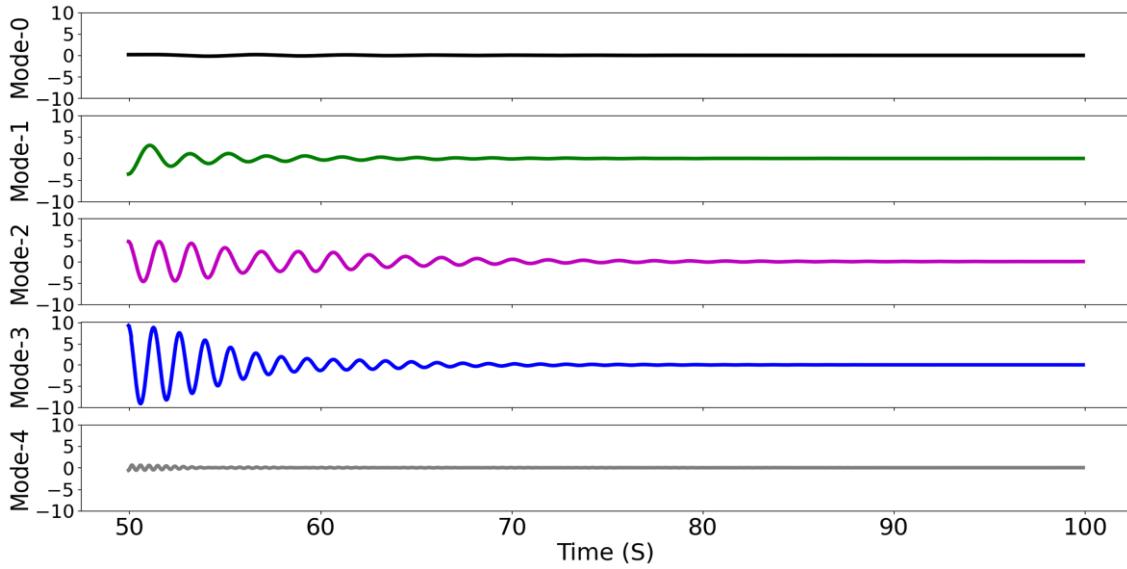


Figure 6.8: Case 1: Decomposition of PC1 voltage signal using SEWT.

are used from the 20 synchronous generators. A window of 20 s with a step size of 1 s is chosen (19 s overlap). The proposed scheme utilizes a window size of 20 seconds for data from Phasor Measurement Units (PMUs). This window size is chosen to allow for the analysis of oscillation cycles while ensuring the continuity of data. The window operates with a step size of 1 second, resulting in a 19-second overlap between consecutive windows.

The 19-second overlap indicates that once a 20-second window of data is stored, the window moves forward by 1 second at a time. This rapid movement of the window ensures that new data is continuously incorporated into the analysis, providing real-time insights into the system behavior. The window size of 20 seconds is selected to accommodate the analysis of oscillation cycles. It allows for capturing multiple cycles within the window, facilitating the examination of the characteristics and dynamics of these low-frequency oscillations. By analyzing longer time intervals, more comprehensive information about the system's behavior and stability can be obtained. This choice of window size, combined with the 1-second step size, strikes a balance between capturing sufficient data for accurate analysis and ensuring a quick update rate to facilitate real-time monitoring and decision-making.

Two case study scenarios were investigated a) **Stable oscillations model (C1)** and b) **Unstable oscillations model (C2)**. For C2, several generator PSS controllers were deactivated to destabilize the system. In both cases, the modes are excited by applying a three-phase fault at bus 4072 at time  $t = 50$  s, which is automatically cleared after 100 ms. All the measurements

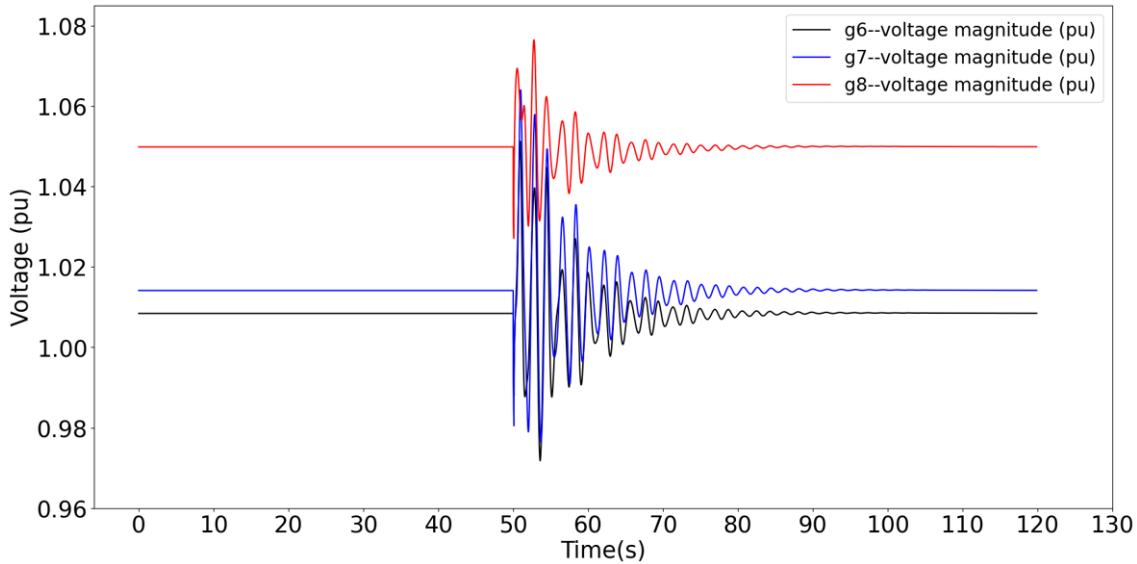


Figure 6.9: Case 1: Voltage signal for three generator buses, g6,g7 and g8 after the disturbance is introduced at time  $t = 50$  s

are sampled at 50 Hz. Gaussian white noise is added to the voltage signals with a signal-to-noise ratio (SNR) of 40 dB.

We performed a small-signal stability analysis on the original differential-algebraic model used for the above two cases and identified the main electromechanical modes as detailed in Table 6.1. These are used as a benchmark (BM) for our real-time detection algorithm that only uses the data flow.

### 6.5.2 Case study 1(Stable Model):

The first step applies PCA to the normalized voltage magnitude data. The first principal component (PC1) accounts for 83% of the variation, while the second component (PC2) accounts for 8% of the total explained variance. Therefore, only the first two PCs are retained, capturing a total of 91% variation. Fig. 6.9 shows three voltage signals from generator buses g6, g7, and g8, after the disturbance is introduced at time  $t=50$ s. Fig. 6.10-a shows the same voltage signal with an SNR of 40 dB. Fig. 6.10-b shows the reconstructed signal obtained through PCA for 20 voltage buses. PCA is a powerful statistical technique that can be used to determine lower dimensional representations of sampled data and reduce noise from the data. It can be noticed from Fig. 6.10-b that PCA has successfully reduced the signal's noise level.

For case study 1, a change in the variance level  $\sigma$  of the window was experienced on the system, followed by an introduction of a disturbance at time 50 s. At this point, the variance reached

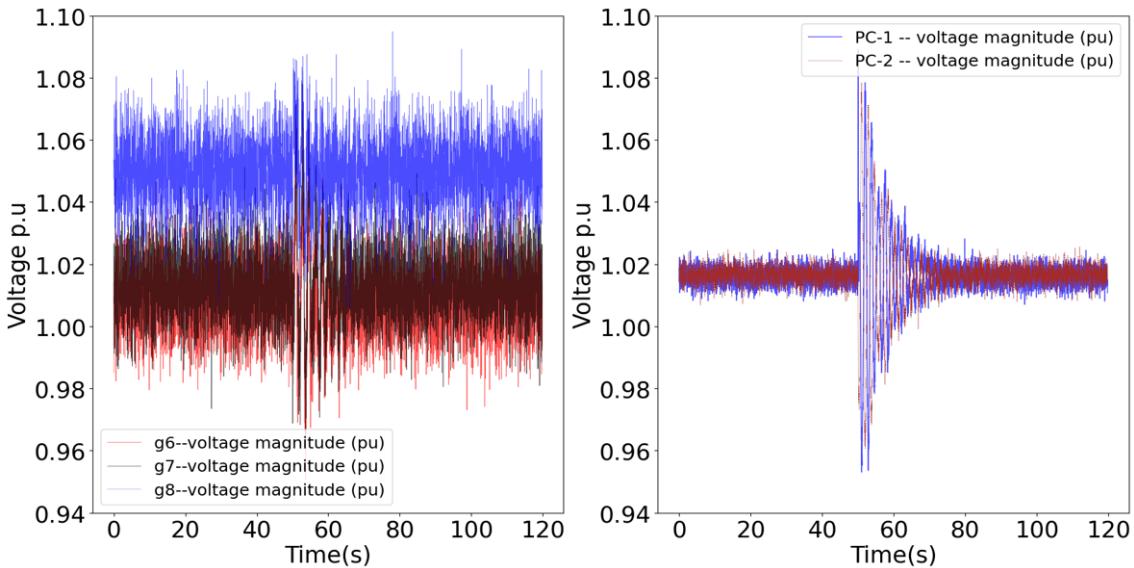


Figure 6.10: Case 1: On the left voltage signal for generator buses g6, g7 and g8 with the SNR of 40 dB. On the right, PC1 and PC2 for the 20 generator buses from Nordic-32 test system. PC1 captures 83% of the variation, while the PC2 captures 8% of the total explained variance.

its threshold level of 0.05. Once the event is detected, SEWT is applied, and the signal is decomposed. Fig. 6.7 shows the results of the segmentation of the Fourier spectrum using three different threshold  $\alpha$  values and a step size of 0.04 Hz. Fig. 6.8 illustrates the modes extracted using SEWT with  $\alpha = 15\%$ . The method analyses all the modes. However, the signal power of Mode-0 and Mode-1 is negligible and therefore discarded. We only consider Mode-2 and Mode-3 as these modes contain more abundant impulse information. Therefore, HT is applied to Mode-2 and Mode-3 only. Their corresponding frequencies are estimated 0.54 Hz and 0.77 Hz. Meanwhile, the noise is separated from the dominant components, represented by Mode-4 in Fig. 6.8.

Fig. 6.11 shows the decaying amplitude from HT for Mode-2. The envelop obtained from the application of the HT is fitted to the reconstructed EWT signal. At the beginning and end of the signal there is a difference due to the end effect [185] of HT as a result of a finite time series. In a real time, analysis, we have a continuous stream of data therefore to overcome this issue the left and right tails of the IA envelop are discarded without extending the ends. In our analysis, we utilize a window size of 20 seconds with a step size of 1 second. As part of our methodology, we purposefully exclude the segment extending 0.5 seconds from the beginning and 0.5 seconds from the end of each window. This selective exclusion is achieved through a binary masking technique, where the mask value is set to 0 within the portion that needs to be discarded and 1 elsewhere. No information is lost as the discarded data is overlapping with

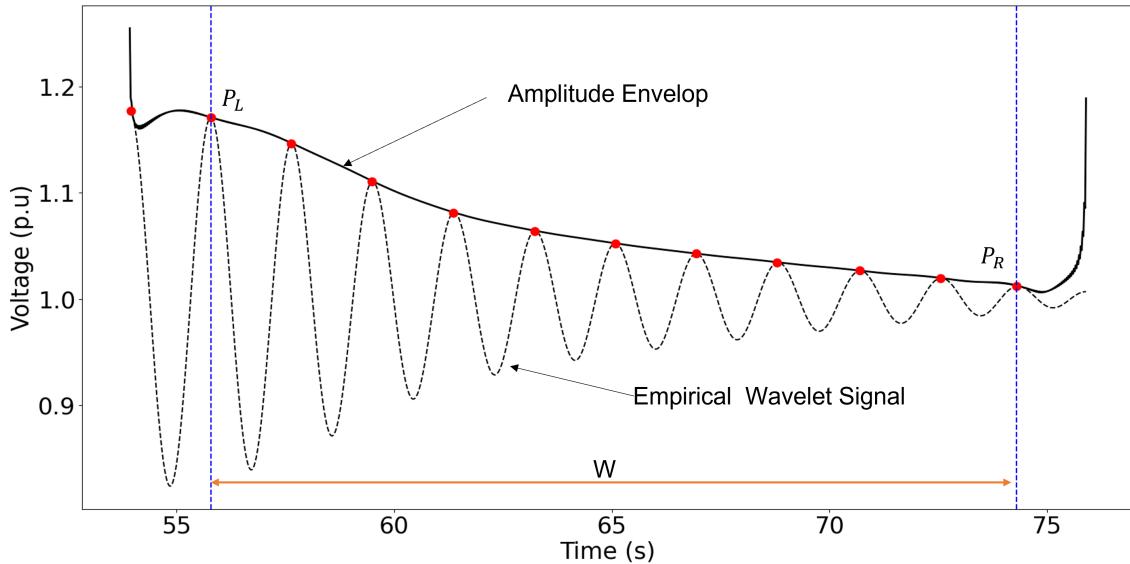


Figure 6.11: IA envelop fitted with the reconstructed EWT signal.

the next/previous sliding windows. The average DR and the IDR of the truncated envelop are computed using (6.18) and (6.14) respectively.

The HT method is applied to each selected mode to acquire the instantaneous components for feature extraction. A deeper insight into the nature of temporal behavior of the oscillations is obtained from analyzing these instantaneous attributes for each EMF. The IF and ID for Case-1 are shown in Fig. 6.12. Simulation results show that the instantaneous attributes of the oscillatory response exhibit increased damping, thus indicating the system approaching stability after the disturbance. Of particular interest and practical significance is the analysis of the IF in Fig. 6.12, which shows that the transient signal has a frequency of 0.54 Hz and 0.76 Hz for mode-2 and mode-3, as expected from conventional linear analysis techniques.

Fig. 6.13 shows the time evolution of the average DR for different window sizes. The proposed scheme can estimate the damping both during ambient conditions and ring down (i.e., before and after the tripping of the branch) rather accurately. However, when a sudden disturbance occurs, resulting in a ringdown oscillatory response, the algorithm will see that mode as negatively damped for the time windows that have the event start time in the middle of the window. For example, if the mode is not excited in the first 15 s of the window and is excited in the last 5 s after the event (Fig. 6.14), the peak of that mode will increase in magnitude throughout that window. To solve this issue, the program skips the analysis windows where a high jump in the signal occurs and starts the analysis once that sharp jump leaves the window. The channel is

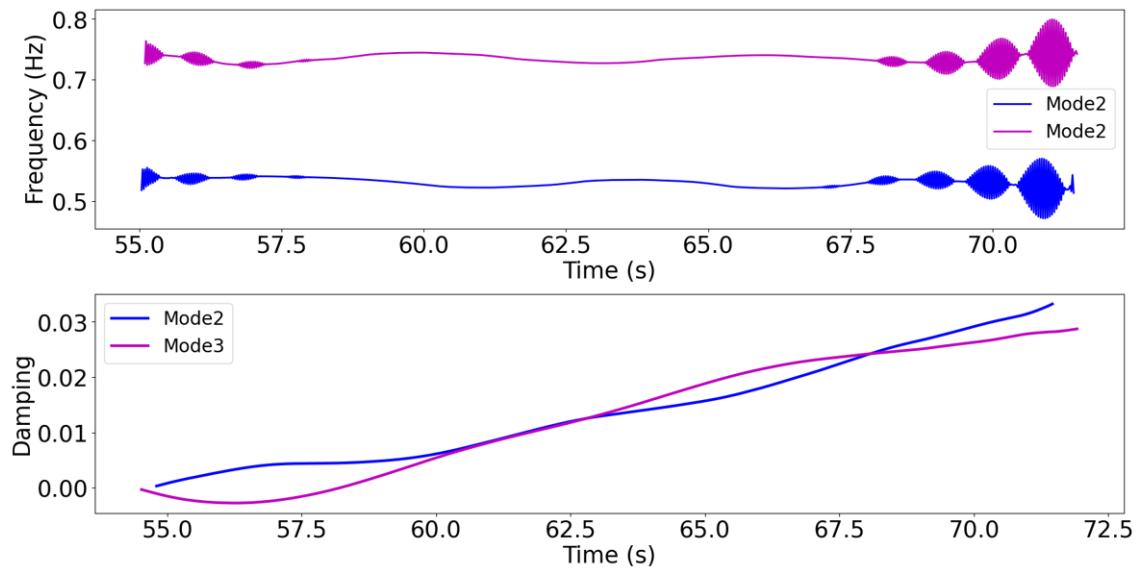


Figure 6.12: Case 1: Top panel shows IF  $f(t)$  and the bottom panel shows IDR  $\zeta(t)$  of extracted modes from PC1.

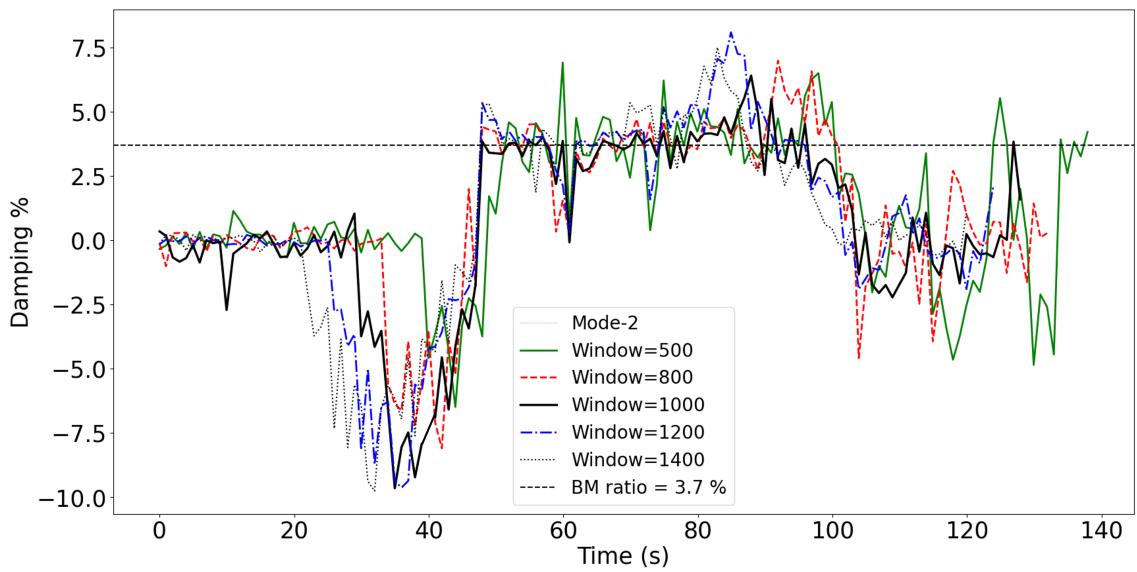


Figure 6.13: Case 1: Estimates for the average DR  $\zeta_{avg}$  for different window sizes.

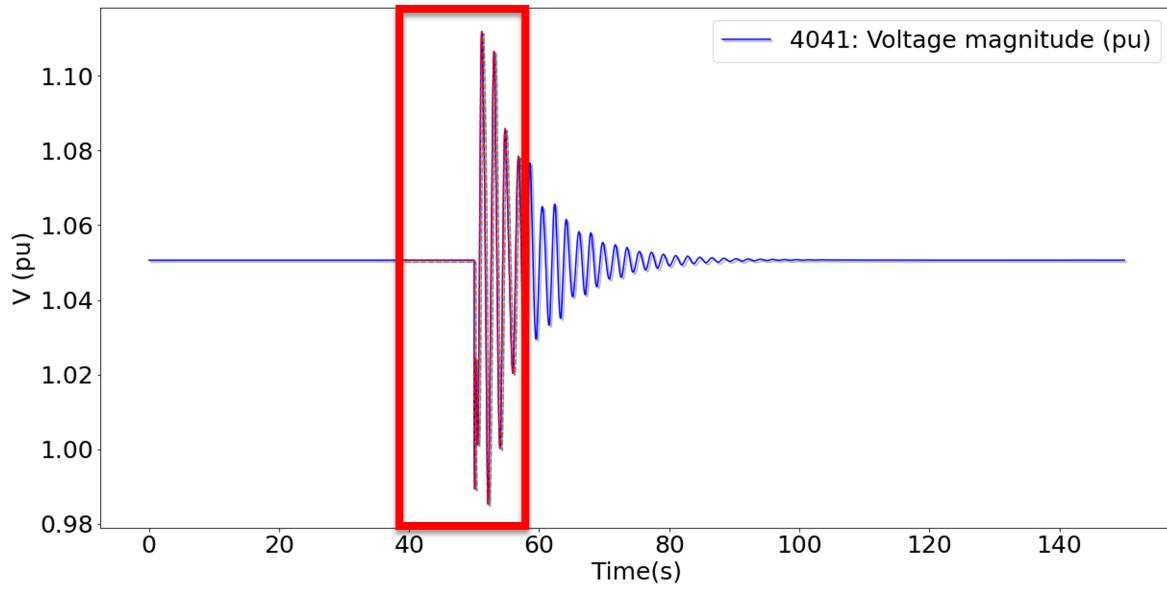


Figure 6.14: Excitation of event in the middle of the window

discarded if the  $\sigma$  of the window is below a preset threshold.

To ensure the threshold for detecting jumps or events in the signal is carefully chosen, a rigorous approach is followed, which involves running simulations multiple times and analysing the resulting data. This process helps determine the level of jump expected when an actual event occurs and establishes the baseline voltage magnitude when there is no event. Simulations are conducted with different event scenarios that are representative of the system's behaviour and potential disturbances.

These scenarios include various types of events, such as faults, trip branch, and load changes. Each simulation run generates a dataset that captures the response of the system under different conditions. The datasets obtained from the simulation runs are then analyzed to identify and quantify the jumps or changes in voltage magnitude associated with each event. Statistical techniques were employed to determine the average jump level and its variability for different event types. This analysis helps establish a confidence interval for the expected jump in voltage magnitude when an event occurs. Similarly, the datasets obtained from the simulations without any events are analysed to understand the baseline voltage magnitude when no disturbances are present. Statistical analysis can be performed to determine the average voltage magnitude and its variability during non-event periods. The threshold for detecting jumps or events was set based on the findings from the above analyses. The threshold can be established as a level above the expected baseline voltage magnitude during non-event periods, taking into account

the confidence interval of the jump level during event scenarios. This ensures that jumps above the threshold are indicative of actual events, while smaller fluctuations are considered as noise or insignificant variations.

In order to provide further clarification regarding the method employed to address the aforementioned issue, the events illustrated in Fig. 6.14 should be taken into consideration. Two modes excited by this event is estimated to be 0.54 Hz and 0.77 Hz with a decay rate of 3.7% and 3.9% respectively. If the low  $\sigma$  window is processed, the program would see the mode negatively damped for the first part of the event. The output of the SEWT algorithm during this phase would result in a pseudo negative, as shown by the green and red dots in Fig. 6.15.

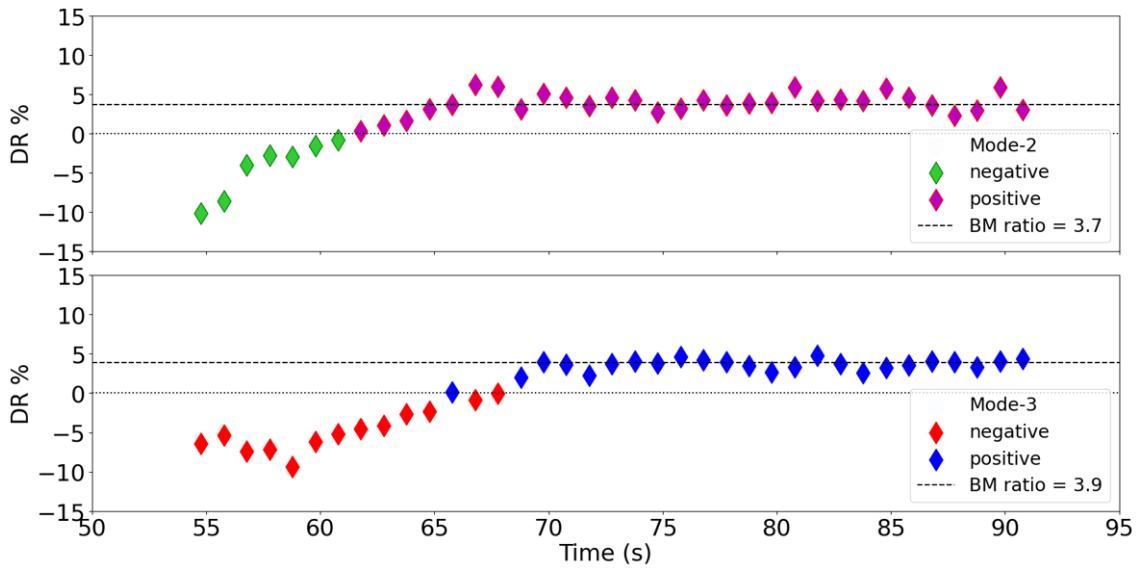


Figure 6.15: Case 1: Estimates for the average DR.

As the ringdown begins to pass through the analysis window, the damping estimate approaches its correct value and the algorithm eventually finds the correct DR. To avoid the pseudo-negative DR, only the high variance window is processed by using a threshold, and the low variance window is ignored. Thus, the SEWT will only start reporting when the sliding window moves into the oscillatory zone and the pre-event results are ignored. This can be seen in Fig. 6.16.

### Effect of window size on the detection process

The window length defines the length of the data over which the algorithm computes DR and frequency. The window moves as the new data comes in. If the window is large, the parameters calculated for oscillatory modes are closer to the data's stationary statistics. A longer window gets smoother statistics for data that does not change rapidly. For dynamic

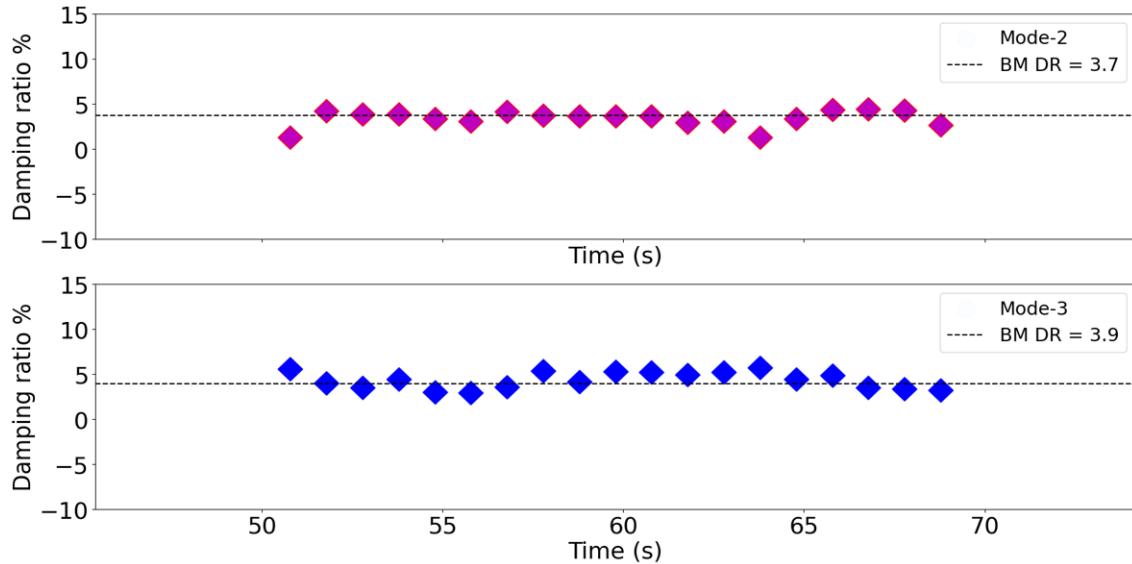


Figure 6.16: Case 1: Estimation of Damping ratio for mode-2 and mode-3.

data, a smaller window is preferred. There is a trade-off in the choice of the size of the sliding time window and accuracy. More precision is achieved with a larger time window, but more information must be stored. Moreover, the standard duration of the oscillation and resolution of frequency must be considered in the choice of the window size. In the case of oscillation, where the oscillation frequency is taken as a criterion, the window length can be calculated as proportional to the period of the oscillation. Specifically, for our analysis (interarea), the window length is considered proportional to the modal frequency.

There are two main parameters to handle the speed and accuracy of the window: the window size ( $w$ ) and sliding size ( $s$ ). The window size determines the number of data points the detection algorithm will get to apply the statistics, and the sliding size is the period to which the model is updated. In Fig. 6.17, various windowing detection procedures for mode-3, corresponding to different window sizes, are depicted for case-C1. The process is initiated by inserting a disturbance into the system at time  $t=50$  s in the central area, generating an oscillation in the model. The machine in the southern area oscillates as a result of this disturbance. Subsequently, a dominant mode extraction process is carried out using SEWT to detect the DR of the signals for low-frequency electromechanical oscillation.

In each window, the sliding size is kept constant, i.e.,  $s = 50$  samples equivalent to one second, while the window size increases between the sample range of  $w=200$  (four seconds) to  $w=1400$  (twenty-four seconds). For each window size, the average DR is calculated. It can be noticed from Fig. 6.17 that when  $w$  is minimal ( $w=200$ ), the accuracy is very low. As the window

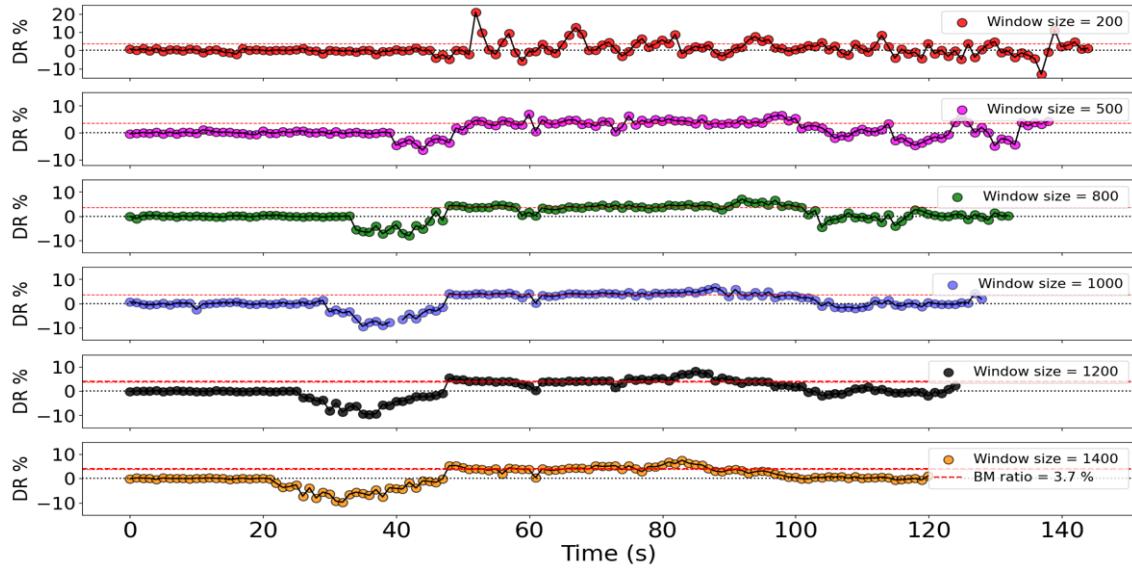


Figure 6.17: Case 1: Estimates for the average DR  $\zeta_{avg}$  for different windows with different sample size.

sample size increases, the accuracy is improved. Fig. 6.18 shows the RMSE for each window size. Notice that the error is minimum when the window size is 1000 samples, which is approximately proportional to the period of the oscillation.

### 6.5.3 Case study 2 (Unstable Model) :

For C2 the voltage response of the system is displayed in Fig. 6.19. The total variance accounts for 93% of the information (85% for PC1 and 8% for PC2). Note that with this disturbance, in addition to Mode-1 and Mode-2 discussed in case C1 there is also another mode detected around 0.9 Hz. Analysis of the voltage variables for C2 indicates that these modes are visible in all the busses. Therefore these modes are adequately captured by both PCs leading to a low MSE in Table 6.4.

Fig. 6.20 displays the SEWT estimation of the average DR and average dominant frequency (NF) for case C2. The red dots indicate a negative DR, while the blue dots represent positive DR estimates. As observed in Fig. 6.20, the algorithm provides dynamic DR estimates of the mode during the disturbance. These findings demonstrate that the proposed approach can offer operators useful information regarding the changing modal parameters of inter-area oscillation during this intricate disturbance process.

In particular, the average DR of mode-2 is noteworthy. The benchmark DR for mode-2 is -0.4, indicating a relatively high damping level. However, due to the continuously moving

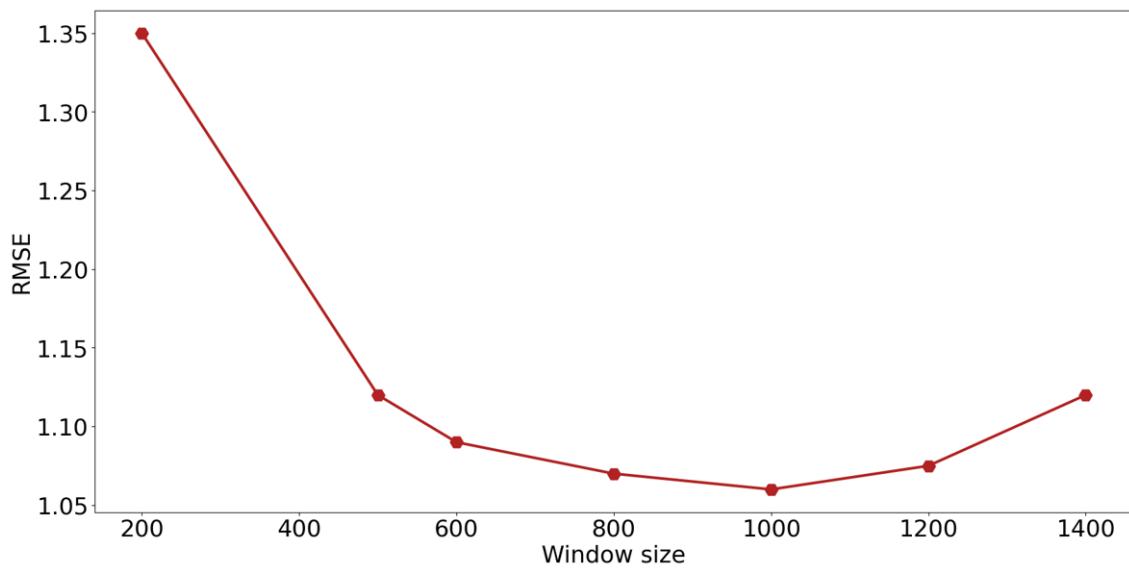


Figure 6.18: Case 1: RMSE for different window size

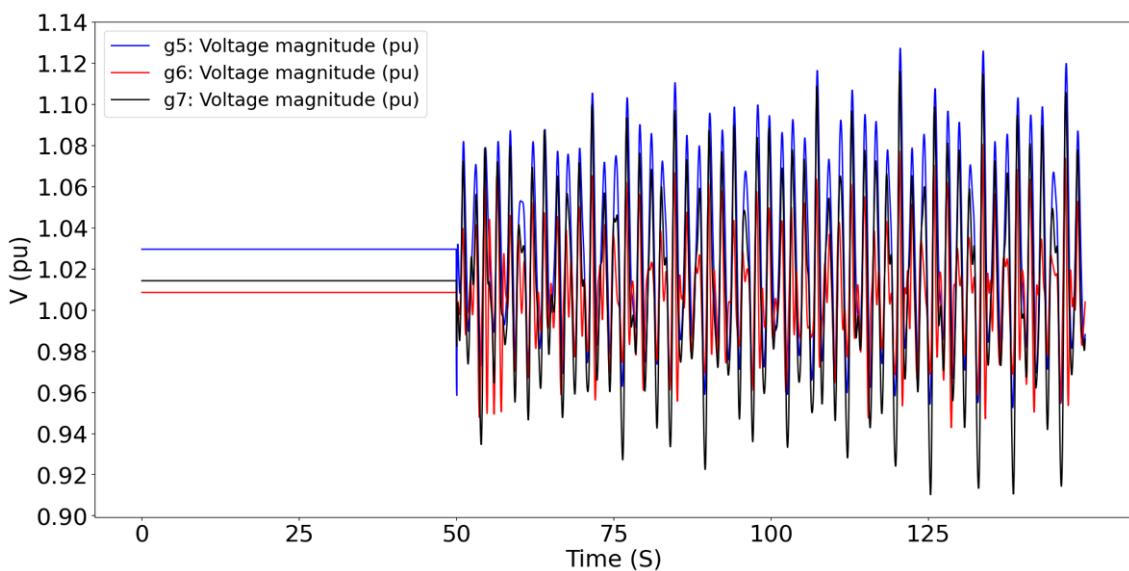


Figure 6.19: Case 2: Temporal evolution of voltage signals.

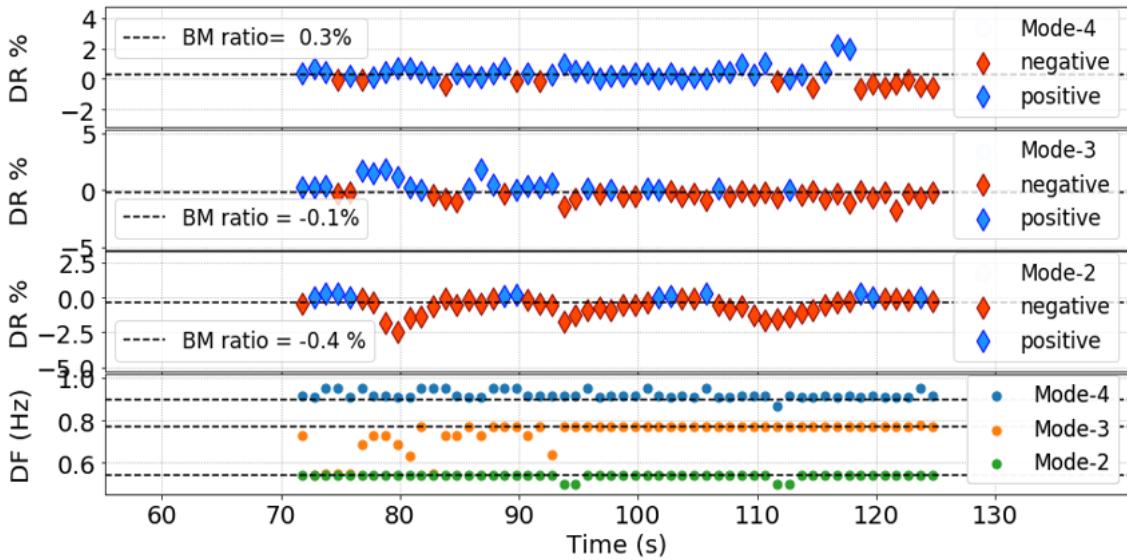


Figure 6.20: Case 2: Estimates of average DR  $\zeta_{avg}$  and average frequency  $F_{avg}$  from PC1.

window used in the SEWT algorithm, the average DR varies as the window progresses along the signal. It is important to highlight that the observed variation primarily stems from the calculation of the average DR based on the instantaneous parameters. Unlike methods such as matrix pencil, or Prony, the SEWT is an adaptive signal processing technique that excels at capturing instantaneous attributes. This attribute endows the SEWT with a unique capability to provide real-time insights into the dynamic behavior of the system, enhancing its effectiveness in practical applications.

Overall, the findings highlight the utility of the proposed approach in offering operators valuable information regarding the changing modal parameters of inter-area oscillation. The average DR, although subject to variation as the window moves along the signal, provides insights into the damping characteristics of the mode when the oscillation is fully within the window. This knowledge can aid in assessing system stability, making informed decisions, and implementing appropriate control strategies during complex disturbance events.

#### 6.5.4 Detection of false alarm

To create a reliable online oscillation detection system, it is crucial to minimize false alarms and only trigger oscillation alerts when the identified oscillation modes are consistently present. Grid operators would not want a method that is highly sensitive and generates a large number of false oscillation events. Once an oscillation mode of interest is detected, it should display consistent modal features during a specific time interval.

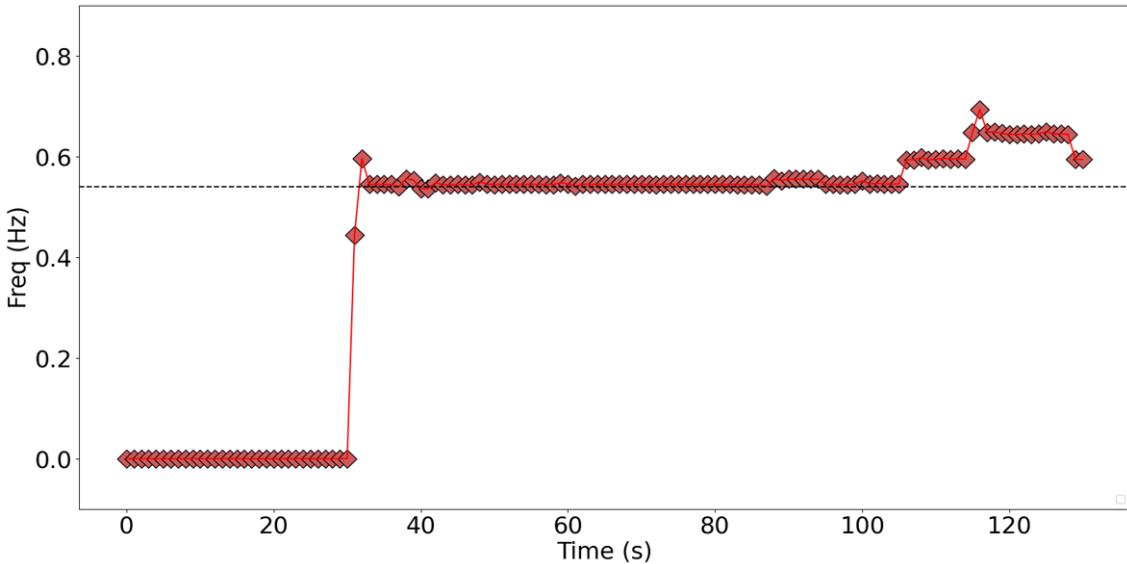


Figure 6.21: Case 2: Estimates of average frequency

In order to assess the accuracy of the proposed algorithm in detecting oscillations with minimal false positives, measurements of the signal frequency without noise and with varying noise levels were taken. The experiment was repeated 50 times, and the mean and confidence interval were calculated. Fig. 6.21 illustrates the average NF for the dominant mode of case-study C1 in the absence of noise. The mean and confidence interval remain the same when there is no noise. As can be observed from the figure, the algorithm consistently estimates the dominant frequency of the voltage signal to be zero under normal operating conditions (first 30 time windows). However, at time  $t=30$  s, the branch between bus 4042 and bus 4044 was tripped, introducing a disturbance into the system. As a result, the dominant frequency of the signal, as estimated by the algorithm, increased to 0.54 Hz.

Table 6.2: Average frequency without noise

	Natural Frequency		Damping Ratio	
<i>Before disturbance</i>	<b>True Negative = 30</b>	0 Hz	<b>True Negative = 30</b>	0 %
<i>After disturbance</i>	<b>True Positive = 40</b>	0.54 Hz	<b>True Positive = 40</b>	3.5% - 4.0%

Table 6.2 presents the outcomes for average DR and average NF in the absence of any noise interference. The observations were recorded for a total of 130 instances, out of which 70 observations were considered as a reference or ground truth. These observations were divided into two sets: 30 observations before the disturbance and 40 observations after the disturbance. The observations before the disturbance, occurring in the time windows from  $T_w=0$  seconds to  $T_w=30$  seconds, were labelled as true negatives since they had an NF of 0 Hz and a DR of 0%.

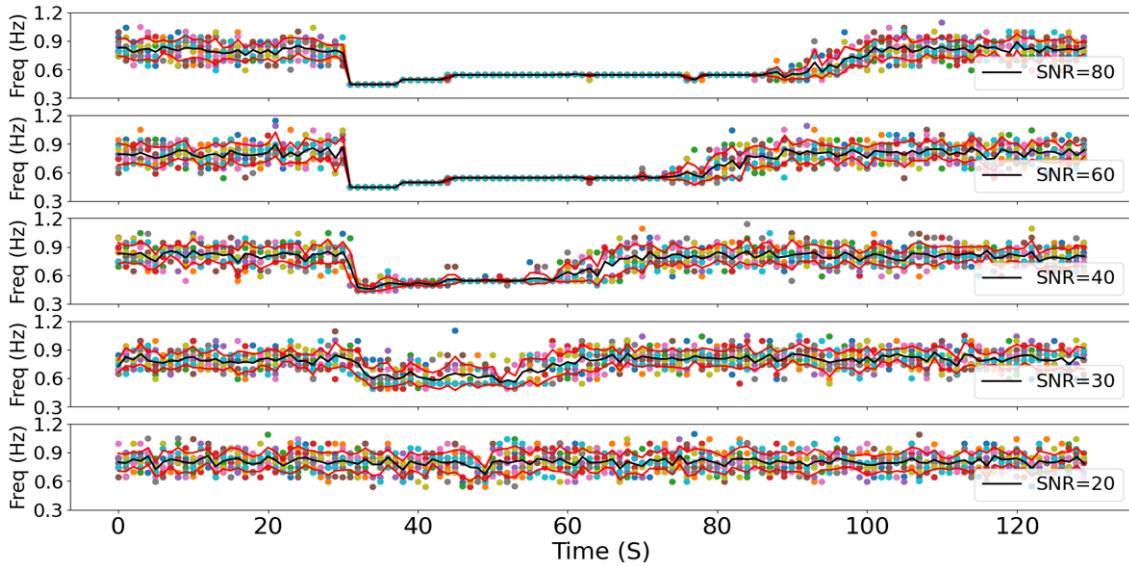


Figure 6.22: Case 1: Estimates of average average frequency with different SNR

The observations after the disturbance, occurring in the time windows from  $T_w = 40$  seconds to  $T_w = 80$  seconds, were labelled as true positives. These observations had an NF of 0.54 Hz and a DR ranging from 3.5% to 4.0%.

To ensure accurate measurements and reliable classification, the transition windows from  $T_w = 30$  seconds to  $T_w = 40$  seconds were excluded from the analysis. These windows were omitted because during this time frame, the oscillations were entering the window, and the true DR values may not have been detected accurately. Similarly, observations from  $T_w = 80$  seconds onwards were not considered due to fluctuations in the signal. These observations were excluded to maintain the data quality and focus on the reliable measurements within the specified time range.

By categorizing the observations into true negatives and true positives, the evaluation of the SEWT algorithm's performance can be conducted. This classification provides insights into the algorithm's ability to accurately detect the presence or absence of oscillatory behaviour and estimate the corresponding modal parameters (NF and DR).

Fig. 6.22 shows the NF measurement for the same voltage signal but with SNR of 80 dB, 60 dB, 40 dB, 30 dB, and 20 dB. It can be noticed from Fig. 6.22 that due to the presence of noise, the NF is in the range of 0.8 Hz to 1 Hz under the normal operating conditions. Whereas, without the noise interference, it should be zero thus it is marked as a false positive. However, when the disturbance is introduced at  $t=30$  seconds the NF converges to 0.54 Hz. The red line shows

Table 6.3: Comparison of Damping Ratio and Damped Natural Frequency Metrics with Various SNR Levels

Metrics	Damping Ratio			Damped Natural Frequency			Actual
	SNR 60	SNR 40	SNR 30	SNR 60	SNR 40	SNR 30	No Noise
<b>True Negative</b>	22	0	21	2	18	0	30
<b>False Positive</b>	8	30	9	28	12	30	0
<b>False Negative</b>	2	9	0	1	7	19	0
<b>True Positive</b>	38	31	40	39	33	21	40
<b>Total</b>	70						

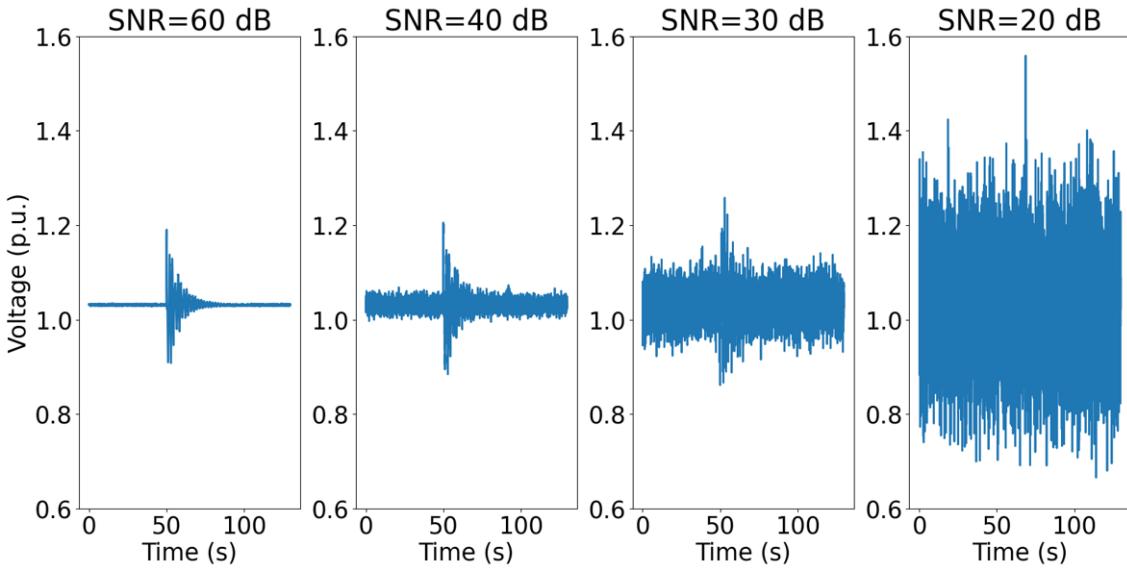


Figure 6.23: Case 1: Estimates of average average frequency with different SNR

the confidence interval and the black line represent the mean of the measured signals. It can be noticed that when the noise level is low ( $\text{SNR} = 80 \text{ dB}$ ) the probability of detection  $P_d$  is high compared to a high noise level ( $\text{SNR} = 30 \text{ dB}$ ). When the SNR is 20 dB, the noise fully takes control of the signal and no event is detected unless some noise-reduction technique is launched.

The metrics resulting from three different SNRs are presented in Table-6.3. As shown in the table, increasing SNR leads to an increase in false positive due to noise randomness. However, detecting an oscillation requires consistent characteristics in a mode of interest, particularly its oscillation frequency, over a time window. If the frequency changes significantly over three or four consecutive detection windows, it indicates that the previous oscillation mode has disappeared or merged with the current mode. In either case, triggering an oscillation alarm is not necessary, as it would result in numerous false alarms and significantly impact system performance.

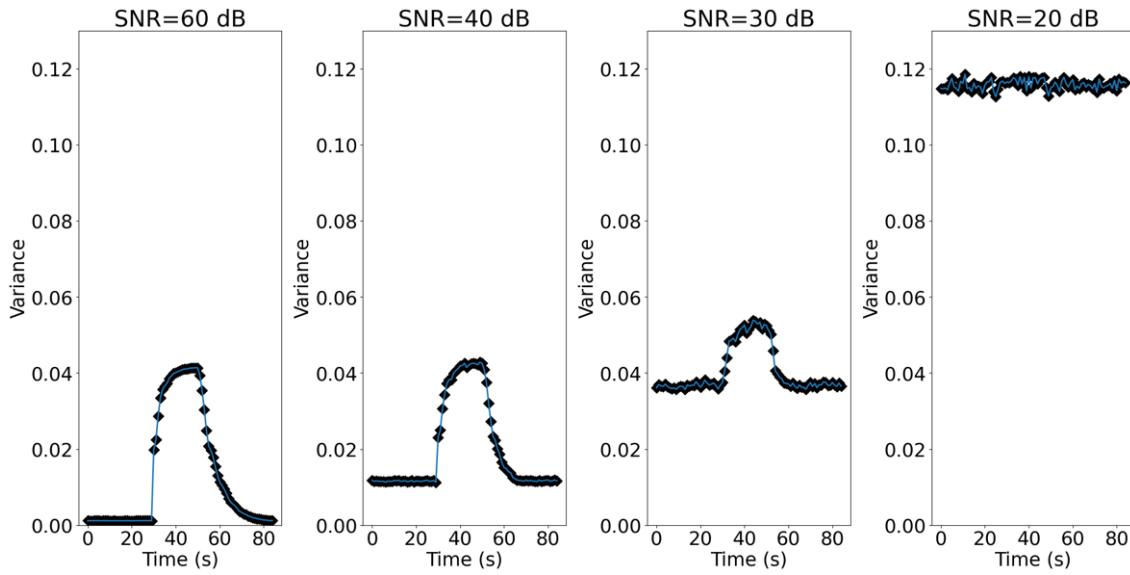


Figure 6.24: Case 1: Estimates of average average frequency with different SNR

To effectively reduce false alarms in the proposed algorithm, the variance of the voltage signal for each time window is compared to a pre-computed variance threshold during normal conditions. If the variance of a time window is above the pre-computed threshold, it is assumed that a disturbance is present in addition to noise interference, and a detection or "hit" is declared. The proposed algorithm determines whether the energy and variance in each received signal sample are too large to have likely resulted from interference alone, and assumes that a disturbance has contributed to that sample. The concept of window variance is illustrated in Fig. 6.23 and Fig. 6.24, where the top panel shows the transient voltage signal with increasing noise level, and the bottom panel shows the change in signal variance. The algorithm can accurately identify the NF and DR up to a noise level of 30 dB, beyond which the signal variance increases significantly, and the modal parameters cannot be identified correctly.

## 6.6 Performance evaluation and comparison

The performance of the proposed algorithm is assessed using the Mean Squared Error (MSE). The estimates of average DR and average frequency for each sliding window are compared with the BM values from the small-signal analysis. The MSE is found by averaging the squared error over the sampled window length. Table 6.4 provides the results of the MSE for SEWT compared to EWT. SEWT PC1 and SEWT PC2 indicate the results obtained from the PC1 and PC2, respectively. Correlation analysis of the PCs suggest that PC1 is positively correlated with all the twenty generators except g4. Thus, there is a direct relation between all the voltage

Table 6.4: Mean squared error

<b>Method</b>	<b>Mode</b>	<b>C1 MSE</b>		<b>C2 MSE</b>	
		<b>FREQ</b>	<b>DR</b>	<b>FREQ</b>	<b>DR</b>
<b>EWT</b>	Mode-2	0.024	1.041	0.053	0.905
	Mode-3	0.088	2.447	0.007	0.551
	Mode-4	-	-	0.013	3.168
<b>SEWT PC1</b>	Mode-2	0.001	0.144	0.06	0.092
	Mode-3	0.01	1.059	0.001	0.535
	Mode-4	-	-	0.001	0.15
<b>SEWT PC2</b>	Mode-2	0.009	1.269	0.001	0.318
	Mode-3	0.02	2.126	0.006	0.857
	Mode-4	-	-	0.147	0.86

buses and all of them participate in the oscillations. PC2 has large negative associations with busses g1, g2, g19, and g20, therefore the larger part of variations are coming from them. Further analysis of the voltage variables indicates that the negatively correlated generators mainly contribute to the strong 0.54 Hz mode while the rest of generators contribute to the 0.54 Hz mode and a weak 0.77 Hz mode. This mode is vaguely visible in the PC2 as reflected by the high MSE of the DR  $\zeta_{avg}$  in Table 6.4.

## 6.7 Conclusion

This thesis describes the real-time application of a proposed EWT-based technique for estimating the modal frequency and DR. The algorithm is highly adaptive to the signal's oscillation characteristics. The proposed SEWT method automatically estimates the number of modes based on the frequency contents of the signal. The window-based automatic mode detection method locates the local maxima, and the use of amplitude threshold avoids any unnecessary segmentation of the Fourier spectrum. Unlike conventional single-channel methods, the proposed algorithm is based on multi-variate data analysis and captures the global dynamic features. In addition, through data compression, it is effective in reducing noise errors. Test result evaluation and comparison with existing methods reveal that the proposed method shows excellent potential for real-time monitoring and identification of inter-area oscillations.

## Chapter 7

# Estimating interarea oscillation with machine learning

In the previous chapter, we explained how the Natural Frequency (NF) and Damping Ratio (DR) were extracted from voltage signals using the Sliding Window based Empirical Wavelet Transform (SEWT) technique. In this chapter, we extend our analysis by incorporating machine learning (ML) algorithms to estimate the NF and DR from the voltage signals.

We measured the voltage magnitude from the Nordic Test system and employed the LASSO method to select relevant features. The selected voltage magnitude from LASSO serves as inputs for the ML algorithm, which predicts the NF and DR as target variables. This study introduces a multi-channel ML model for identifying inter-area oscillation modes in power systems, achieving online and real-time mode identification. This multichannel approach enables analyzing data from multiple nodes, accurately and identify buses experiencing voltage instability due to oscillation.

By combining SEWT and ML techniques, our goal is to precisely estimate the NF and DR from voltage signals, leveraging SEWT's strengths in signal decomposition and feature extraction, and ML's effectiveness in prediction tasks. The proposed approach offers improved accuracy and reliability in identifying dominant inter-area modes, beneficial for power system operators in maintaining grid security.

The chapter is organized as follows. Section 7.1 gives an introduction. Section 7.2 presents the overall algorithm and methodology. Section 7.3 provides numerical results to verify the proposed

methods' performance, followed by conclusions in Section 7.4.

## 7.1 Introduction

Recent research works have focused on merging measurement-based techniques with machine learning (ML) frameworks. Nonetheless, most existing methods can only process one signal or they analyze multiple signals separately. These methods usually require measurements from critical devices, e.g., bus frequency, which has good observability for the specific inter-area oscillation modes. If the measurement that has good oscillation observability is not available in some areas, single-channel methods may not be able to provide oscillation information in these areas. As a result, a multi-channel approach becomes necessary to extract complex relationships between multiple feature/target variables. Multi-channel/multi-variate methods can capture the underlying relationships between features and targets. Moreover, adopting multi-target/multi-variate approaches could lead to simpler models with improved computational efficiency.

In dealing with these challenges, this study aims to introduce a multi-channel machine learning model (MLM) to identify inter-area oscillation modes. This approach incorporates real-time mode identification through the integration of MLM with adaptive signal processing tools like EWT and HT. The key contributions of this chapter can be summarized as follows:

- Utilizing SEWT for feature extraction and employing the extracted features as target variables.
- Employing a multivariate MLM approach, utilizing features selected (voltage magnitude) via Lasso, and associating them with the target variable (NF and DR).
- Real-time implementation of the MLM method to estimate the NF and DR.

### 7.1.1 Fundamentals and proposed algorithm

ML techniques approximate the unknown relationship between the multiple input voltage magnitudes and the output, in our case, instantaneous damping ratio (IDR) and instantaneous frequency (IF). Three disturbance scenarios are considered, 1) stable oscillations, 2) unstable oscillations, and 3) sustained oscillations. Their definitions are briefly given below:

**Stable Oscillations:** Oscillations have been observed in power systems from the moment networks comprising multiple generators were interconnected to provide more power capacity

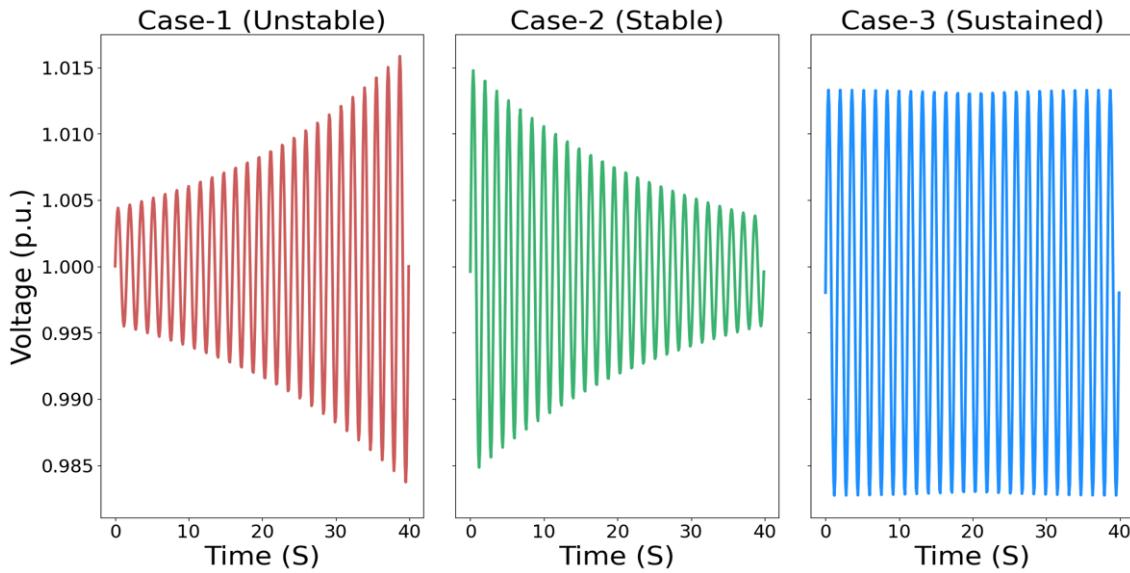


Figure 7.1: Three different operating scenarios. Case-1 represents the unstable operating conditions, Case-2 represents the decreasing oscillations signifying the system is approaching stability

and reliability. It is customary to operate the system or design the system controls so that each mode decays by at least  $\exp(-0.1\pi)$  during every oscillation cycle. The system is considered sufficiently stable at that operating point if all modes have real parts more negative than the limit.

**Unstable Oscillations:** When a power system is pressed to provide additional load, unstable oscillations often arise. As transmission lines become more heavily loaded, generators must depend more on their excitation systems to maintain synchronism. However, without supplementary control, the synchronizing oscillations may eventually become unstable. These oscillations are identified by modes with positive real parts.

**Sustained Oscillations:** Slowly damped or sustained oscillations in power systems can damage equipment, reduce the power transfer limit, and even result in cascading blackouts. Several mechanisms can give rise to these oscillations, including improper operating conditions, periodic disturbances, or malfunctioning controllers.

## 7.2 Methodology

The development of the proposed oscillatory assessment system consists of several steps, including a) generating training data, b) extracting the features, c) selecting the important features,

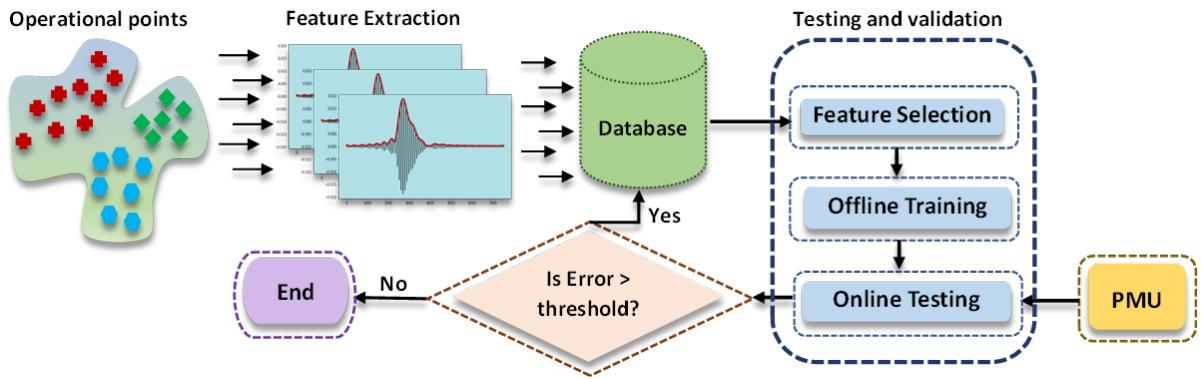


Figure 7.2: Flow diagram of the proposed methodology.

d) training and evaluation of the ML algorithm, and e) addressing real-time implementation issues. These steps are shown in Fig. 7.2 and described in the following sections.

### 7.2.1 Data Generation for machine learning

The data generation process details have been provided in Section 2.4. Operating point A of the Nordic test system [188]. is used to analyse oscillatory behaviour. For our purpose, a sufficiently but realistically extensive range of contingencies and operating scenarios are considered to ensure the generality of the predictive model and incorporate uncertainty.

The data generation process begins by extracting a comprehensive set of operating points from the historical electricity consumption load profile dataset, as referenced in Chapter 3, [134]. The scenarios are constructed through a DBSCAN clustering technique.. However, prior to clustering, the load profiles are first divided based on the seasons of the year, as depicted in Fig. 7.3. This stratification ensures that each season is represented and reduces bias by organizing the load profiles into homogeneous subgroups.

The clusters obtained from the stratified groups are validated using the Silhouette index method. This validation step ensures that the resulting clusters accurately represent the characteristics of the load profiles and further enhances the reliability of the data generation process.

Based on the extracted load operating points, datasets are generated for three cases mentioned in section 7.1.1. These cases represent unstable, stable, and sustained oscillation scenarios. The output/target variables comprising instantaneous frequency (IF) and instantaneous damping ratio (IDR) are generated through the feature extraction process described in the next section.

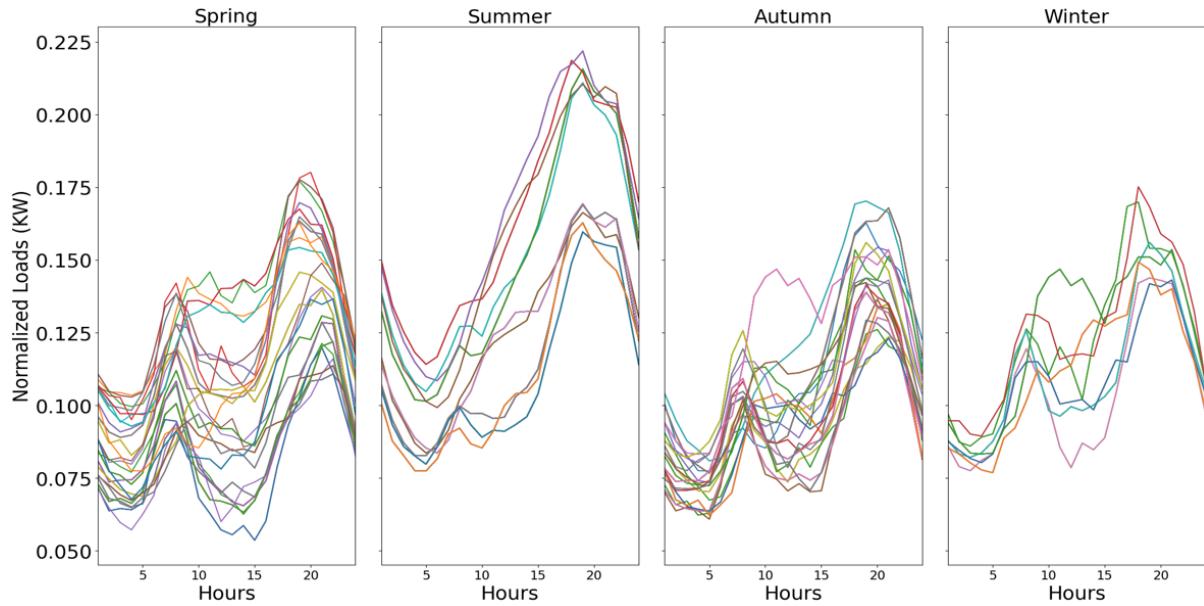


Figure 7.3: Stratification of load profiles based on seasons of the year

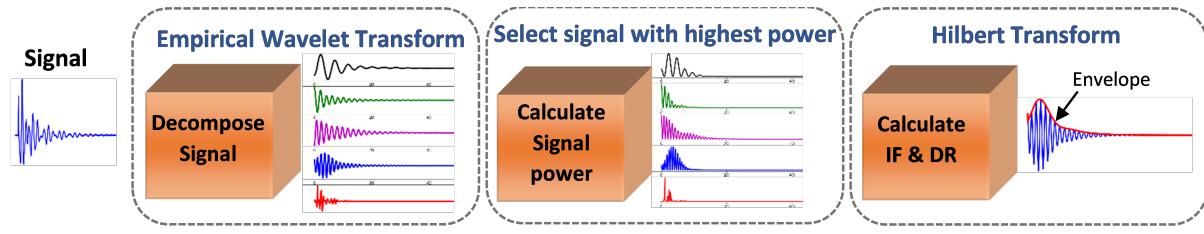


Figure 7.4: Feature extraction process.

### Feature Extraction using EWT and HT

The purpose of using a feature extraction technique is to obtain a subset of features from the original feature set while retaining as much information in the data as possible through a functional mapping. Power system oscillation signals consist of multiple modes, and due to the presence of noise, it can be challenging to determine the number of segments in advance. This study proposes a technique called Sliding Window based Empirical Wavelet Transform (SEWT) for mode separation to segment the Fourier spectrum  $X_\omega$ , as explained in section 6.3.4. The goal is to extract features using SEWT and HT to classify the signal in terms of IF and IDR. The entire process is illustrated in Fig. 7.4. First, SEWT decomposes the frequency spectrum of the signal into an empirical mono-component, and the signal with the highest energy level is identified as the critical mode. HT is then used to track the amplitude envelope of the critical signal, from which the IF and DR can be calculated.

### 7.2.2 Feature Selection with LASSO

Building a helpful feature set with a lower dimension for ML is significant. Moreover, the redundancy of features is also an inevitable problem. Thus, in this work, a least absolute shrinkage and selection operator (LASSO)-based [125] feature selection framework is proposed to achieve efficient feature selection and to decrease the features redundancy. There are many advantages to using the LASSO method; first, it can provide outstanding prediction accuracy because shrinking and removing the coefficients can reduce variance without a substantial increase in the bias. The regularization process updates the coefficient values of the regression variables by reducing few to zero, meaning that it can nullify the impact of irrelevant features in the data [126].

### 7.2.3 Training and evaluation

This study uses the data from PMU measurements to conduct an ML-based multi-variate analysis. Using a ML, multiple target variables are predicted simultaneously.

In multivariate analysis, given the dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  of  $n$  training examples,  $x = [x_1, x_2, \dots, x_d]$  is the input attribute vector consisting of the voltage magnitudes at the generator buses for the  $i$ -th operating point of the  $n$  samples. The output variable  $y$  consists of multiple  $m$  target  $y_1, y_2, \dots, y_m$  and is the associated target values of instantaneous parameters (IF and IDR) that correspond to the  $m$  size of the training data.

The goal in multi-input multi-output (MIMO) regression is to learn a model  $h : x \rightarrow y$ . For a given input vector  $x$ , the model can predict an output vector  $\hat{y} = h(x)$  that best approximates the actual output vector  $y$ . After the training phase, the constructed multi-target model will be used to simultaneously predict  $m$  target variables based on  $d$  input variables of the new incoming instance.

To put it briefly, if  $m$  is the number of target attributes, single target regression considers domains where  $m = 1$ , while MIMO considers domains with  $m \geq 2$ .

### Decision Tree

A DT is a graphical representation of possible solutions to a decision based on certain conditions. It is a tree-like model of decisions and their possible consequences. In a DT, an internal node represents a decision, and the branches represent the possible outcomes of that decision. Each

leaf node represents a final decision or outcome. DTs are useful because they allow you to visually and explicitly represent decisions and their potential consequences, and they can help make decisions under uncertainty.

There are many different algorithms for constructing DTs, such as ID3, C4.5, and CART (Classification and Regression Trees). These algorithms use training data to learn the relationships between different features in the data and the target variable. They use this information to create a tree that can be used to make predictions.

In this study, CART is used for building DT models as they can be used both for classification and regression tasks. They are relatively fast to train, especially compared to more complex algorithms such as neural networks.

### **Random Forest Regressor**

Random Forest algorithm is used for both classification and regression in ML. It is an ensemble method, which means that it is based on the combination of several models to make a final prediction. In a Random Forest model, a large number of decision trees (DT) are trained on a randomly selected subset of the training data. Each DT is trained to make predictions independently, and the final prediction is made by taking the majority vote of all the DTs. This process helps to reduce overfitting and improve the generalization of the model. Training samples are randomly selected with replacements. The size of each new training set is the same as the original dataset. As input parameters, the number of trees in the algorithm and maximum depth should be determined initially. The change in their values may affect the performance and predictive power of the algorithm. Therefore, all possible parameters in the dataset's size range are given to the method and tested. The parameters leading to the best results become candidates to be used.

### **K-Nearest Neighbours**

K-nearest neighbour (KNN) is a supervised ML algorithm for classification and regression. The idea behind KNN is to determine the class of a given data point by looking at the class of the data points that are closest to it based on a distance measure.

In KNN, the value of  $K$  determines the number of nearest neighbours that will be considered when making a prediction. For example, if  $K$  is set to 3, the algorithm will consider the 3

nearest data points and make a prediction based on the class of those data points.

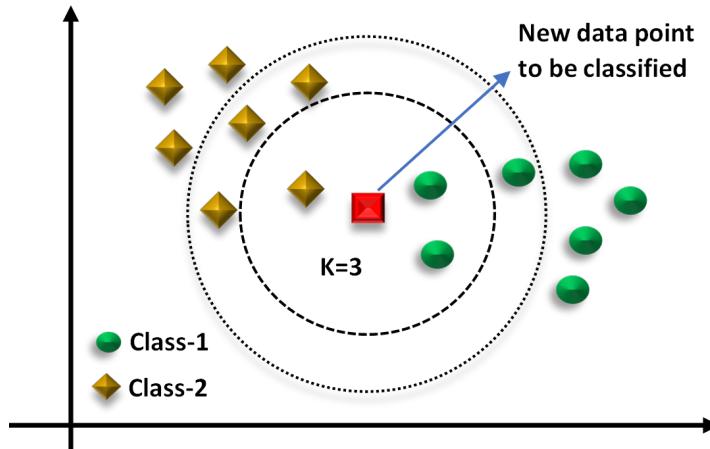


Figure 7.5: K-nearest-neighbor

To determine the class of a given data point, the KNN algorithm calculates the distance between that point and all other data points in the training set. It then selects the K data points closest to the given data point and uses the majority class among those K data points to make a prediction.

KNN is particularly useful when the training data is very large, as it is a non-parametric method and does not require any assumptions about the underlying distribution of the data. However, it can be computationally expensive to calculate the distance between a given data point and all other data points in the training set, especially for large datasets.

### Motivation for using machine learning models

The motivation for using Random Forest Regressor, Decision Tree, and KNN for estimating interarea oscillation, while extracting features such as damping ratio and natural frequency using empirical wavelet transform, can be attributed to several factors:

- Nonlinearity and Complex Relationships: Interarea oscillations in power systems often exhibit nonlinear behaviour and complex relationships among the variables involved. Traditional linear models may struggle to capture these intricate dynamics effectively. Decision Tree, Random Forest, and KNN models are capable of capturing nonlinear relationships and can handle complex interactions between variables.
- Flexibility and Adaptability: Decision Tree, Random Forest, and KNN models are highly flexible and adaptable to various data distributions and patterns. They do not impose

strong assumptions on the data structure and can automatically handle feature interactions and non-linearities. This makes them suitable for modelling complex interarea oscillation dynamics.

- Ensemble Learning and Generalization: Random Forest, as an ensemble method, combines multiple decision trees to improve prediction accuracy and reduce overfitting. It benefits from the wisdom of crowds and captures diverse aspects of the data. Similarly, KNN utilizes the nearest neighbors to make predictions, which can help in capturing local patterns. These ensemble and neighbor-based approaches can enhance the generalization ability of the models and provide robust estimation of interarea oscillation.
- Interpretability: Decision Tree-based models offer interpretability, as the decision rules can be easily understood and visualized. This can be valuable for gaining insights into the underlying dynamics of interarea oscillations and identifying important features and their relationships.

By utilizing these machine learning techniques in combination with empirical wavelet transform for feature extraction, we can leverage their strengths in handling complex relationships, non-linearities, and adaptability to estimate interarea oscillation more accurately and potentially gain insights into the underlying dynamics of the system.

#### 7.2.4 Online Assessment

After completing the training stage, the obtained ML is applied for the online assessment of low-frequency interarea oscillations. During the online monitoring stage, when a transient fault/event occurs in the system, multiple trajectories of  $V_{mag}$  are acquired from individual load buses by PMUs. These trajectories are fed into the ML model to monitor and predict the instantaneous frequency and damping ratio. If the system is expected to be unstable, alarm signals will be issued to warn that remedial actions should be taken as soon as possible. Otherwise, the algorithm will keep monitoring the system status.

#### 7.2.5 Database update condition

A feasible online estimation approach should be able to accommodate different unseen network topologies. When some unfavourable system change is incurred, operators will be sent a warning, and the database should be updated to improve the prediction. Notably, more operating

points relevant to the current one should be added to the database. We propose to generate the operating points pertinent to the existing ones by adding critical contingencies to the current operating point and deploying these newly added operating points to the online testing phase. The operational states are assumed to not change too fast concerning the near-real-time scale. If some contingency occurs, it is within the high probability/critical contingencies of current operating points. This study implements the database update scheme using a K-means clustering approach. Details about the k-means clustering can be found in section 4.1.

The steps for database update are provided below:

1. **Initial Clustering:** K-means clustering is initially applied to the training database. This involves partitioning the dataset into a specified number of clusters ( $K$ ) based on the similarity of data points.
2. **Distance Calculation:** For each data point in the training database, the distances between that data point and the cluster centers are computed. The cluster center is essentially the mean or centroid of the data points within a cluster.
3. **Threshold Ratio:** A threshold ratio is set, which serves as a criterion to determine whether a newly arrived data point during online monitoring should be associated with an existing cluster or not.
4. **Online Monitoring:** As new data points arrive from the PMUs (Phasor Measurement Units), the process of clustering continues. For each new data point, its distance to the cluster centers is recalculated.
5. **Threshold Check:** If the calculated distance between a newly arrived data point and the cluster center is within the specified threshold (i.e., the distance is relatively small), it implies that the new data point is similar enough to the existing cluster. This helps maintain the existing clusters and does not significantly impact the prediction accuracy.
6. **Outside Threshold:** On the other hand, if the distance between the newly arrived data point and the cluster center exceeds or falls below the threshold value, it suggests that the new data point is significantly different from the existing clusters. In such cases, the new data point is not merged with any existing cluster. Instead, the operating point represented by the new data point is added back to the dataset, and the training process is restarted offline. This helps in capturing and adapting to new patterns or changes in

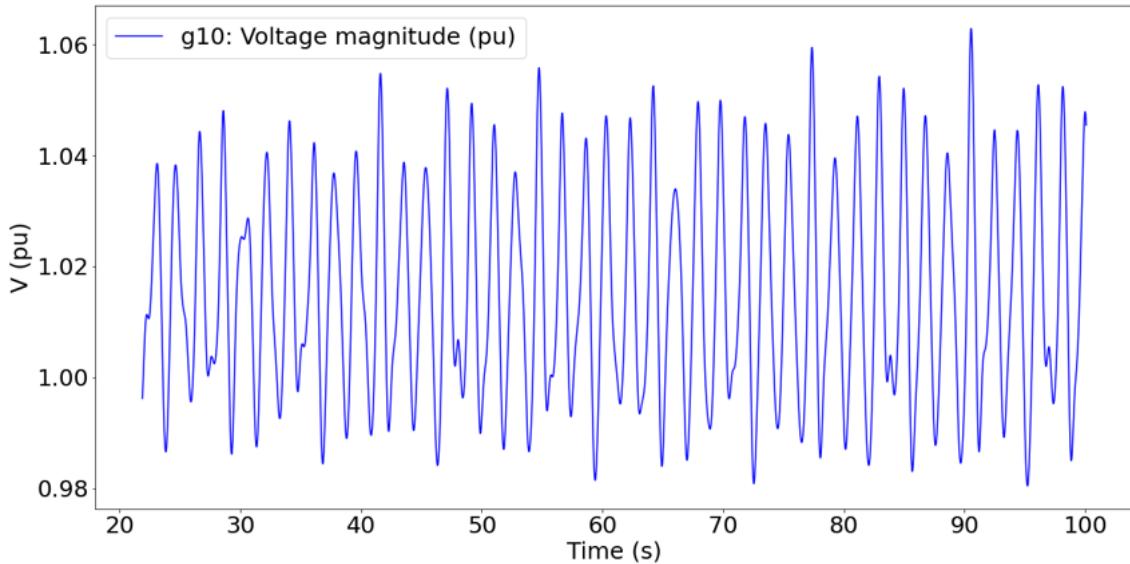


Figure 7.6: C-1: Voltage signal from generator bus g10 with unstable oscillations.

the data that are not well-represented by the existing clusters.

The threshold-based approach allows for the adaptation of the clustering as the data distribution changes over time, ensuring the accuracy of predictions and detecting significant deviations that require dataset updates and retraining.

## 7.3 Numerical results

### 7.3.1 System model and case study description

The Nordic test system proposed by the IEEE Task Force for Voltage stability and security assessment is used to demonstrate the dependability and security of the proposed ensemble method. The detailed data of the test system and the operating points can be found in [135]. A one-line diagram can be seen in Fig. A.1. The system has rather long transmission lines with 400-kV nominal voltage. The model also represents some regional systems operating at 220 and 130 kV, respectively. All 20 generators (19 synchronous generators and one condenser) are represented behind their step-up transformers. The LTCs of the step-down transformers control all 22 loads at the distribution level. This system is made up of the following four areas:

- **North** with hydro generation and some load.
- **Central** with much higher load and thermal power generation.
- **Equiv** is connected to the **North**, which includes a simple equivalent of an external

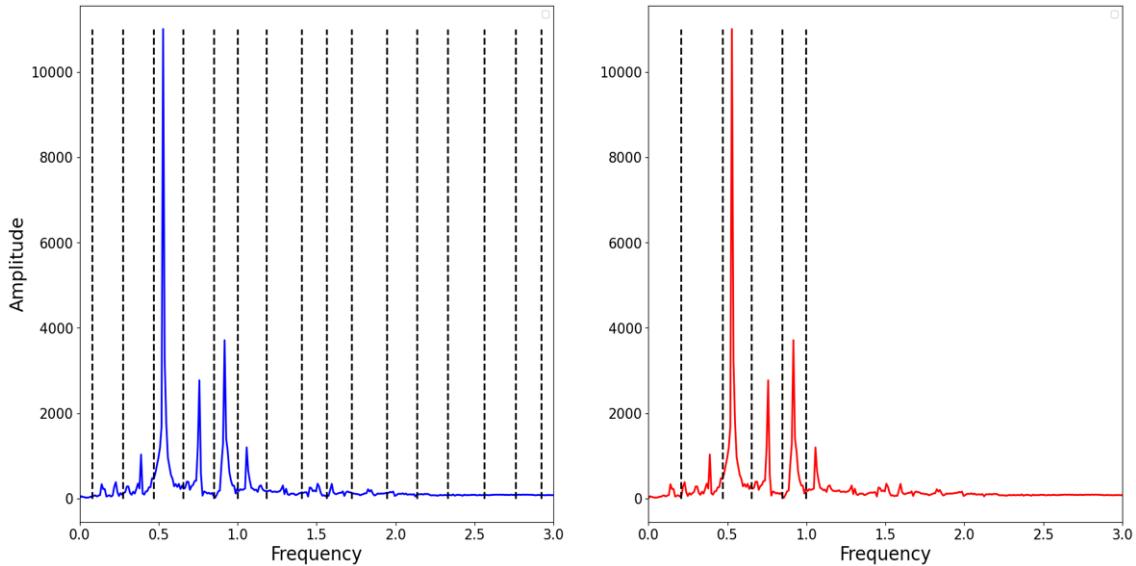


Figure 7.7: C-1: Fourier spectrum with detected boundaries.

Table 7.1: Small-signal stability analysis of the Nordic system.

	C-1		C-2			C-3	
	Mode2	Mode3	Mode2	Mode3	Mode4	Mode2	Mode3
Frequency(Hz)	0.538	0.742	0.533	0.766	0.949	0.533	0.762
DR (%)	3.7	3.9	-0.4	-0.1	0.3	0.0	0.0

system.

- **South** with the thermal generation, loosely connected to the rest of the system.

### 7.3.2 Data Preparation

A large number of training data were simulated based on the Nordic Test system. The simulations were performed for 70 system operating points, including diverse generation patterns and planned transmission outages. Time-domain simulations were conducted using the python-based software PyRAMSES [120], and the results were saved in a database, emulating the PMU measurements.

Three scenarios were investigated a) **Stable oscillations model** (C-1) b) **Unstable oscillations model** (C-2) and **Sustained oscillations** (C-3). For C-2, several generator PSS controllers were deactivated to destabilize the system. In both cases, the modes are excited by applying a three-phase fault at bus 4072 at time  $t = 50$  seconds. The fault is automatically cleared after 5 cycles (100 ms). All the measurements are sampled at 1 cycle (20ms).

We performed a small-signal stability analysis on the original differential-algebraic model used for the above two cases and identified the main electromechanical modes in Table 7.1. These

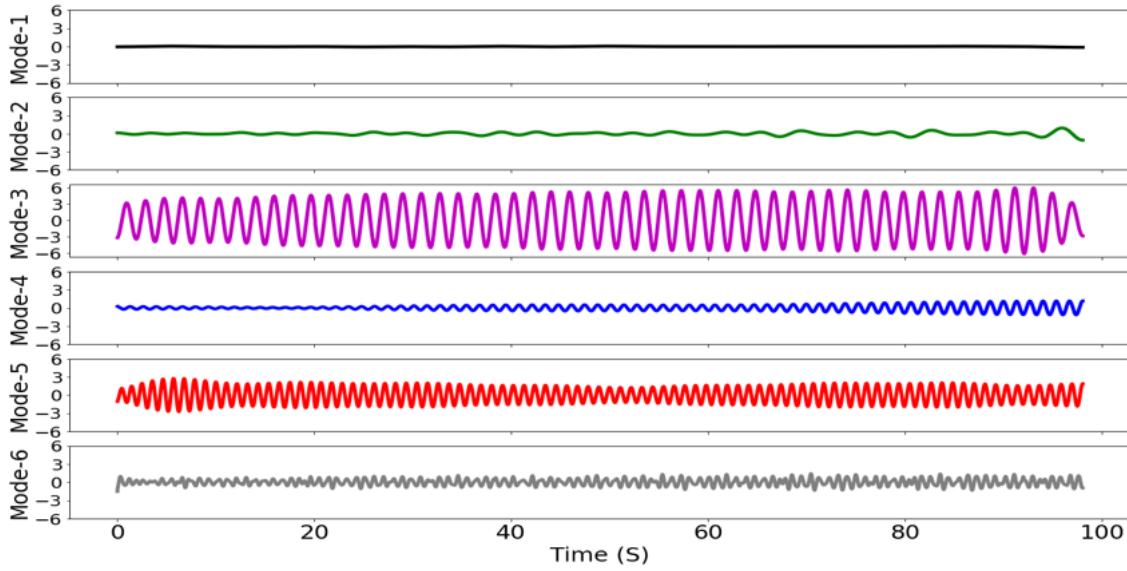


Figure 7.8: C-1: Extracted mono-components of the non-stationary voltage signal.

are used as a benchmark (BM) for our real-time detection algorithm that only uses the data flow. Only voltage magnitude measurements are used from the 20 synchronous generators. The number of samples and features ranged from 1484 to 67557.

### Feature Extraction

The SEWT method was experimentally applied to the voltage signals. Fig. 7.6 shows the temporal voltage signal of the increasing magnitude acquired through simulation. The same procedure described in chapter 6 is used to decompose each signal. Fig. 7.7 shows the FFT spectrum with and without threshold ( $\tau$ ) levels. It can be noticed from Fig. 7.7-(a) that the spectrum with the sliding window approach over the Fourier spectrum but without a threshold shows unnecessary segmentation. Fig. 7.7-(b) represents the spectrum with a threshold level of one standard deviation from the maximum peak and automatically determines the number of segments. According to this spectrum, globally, six peaks can be distinguished, which can be used to compute the support of the filter bank. The support boundaries of the filter bank are the midpoint between successive frequencies associated with these peaks. The black dotted lines in Fig. 7.7 show the estimated boundaries to create the wavelet filter bank.

### Selection of Empirical Mode Functions

After performing SEWT on a voltage signal, we obtain a series of mono-components, each with a single frequency corresponding to one modal response. We call these mono-components Empirical Mode Functions (EMF). Fig. 7.8 shows the EMFs for the voltage signal of increasing

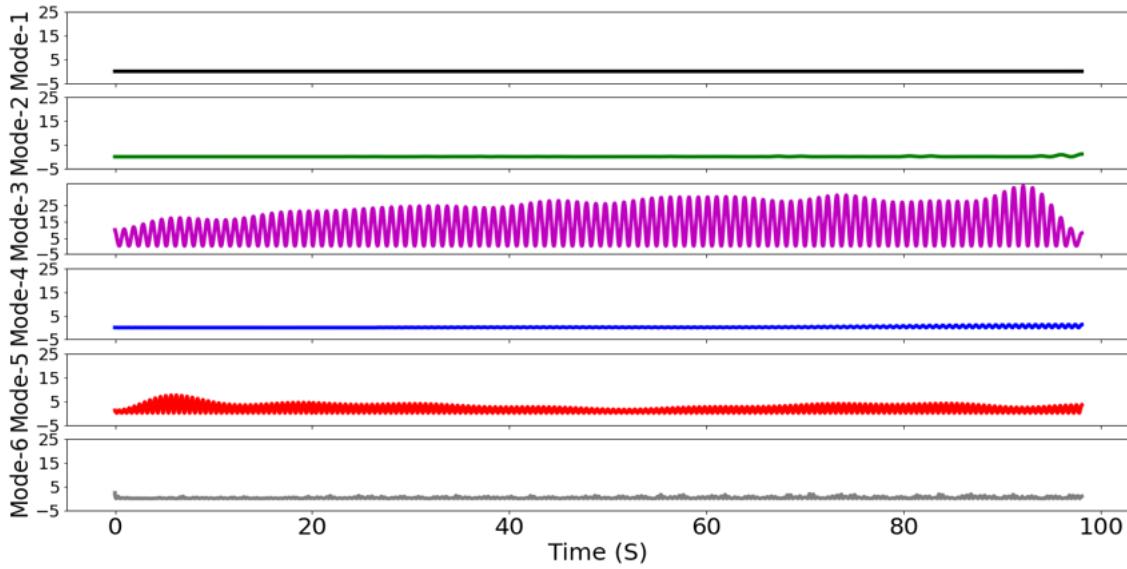


Figure 7.9: C-1: Signal power of extracted mono-components.

magnitude for generator bus g10 and labelled as Mode 1,2,3,4,5 and 6. It can be seen that the SEWT technique provides accurate estimates of the modes present in the signal and decomposes the input signal into different frequency components concerning time. The temporal EMFs in Fig. 7.8 are sorted in increasing frequency order. It can be observed from Fig. 7.8 that the amplitude values of the EMFs do not contribute equally to the analysis of interarea oscillations. Some EMFs are closely related to the original voltage signal, but others contain little information. Thus, in this example, we can observe that EMFs 3, 4, and 5 carry essential information, and their behaviour in the time domain can be associated with critical inter-area modes. Moreover, these EMFs have the highest amplitude values among the extracted EMFs. It can also be seen that EMF-6 contains high-frequency noise.

To construct a more precise Hilbert-spectrum that identifies an oscillation mode with disturbance signatures, it is crucial to choose sensitive EMFs and exclude irrelevant ones by identifying EMFs with higher power. Fig. 7.9 displays the signal's energy over time. Notably, EMF-3 exhibits the highest amplitude (0-20), making it the potential critical mode, while the remaining modes are deemed unimportant and thus discarded.

### Estimation of instantaneous parameters

For all three cases above, C-1, C-2, and C-3, each selected EMF was applied with HT to estimate the instantaneous attributes of desired signals. A deeper insight into the nature of temporal behaviour is obtained from analyzing the instantaneous attributes of each EMF. The IA and

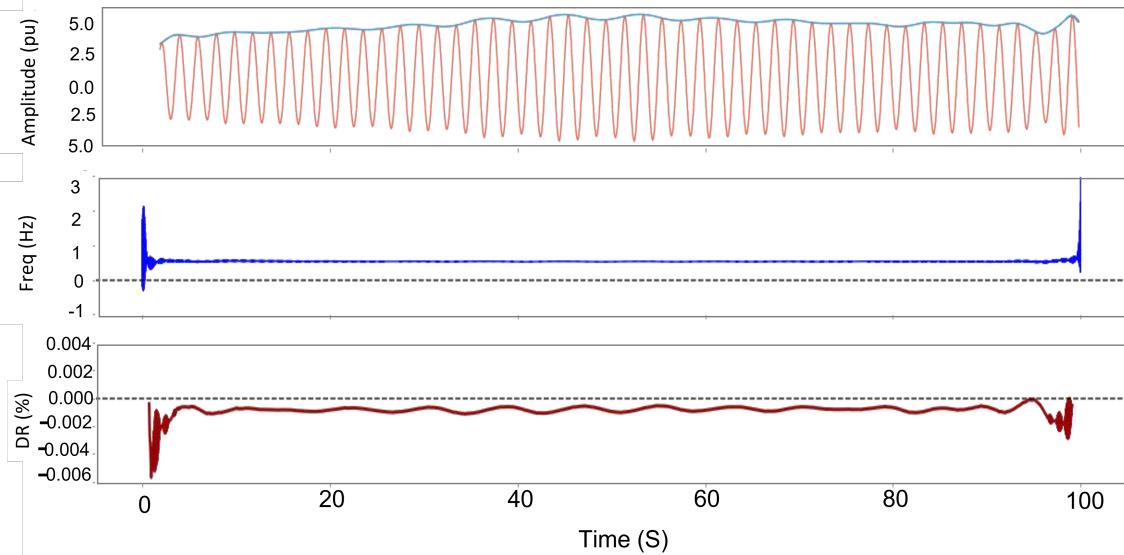


Figure 7.10: C-1: On the top Reconstructed signal with instantaneous amplitude envelop using SEWT, in the middle instantaneous frequency, the bottom panel shows instantaneous damping ratio.

IF of the voltage magnitude from the bus g-12 response are shown in Fig. 7.10. Simulation results show that the instantaneous attributes of the generator bus response to disturbance exhibit increasing amplitudes, with an average IF of 0.53 Hz, thus indicating the presence of low-frequency interarea oscillations. Analysis of the IDR suggests that the proposed approach can determine the precise time evolution of the relevant system parameters, thus leading to an improved assessment of system behaviour.

### 7.3.3 Comparative Performance Analysis: Empirical Wavelet Transform vs. other adaptive methods

In this section, we compare the performance of EWT with other adaptive signal decomposition methods. We evaluate the DR, NF and root mean square error (RMSE) for each method. Our objective is to assess how EWT performs compared to other modal decomposition techniques.

The following methods are considered for comparison:

- 1. Empirical Mode Decomposition (EMD):** EMD [166] is a widely used mode decomposition method that extracts a set of Intrinsic Mode Functions (IMFs) from a signal. IMFs are functions with localized oscillatory behavior and are obtained by iteratively identifying local extrema in the signal. EMD is versatile and can be applied to various types of data without requiring predefined filters or assumptions.

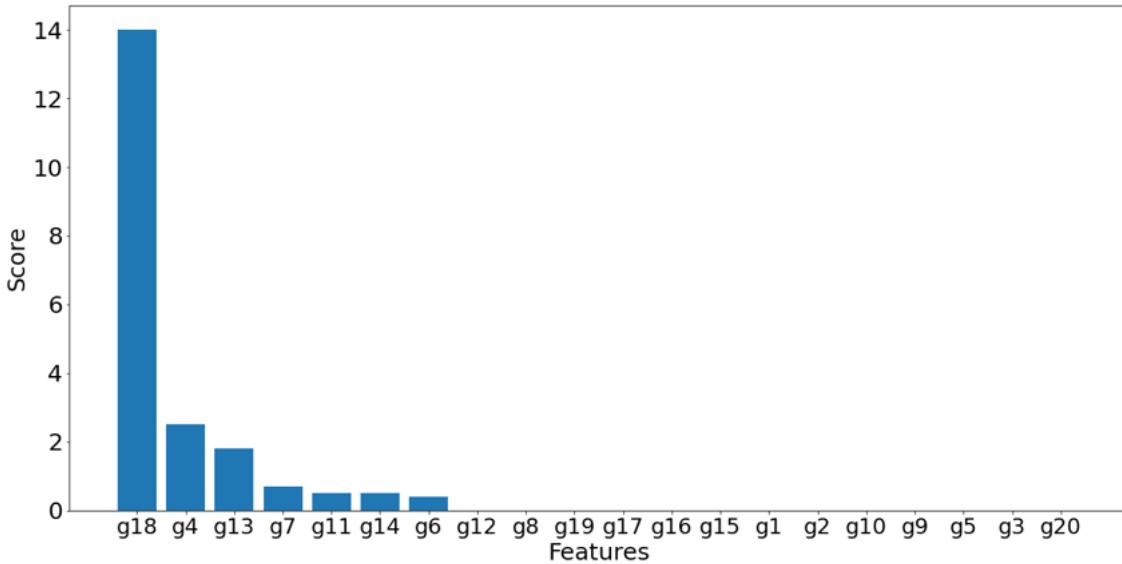


Figure 7.11: Selected features for case study C-1, C-2, and C-3

2. **Variational Mode Decomposition (VMD):** VMD [189] is an adaptive signal decomposition approach capable of decomposing a multi-component signal into several quasi-orthogonal intrinsic mode functions in a non-recursive manner. VMD aims to identify a set of modes, each with its corresponding centre frequency, in order to collectively replicate the input signal while ensuring that each mode exhibits smooth behaviour after demodulation into baseband. The key advantage of VMD is its ability to eliminate residual noise in the modes and reduce redundant modes compared to other techniques such as EMD.

3. **Local Mean Decomposition (LMD):** LMD [190] is an adaptive method for decomposing complex signals into a sequence of mono-components known as product functions. These PFs are formed by combining a frequency-modulated signal with an envelope signal. The fundamental principle of LMD involves iteratively applying moving average operations to gradually smoothen the signal. This process enables the estimation of the IF associated with each product function. By employing LMD, a complex signal can be effectively disassembled into multiple product functions, facilitating a detailed analysis of its frequency characteristics.

The performance of these methods are summarized in Table 7.2, and Table 7.3 respectively.

Detail discussion is provided below:

- On comparison of RMSE, EWT appears to be the most consistent and effective method across stability categories and modes for both DR and NF. VMD also perform well, while LMD and EMD has higher RMSE values compared to EWT.

Table 7.2: RMSE FREQUENCY

	RMSE					
	Stable		Unstable		Sustained	
Method	Mode-1	Mode-2	Mode-1	Mode-2	Mode-1	Mode-2
<b>EWT</b>	0.0194	0.0634	0.0163	0.0531	0.0219	0.1315
<b>VAR</b>	0.0210	0.3946	0.0206	0.1759	0.0244	0.3220
<b>LMD</b>	0.0222	0.3656	0.1617	0.0504	0.0280	0.3535
<b>EMD</b>	0.0208	0.4765	0.2205	0.0687	0.0236	0.2466

Table 7.3: RMSE DAMPING RATIO

	RMSE					
	Stable		Unstable		Sustained	
Method	Mode-1	Mode-2	Mode-1	Mode-2	Mode-1	Mode-2
<b>EWT</b>	0.415	0.938	0.449	0.337	0.150	0.545
<b>VMD</b>	0.793	2.655	0.443	0.709	0.399	0.652
<b>LMD</b>	1.849	1.283	1.445	0.632	0.621	1.955
<b>EMD</b>	2.216	4.316	2.873	0.910	0.203	0.892

- Fig. 7.12 and Fig. 7.12 depict the evaluation of average Damping Ratio (DR) and average Natural Frequency (NF) using EWT, VMD, LMD, and EMD. Among the various methodologies, the EWT stands out as the top-performing approach, particularly in the realm of ratio estimation. It adeptly approximates the ratios for both mode-1 (3.7%) and mode-2 (3.9%). It becomes evident that the EWT's estimation of the average DR closely aligns with the baseline ratio, whereas the other methods encounter challenges. EWT exhibits superior performance in frequency estimation compared to ratio estimation. This observation is supported by Fig. 7.13, which illustrates EWT's consistent accuracy in identifying the NF of Mode-2 for the unstable case. In contrast, VMD's accuracy diminishes beyond a certain point, while LMD and EMD exhibit substantial deviations.

A deeper understanding of the comparative analysis can be gleaned through mode separation process. This concept is demonstrated in Fig. 7.14, showcasing the mode separation accomplished by EWT for the unstable mode. The decomposition of the signal is strikingly accurate. The Fourier Frequency graph on the right vividly exhibits the distinct segregation of modes at 0.54 Hz, 0.77 Hz, and 0.9 Hz. Notably, the reconstruction of the signal with the 0.54 Hz mode faithfully reproduces the original signal, facilitating the accurate construction of the envelope through the Hilbert Transform. This precision ultimately contributes to the accurate estimation of the DR.

- Variation Mode Decomposition (VMD) demonstrates relatively good performance, espe-

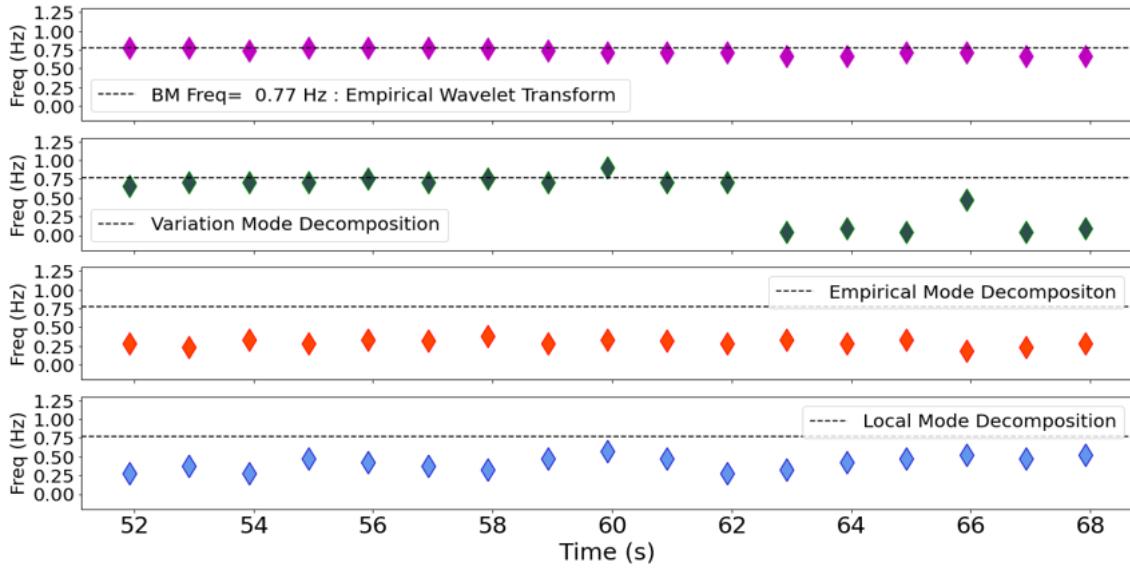


Figure 7.12: Case 2: Left)Decomposition of the unstable signal using EWT, Right) Fourier frequency of the decomposed modes.

cially in the sustained scenario where the results are similar across frequencies. However, VMD its ratio estimation may not be as accurate as EWT. Fig. 7.15 displays the decomposition of the unstable mode using VMD, showing clear identification of the modes at 0.54 Hz and 0.9 Hz, but it fails to detect the 0.77 Hz mode accurately.

- Local Mean Decomposition (LMD) performs relatively poorly compared to the other methods. LMD may suffer from mode mixing, resulting in inadequate separation of modes and less accurate ratio estimation. Although LMD's performance is similar to VMD in the sustained scenario, it may not effectively capture dynamic changes. Fig. 7.16 illustrates the decomposition performed by LMD for the same unstable mode, highlighting the struggle with mode mixing issues and its impact on the accuracy of the DR.
- Similar to LMD, Empirical Mode Decomposition (EMD) does not exhibit strong performance compared to EWT and VMD. EMD may also struggle with mode mixing, leading to incomplete separation of modes and less accurate ratio estimation. Fig. 7.17 showcases the decomposition conducted by EMD for the same unstable mode, revealing similar difficulties with mode mixing and incomplete mode separation.

In summary, EWT has performed the best among the methods, followed by VMD. EWT shows good accuracy in ratio estimation and its frequency results are better overall. LMD and EMD have shown weaker performance, possibly due to mode mixing issues and incomplete separation of modes.

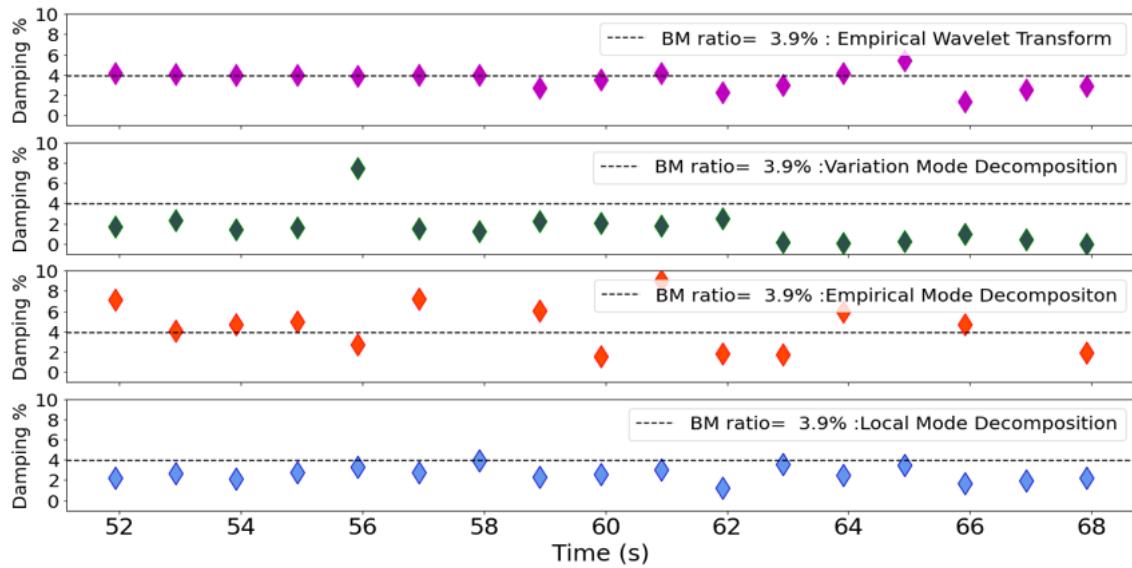


Figure 7.13: Case 1 Stable Scenario:Comparison of average damping ratio estimation by different methods.

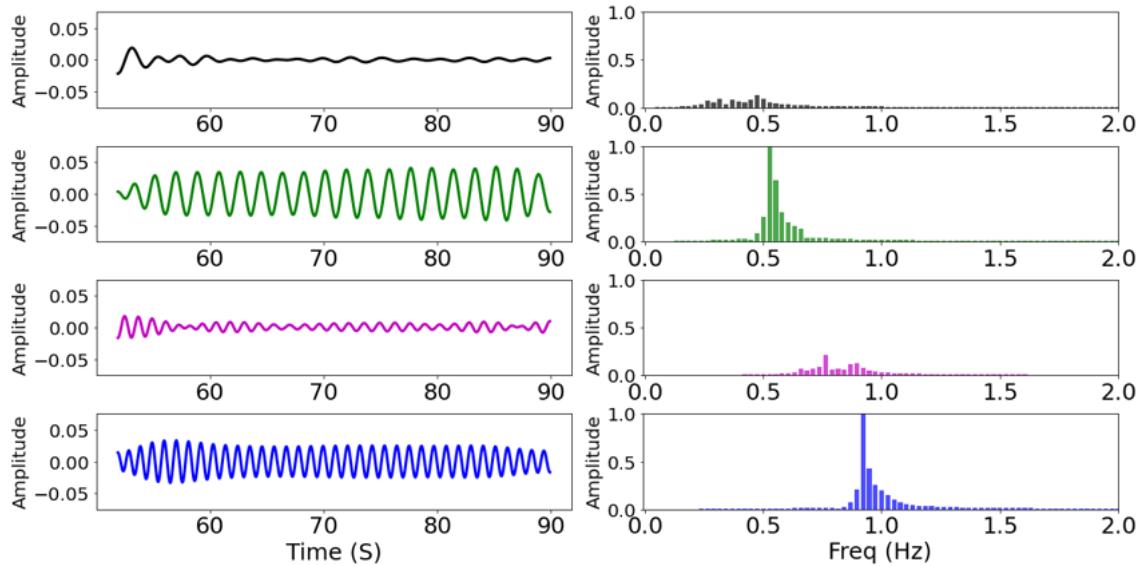


Figure 7.14: Case 2: Left)Decomposition of the unstable signal using EWT, Right) Fourier frequency of the decomposed modes.

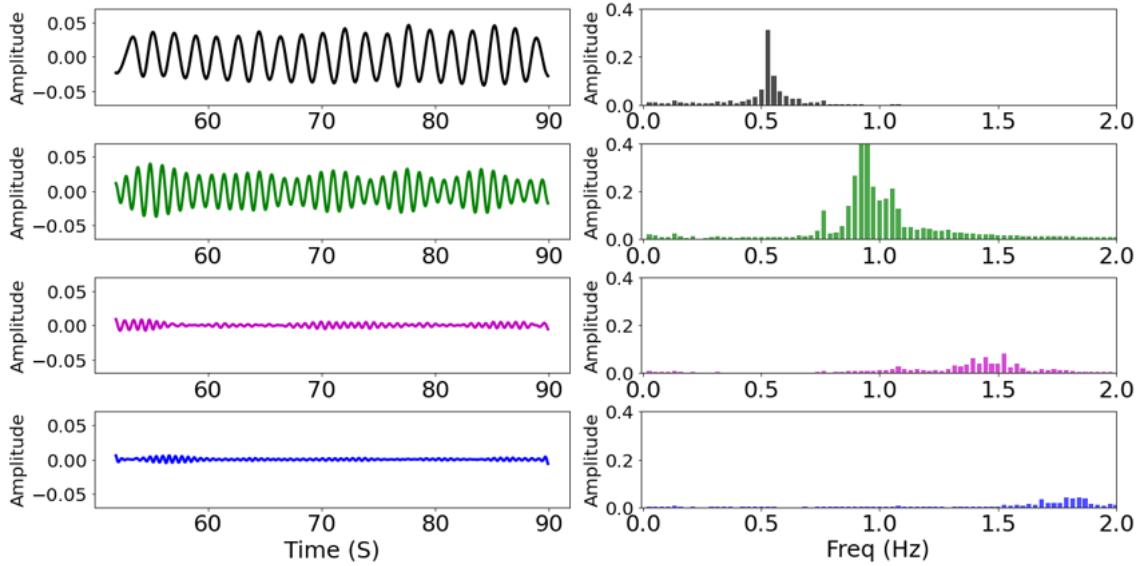


Figure 7.15: Case 2: Left) Decomposition of the unstable signal using Variation Mode Decomposition (VMD), Right) Fourier frequency of the decomposed modes.

### 7.3.4 Feature Selection

Table 7.4 shows the feature importance score calculated using LASSO regression. All the features are obtained using the penalization factor of  $\lambda = 1$ . Our model eliminates features that have negligible importance. The final score for the feature selection process is the average score of all the features that survived the LASSO regression. These features are displayed in Fig. 7.11.

### 7.3.5 Model training and testing

The designed feature selection framework selects seven features for training and testing. The dataset is split into two parts: 80 % of the generated cases are randomly selected for training, and the rest are used for the performance test. The input variables for this analysis consist of voltage magnitudes, while the output comprises the features extracted from the Enhanced Empirical Wavelet Transform (SEWT). These extracted features encompass both Instantaneous Frequency and Instantaneous Magnitude. For the analysis, a time window of 20 seconds is employed, with a step size of 1 second. A 5-fold cross-validation method is employed, wherein the training and testing processes are iterated five times to mitigate the potential variability and randomness in the evaluation process, leading to more reliable and consistent performance assessments.

Table 7.5 and Table 7.6 show the evaluation results. The root means square error (*RMSE*)

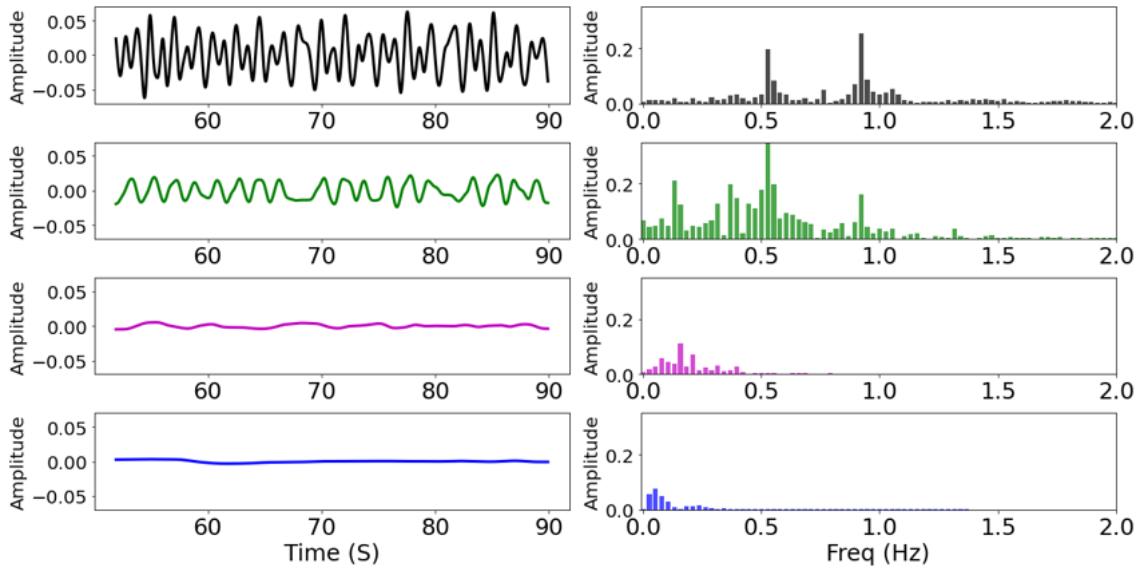


Figure 7.16: Case 2: Left)Decomposition of the unstable signal using Local Mean Decomposition (LMD), Right) Fourier frequency of the decomposed modes.

and the coefficient of multiple determination ( $R^2$ ) of each ML model under training and testing phases with the three algorithms are presented for MIMO. The graphical analysis of the RMSE results is shown in Fig. 7.18. The  $RMSE$  and  $R^2$  are calculated considering all the test samples.

Table 7.5, shows varying accuracy levels for different stability states and machine learning models. The Random Forest (RF) model demonstrates relatively better accuracy for stable and sustained states in both training and testing. The Decision Tree (DT) model performs well in stable and unstable states, particularly in the training phase. The KNN model achieves competitive results but exhibits slightly lower accuracy compared to RF and DT models.

Table 7.7 displays the training and testing duration for the execution of each ML model. Particularly noteworthy is the significantly shorter training time of the DT model when compared to the other two models. For example, with an average execution time of 0.003017 s, 1.2100 s, and 1.9400 s for stable, unstable, and sustained cases, respectively, the DT model proves to be well-suited for near real-time voltage stability assessment. The results are shown in Fig. 7.19.

### Noise impact

Noisy measured data from PMUs may introduce errors. Gaussian white noises with different signal-to-noise ratios (SNRs) are added to the PMU data to demonstrate the effect of noise on oscillation monitoring. A smaller SNR indicates a higher noise level. Three scenarios with SNRs of 60 dB, 40 dB, and 30 dB are respectively tested. First, the SEWT is utilized to decompose

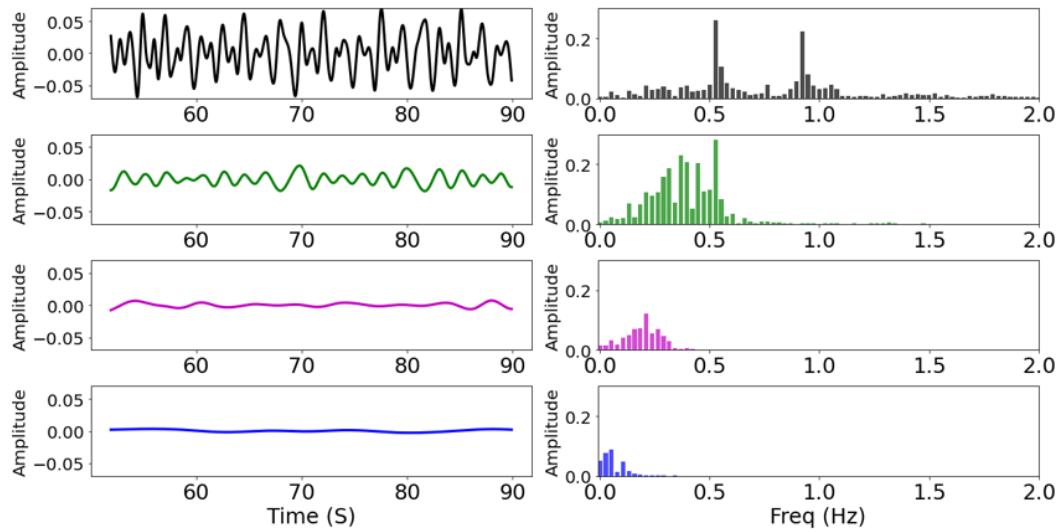


Figure 7.17: Case 2: Left)Decomposition of the unstable signal using Empirical Mode Decomposition (EMD), Right) Fourier frequency of the decomposed modes.

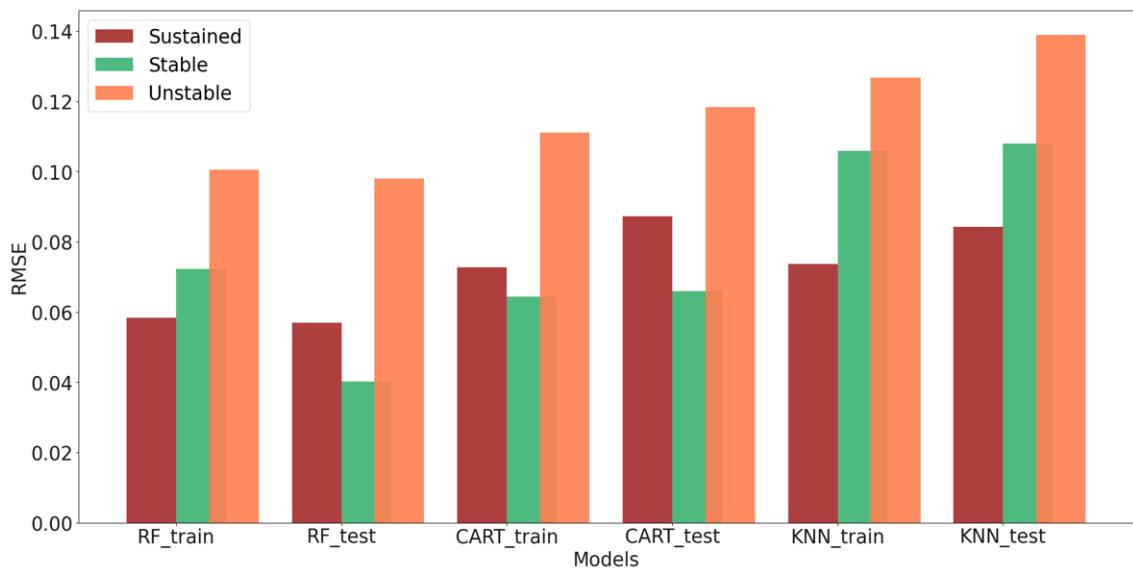


Figure 7.18: Root mean squared error.

Table 7.4: Feature Score

buses	Stable	Unstable	Sustained
g1	0	0	0
g2	0	0	0
g3	0	0	0
g4	0.165	0	9.67
g5	0	0	0
g6	0.209	0.429	0
g7	0	1.677	0
g8	0.028	0	0
g9	0	0	0
g10	0	0	0
g11	0	0.041	0.681
g12	0	0	0.048
g13	0.488	0.072	4.750
g14	0	0.740	0
g15	0	0	0
g16	0	0	0
g17	0	0	0
g18	0	47.83	0
g19	0	0	0
g20	0	0	0

Table 7.5: Value of RMSE

ML	model	Stable	Unstable	Sustained
<b>RF</b>	<i>Train</i>	0.072257	0.100567	0.058324
	<i>Test</i>	0.040147	0.09795	0.056998
<b>DT</b>	<i>Train</i>	0.064289	0.111147	0.072682
	<i>Test</i>	0.065869	0.11835	0.087194
<b>KNN</b>	<i>Train</i>	0.105839	0.126707	0.073614
	<i>Test</i>	0.107989	0.138927	0.084244

the simulated signal adaptively. Then the amplitude envelope of these modes is calculated at each noise level. The waveforms of the generator bus g-7 with and without noise are shown in Fig. 7.20. As the noise level increases, the frequency of 0.54 Hz is buried in solid background noise, and it is hard to distinguish damped frequency only by the amplitude of the spectrum. Despite the noisy signal, the SEWT method successfully separated the dominant modes from the intense noise and extracted non-stationary features of the transient signal. Regarding accuracy, Table 7.8 lists the test results for RMSE and  $R^2$  under noisy environments. It can be seen from Table 7.8 that the assessment accuracy remains good till 30 dB, which thoroughly verifies the excellent robustness of the proposed method against noise.

Table 7.6: Value of  $R^2$ 

ML model	Type	Stable	Unstable	Sustained
<b>RF</b>	<b>Train</b>	0.887	0.828	0.808
	<b>Test</b>	0.873	0.778	0.744
<b>DT</b>	<b>Train</b>	0.888	0.885	0.897
	<b>Test</b>	0.820	0.880	0.897
<b>KNN</b>	<b>Train</b>	0.575	0.789	0.831
	<b>Test</b>	0.143	0.704	0.769

Table 7.7: Training and testing time

Model	CPU TIME (S)	Stable	Unstable	Sustained
<b>RF</b>	<i>Train time</i>	17.90000	62.65300	89.00000
	<i>Test_time</i>	0.00106	0.00040	0.00060
<b>DT</b>	<i>Train time</i>	56.67200	24.80900	43.32400
	<i>Test.time</i>	0.00302	1.21000	1.94000
<b>KNN</b>	<i>Train time</i>	12.98100	135.47500	221.63500
	<i>Test_time</i>	61.67000	65.00000	97.00000

### 7.3.6 Database update condition

A robust online estimation approach should be able to accommodate different unseen network topologies. Suppose the operating point A was the starting point of an online assessment. Two typical scenarios of operation changes were considered here for the adaptability tests of online applications.

- Scenario-1: Fault at line 4022 is triggered at t=100 s.
- Scenario-2: Line 4031–4041 is tripped due to accidental failures.

Given these scenarios, the proposed assessment scheme is tested for database updates and online assessments. Based on each scenario's current and ongoing conditions and credible contingencies, 100000 (5000 samples based on 20 operating points) synthetic samples were generated and fed into data update learning. On the other hand, these new samples were exploited to test the

Table 7.8: Noise Impact

Noise Level	Type	Stable		Unstable		Sustained	
		RMSE	R2	RMSE	R2	RMSE	R2
<b>40 dB</b>	<i>Train</i>	0.2226	0.5160	0.6184	0.4965	0.5714	0.6100
	<i>Test</i>	0.4557	-0.3468	0.8104	0.1187	0.8563	0.3747
<b>30 dB</b>	<i>Train</i>	0.3005	0.5099	0.4576	0.4093	0.5707	0.6003
	<i>Test</i>	0.1381	-0.9684	0.7730	0.0511	0.8844	0.3426
<b>20 dB</b>	<i>Train</i>	2.1210	0.4906	0.3265	0.5341	0.5692	0.6055
	<i>Test</i>	2.8493	0.2268	0.4601	0.0706	0.8749	0.3416

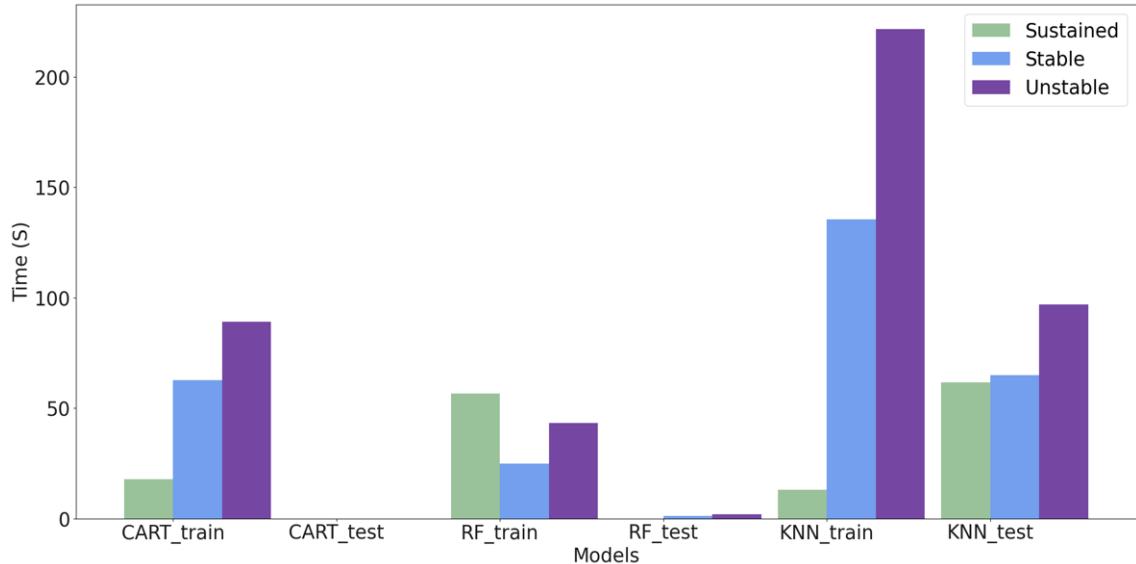


Figure 7.19: Computational time

online assessment performances of the old ML models.

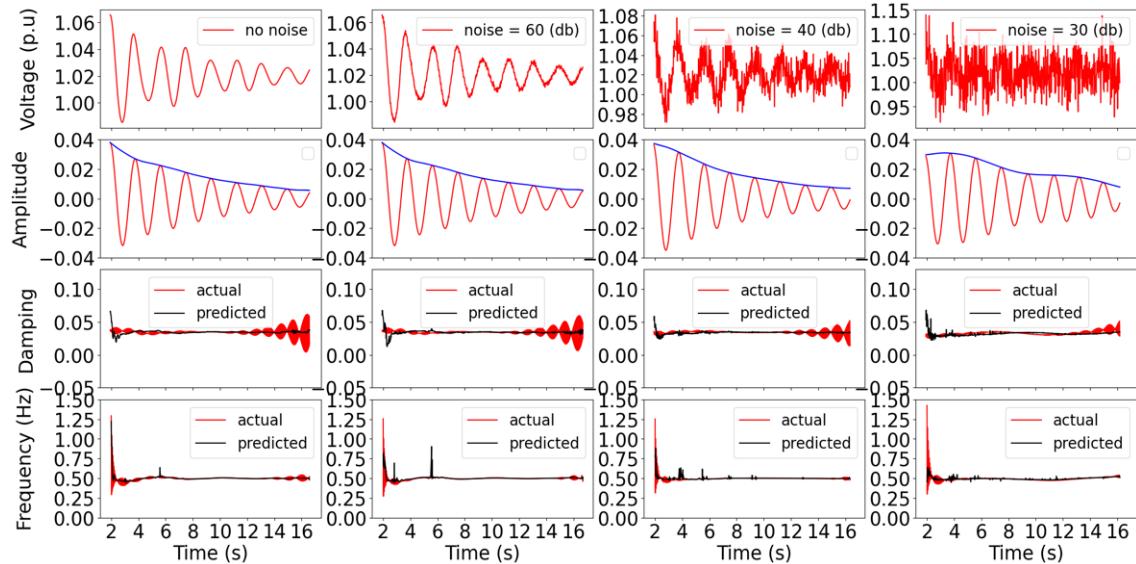


Figure 7.20: Noise impact.

Online feature extraction is done on the new samples with SEWT and HT to evaluate the effectiveness of the proposed online update scheme. Over the whole simulation, the RMSE of the predicted data is calculated. The process of online database update is explained with the help of Fig. 7.21. Before the black dotted, the system topology remains constant, and the RMSE is below 0.1. At  $t=50$  s, fault-1 is introduced, and at  $t=60$  s, fault-2 is presented. The system topology changes due to these faults.

In Fig. 7.21-a, the ML model was not trained on these faults; therefore, the RMSE increased above 0.3. At the same time, K-means clustering is applied to data points that arrive during the

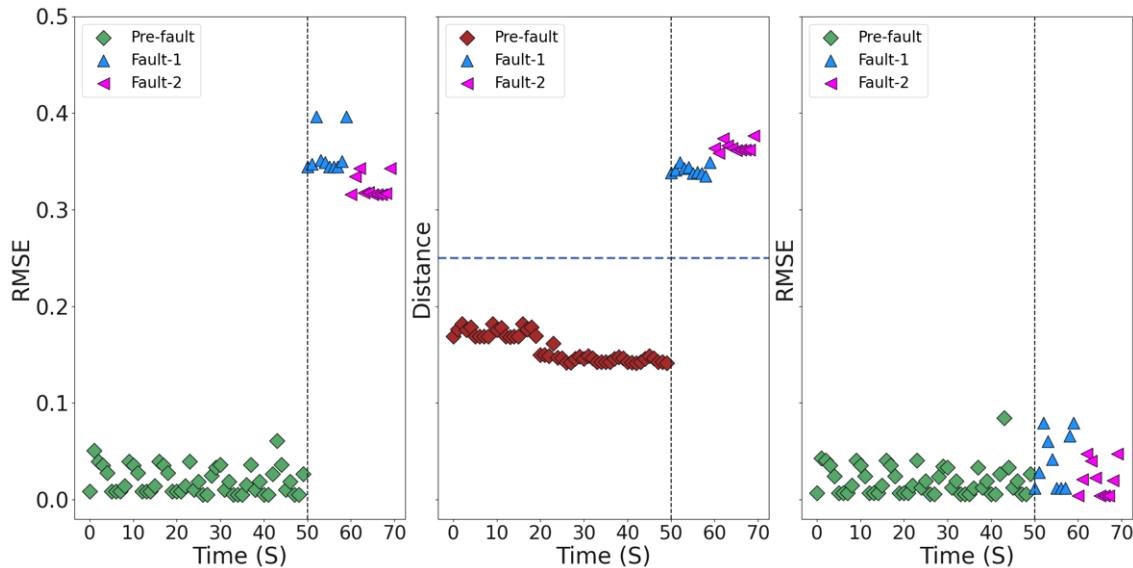


Figure 7.21: Root mean squared error.

online process, and distances from the cluster centres are calculated. K-Means Clustering aimed to handle data drift based on the system's topological changes. Fig. 7.21-b shows the distances from the cluster centres before and after introducing faults. The dotted blue line represents the distance threshold ( $\tau_d = 0.25$ ). After introducing the faults, the distance from the cluster centres crosses the threshold value triggering the database update. Once the database is updated, the ML model is retrained. Fig. 7.21-c shows the new RMSE after the retraining of ML model. After retraining, the average reduction in the RMSE of the post-fault data is approximately 80%.

## 7.4 Chapter Conclusion

Although much research has been conducted on estimating low-frequency electromechanical oscillation in power systems, most signal-processing approaches are limited to single-channel identification algorithms and cannot handle multichannel measurements simultaneously. While single-channel estimation methods have been well-developed, they suffer from potentially estimating the same dominant mode with similar but non-identical values in different channels due to differences in observability and embedded noise in the oscillation mode. On the other hand, multichannel measurements provide a more comprehensive description of the system dynamics than single-channel measurements. This study proposes an ML-based multichannel approach for estimating and monitoring interarea oscillation. To estimate the oscillation mode from measurement data, the enhanced Empirical Wavelet Transform (SEWT) is applied to the

measurement data from each bus to extract the instantaneous features of the interarea modes. The extracted features from each channel are stored in a database, and the ML algorithm is trained on these features. Finally, the trained model is applied online to identify and monitor the dominant modes from multiple channels. This approach provides an alternative method to the complicated and time-consuming power system mathematical model whilst offering a highly accurate solution compared to existing methods.

## Chapter 8

# Conclusion and future work

Due to economic and environmental constraints, power systems have operated closer to their stability limits. Voltage stability has become an increasingly important factor in the planning and operation of electric power systems. Due to the continuously-changing operating conditions and various unforeseen factors associated with large power systems, offline stability studies can no longer ensure a secure power system operation. Online stability assessment based on real-time measurements can better estimate the power system state.

While various tools have been proposed for voltage stability analysis, such as P-V curves, Q-V curves, voltage stability indices, and continuation power flow, their online implementation is challenging as they require precise mathematical modelling of the system and are computationally intensive. This research has a primary objective of monitoring and assessing the online voltage stability of the power system, which can be categorized into two parts.

The first part involves monitoring and evaluating the long-term voltage stability, which occurs over a time scale of minutes. The second part is focused on detecting and monitoring low-frequency electromechanical oscillation. The inability of the power network to provide adequate reactive power support to the system is the underlying cause of long-term voltage stability. The focal point of this research is the online monitoring and evaluation of the power system's stability using machine learning models (MLM). In recent years, machine learning methods have gained considerable attention from researchers as they provide an effective tool for online voltage stability assessment, particularly due to the non-linear nature of voltage stability. Machine learning-based techniques are more suitable than conventional analytical methods for voltage

stability analysis. Furthermore, the execution time of the MLM is significantly less once it is trained, which makes it an attractive alternative compared to conventional voltage stability analysis methods.

The second part of the thesis covers the power system oscillation and, more precisely, how to detect and monitor the oscillations. Electromechanical oscillations are characteristic of an interconnected electrical power system. Among different types of electromechanical oscillations, inter-area oscillations, which are the oscillatory modes involving the rotors of generators in different areas, have gained the attention of power system researchers. This is because a stable or a well-damped inter-area mode can become negatively damped or poorly damped due to changes in power system structure, operating conditions, and load characteristics. Due to this reason, the power system may collapse due to gradually increasing rotor oscillations over several seconds. Therefore, continuous monitoring of poorly-damped modes provides information on whether the power system operates closer to the instability and helps initiate preventive control actions.

The main objective of this study is to build a simple and computationally efficient algorithm for improving classification accuracy among different signal processing and machine learning tasks.

## 8.1 Advancements and Contributions

The original contribution of the work lies in the following aspects:

- **Enhanced Feature Ensemble Approach for Voltage Stability Monitoring:** This research introduces a novel feature ensemble approach for online voltage stability monitoring, which demonstrates superiority over conventional feature selection methods. The proposed approach effectively handles feature stability issues, an area of weakness in existing methods, leading to improved performance.
- **Cluster-based Sampling for Enhanced Performance:** - A cluster-based sampling approach is introduced for selecting operating points, effectively reducing computational burden and enhancing practical usability for speedy assessment applications
- **Bayesian hyperparameter optimization**A Bayesian optimization technique is employed to tune the hyper-parameters of machine learning (ML) models and determine the optimal number of features. This contributes to increased model accuracy and effective-

nness in predicting the dynamic voltage stability status.

- **Utilization of Sliding Window-based Empirical Wavelet Transform (SEWT) for Oscillatory Stability:** A significant contribution is made through the application of the SEWT to detect low-frequency inter-area oscillations (LFIO) in power systems.
- **Accurate estimation of instantaneous parameters:** The SEWT approach enables accurate estimation of instantaneous parameters such as amplitude and frequency, enhancing the identification and analysis of low-frequency electromechanical oscillations.
- **Multichannel approach:** SEWT is integrated with machine learning techniques using a multichannel approach, further improving accuracy in estimating inter-area modes.
- **Automated Mode Detection in EWT:** SEWT automatically detect modes without the need for pre-specifying their number. This addresses the challenge of closely spaced modes in power systems and concealed interarea modes caused by signal noise

## 8.2 Demonstrating Superiority: Unveiling the Strengths of the Proposed Approach Over Existing Methods

The superiority of the proposed work compared to existing methods is evident in the following aspects:

- The feature ensemble approach significantly enhances the accuracy of voltage stability prediction by effectively addressing inherent feature stability challenges, a hurdle that conventional methods often encounter. This approach offers a distinct advantage over single feature selection techniques, as it incorporates multiple feature selectors. By doing so, it not only bolsters stability and robustness but also introduces a heightened level of diversity in the selection process. This diversity contributes to mitigating overfitting concerns and enriches the overall selection of features, ultimately leading to improved generalization capabilities of the model.
- The cluster-based sampling technique enhances computational efficiency and captures diverse load characteristics, offering a practical alternative to traditional Monte Carlo simulations.
- The incorporation of Bayesian optimization for hyper-parameter tuning makes a substan-

tial contribution to the enhancement of voltage stability predictions in terms of accuracy and adaptability. This innovative approach not only fine-tunes hyper-parameters for improved model performance but also introduces a unique advantage in the feature ensemble context. Specifically, Bayesian hyper-parameter optimization facilitates the selection of the optimal number of features for the ensemble without necessitating the explicit specification of the number of features to be chosen. This adaptive feature selection process further amplifies the model's capability to adapt and yield precise voltage stability predictions.

- SEWT emerges as a highly adaptable signal processing technique, presenting a marked advantage over traditional EWT (Empirical Wavelet Transform) approaches. Particularly noteworthy is SEWT's inherent flexibility, eliminating the necessity of specifying the number of modes in advance—a constraint often encountered by EWT. Functioning as an adaptive signal processing tool, SEWT boasts a multifaceted set of advantages that distinguishes it from other established methods, including Prony and matrix pencil techniques.
- SEWT's adaptive nature empowers it with the capacity to seamlessly adjust to varying signal characteristics, making it well-suited for handling non-linear and non-stationary signals—a characteristic that sets it apart from conventional methods. The technique's proficiency extends to managing intricate signals characterized by multiple modes, further enhancing its utility in real-world scenarios. An additional distinctive trait of SEWT is its immunity to mode mixing, a challenge often faced by existing adaptive signal processing techniques.
- Moreover, SEWT excels in the realm of noise compression, a feature that bolsters its effectiveness in practical applications. When compared to established adaptive signal processing techniques, SEWT stands out due to its unique combination of benefits, demonstrating its prowess in accurately estimating modal parameters and contributing to the detection and analysis of low-frequency inter-area oscillations within power systems.
- The SEWT approach coupled with machine learning offers superior accuracy in identifying and analyzing low-frequency inter-area oscillations for online estimation of oscillatory modes.

In summary, we have demonstrated that the improved EWT method can effectively decompose the frequency boundaries and outperforms the original EWT method. It can extract a series of amplitude-modulated-frequency-modulated (AM-FM) signals from the given data. Moreover, the combined scheme based on the enhanced EWT method provided modal estimates (natural frequency and damping ratio) with high accuracy even when the signal was embedded with high noise levels. Due to its simplicity, EWT algorithm could be implemented for online applications efficiently. Such a technique can also be used for big data applications for low computational expenses. However, there also exist certain limitations to the proposed EWT method. The main shortcoming of EWT is that Fourier segmentation strongly depends on the local maxima of the Fourier spectrum's amplitude. This study introduced an enhanced empirical wavelet transform using the sliding window technique over the Fourier spectrum. This technique avoids getting stuck in the local maxima and identifies maxima for optimal spectrum decomposition intervals. Moreover, using this technique, the spectrum is automatically decomposed, and there is no need to specify the number of segments in advance.

### 8.3 Future Work

We envision extending this work in a more robust environment that addresses uncertainties and faults in distribution networks. Our future work will describe the network accurately and present this information in machine learning. Furthermore, issues such as identifying weak buses for emergency demand response-based voltage stability improvement would be given more attention. The proposed MLM could be further improved with more advanced deep learning techniques and applied heuristic search algorithms.

Currently, features are just voltage phase angle and voltage magnitude, which seems relatively straightforward. Next, we will develop more complex features such as current, active power, reactive power, their derivatives, frequency, the flow direction of reaction power, etc. While long-term voltage stability is widely studied, short-term stability receives significantly less attention. While the issue used to be a low-probability event, this might no longer be the case with the changes we see in the systems. It can be expected that local voltage-related issues will be more common and will become a significant focus of system operators throughout the energy transition. Therefore, a critical need exists to investigate and estimate short-term voltage instabilities.

Future investigations of oscillations would focus on identifying power system oscillation modes using ambient data. As ring-down data is not always available, identifying power systems' dynamic characteristics through measured ambient data is an alternative in this situation. Minor stochastic disturbances, such as random load fluctuation, always exist in power systems. There is an important assumption that these disturbances are white noise. Under these stochastic disturbances, the ambient data measured is abundant in dynamic information. Many scholars have conducted effective and productive explorations to obtain dynamic characteristics based on ambient data. The effect of these and other factors can be studied with real and simulation studies, and this area is subject to future work.

Another area of focus would be underlying system conditions and effects related to resonance in power systems caused by forced oscillations and ways to detect such scenarios using synchrophasors. The resonance depends on many factors, but the most important is the location of the forced oscillation source. Accurately locating Forced Oscillations sources in a large-scale power system is a challenging and important aspect of power system operation. Machine learning techniques can be used to localize forced oscillation sources using data from PMU by tracing the source of the forced oscillation on the branch level in the power system network.

# References

- [1] N. Hatziargyriou, J. Milanovic, C. Rahmann, V. Ajjarapu, C. Canizares, I. Erlich, D. Hill, I. Hiskens, I. Kamwa, B. Pal, P. Pourbeik, J. Sanchez-Gasca, A. Stankovic, T. Van Cutsem, V. Vittal, and C. Vournas, “Definition and Classification of Power System Stability - Revisited & Extended,” *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3271–3281, 7 2021.
- [2] B. Bhargava, “Synchronized phasor measurement system project at southern california edison co,” in *1999 IEEE Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No.99CH36364)*, vol. 1, 1999, pp. 16–22 vol.1.
- [3] T. Van Cutsem and L. Papangelis, “Description, modeling and simulation results of a test system for voltage stability analysis,” *Book*, vol. 6, no. 1, p. 49, 2013. [Online]. Available: [https://orbi.uliege.be/bitstream/2268/141234/1/Nordic\\_test\\_system\\_V6.pdf](https://orbi.uliege.be/bitstream/2268/141234/1/Nordic_test_system_V6.pdf)
- [4] A. K. Singh and M. Fozdar, “Event-driven frequency and voltage stability predictive assessment and unified load shedding,” *IET Generation, Transmission and Distribution*, vol. 13, no. 19, pp. 4410–4420, 10 2019.
- [5] L. Rouco and I. J. Pérez-Arriaga, “Multi-Area analysis of small signal stability in large electric power systems by sma,” *IEEE Transactions on Power Systems*, vol. 8, no. 3, pp. 1257–1265, 1993.
- [6] A. M. Vural, “Contribution of high voltage direct current transmission systems to inter-area oscillation damping: A review,” *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 892–915, 5 2016.
- [7] B. C. Pal, “Robust Control in Power Systems,” in *Robust Control in Power Systems*. Boston: Kluwer Academic Publishers, 2006, pp. 1–4.

- [8] R. Witzmann, “Damping of Interarea Oscillations in Large Interconnected Power Systems,” .., 2000.
- [9] C. W. Taylor, N. J. Balu, and D. Maratukulam, “Power system voltage stability,” .., p. 273, 1994.
- [10] G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca, R. Schulz, A. Stankovic, C. Taylor, and V. Vittal, “Causes of the 2003 major grid blackouts in North America Europe, and recommended means to improve system dynamic performance,” *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1922–1928, 11 2005.
- [11] P. Pourbeik, P. S. Kundur, and C. W. Taylor, “The anatomy of a power grid blackout - Root causes and dynamics of recent major blackouts,” *undefined*, vol. 4, no. 5, pp. 22–29, 2006.
- [12] W. O. Committee *et al.*, “Western systems coordinating council disturbance report for the power system outages that occurred on the western interconnection on july 2, 1996 and july 3, 1996,” *Western Systems Coordinating Council*, 1996.
- [13] A. Muir and J. Lopatto, “Final report on the august 14, 2003 blackout in the united states and canada: causes and recommendations,” .., 2004.
- [14] F. Vandenbergh, E. Grebe, D. Klaar, K. Kleinekorte, J. Rodriguez, H. Erven, and L. Tas-san, “Final report of the investigation committee on the 28 september 2003 blackout in italy,” *UCTE, Tech. Rep*, 2004.
- [15] E. van der Vleuten and V. Lagendijk, “Transnational infrastructure vulnerability: The historical shaping of the 2006 european “blackout”,” *Energy Policy*, vol. 38, no. 4, pp. 2042–2052, 2010, energy Security - Concepts and Indicators with regular papers. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0301421509008970>
- [16] K. Yamashita, J. Li, P. Zhang, and C. C. Liu, “Analysis and control of major blackout events,” *2009 IEEE/PES Power Systems Conference and Exposition, PSCE 2009*, 2009.
- [17] M. Klein, G. J. Rogers, and P. Kundur, “A fundamental study of inter-area oscillations in power systems,” *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 914–921, 1991. [Online]. Available: <http://ieeexplore.ieee.org/document/119229/>

- [18] A. M. Vural, "Contribution of high voltage direct current transmission systems to inter-area oscillation damping: A review," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 892–915, 2016.
- [19] S. You, G. Kou, Y. Liu, X. Zhang, Y. Cui, M. J. Till, W. Yao, and Y. Liu, "Impact of high pv penetration on the inter-area oscillations in the us eastern interconnection," *IEEE Access*, vol. 5, pp. 4361–4369, 2017.
- [20] A. B. Iskakov, E. Y. Kutyakov, N. V. Tomin, D. A. Panasetsky, A. N. Abramov, and S. V. Dushin, "Estimation of the location of inter-area oscillations and their interactions in electrical power systems using lyapunov modal analysis," *International Journal of Electrical Power & Energy Systems*, vol. 153, p. 109374, 2023.
- [21] M. Klein, G. Rogers, and P. Kundur, "A fundamental study of inter-area oscillations in power systems," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 914–921, 1991.
- [22] D. Eltigani and S. Masri, "Challenges of integrating renewable energy sources to smart grids: A review," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 770–780, 2015.
- [23] M. Weiss, B. N. Abu-Jaradeh, A. Chakrabortty, A. Jamehbozorg, F. Habibi-Ashrafi, and A. Salazar, "A wide-area svc controller design for inter-area oscillation damping in wecc based on a structured dynamic equivalent model," *Electric Power Systems Research*, vol. 133, pp. 1–11, 2016.
- [24] S. Al-Ali, I. Nassar, and H. Weber, "Interconnection of the european entso-e-ce system with the turkish system: investigation of the expected inter-area-oscillations behaviour," *Romania*, vol. 50, pp. 49–99, 2011.
- [25] P.-A. LÖf, D. J. Hill, S. Arnborg, and G. Andersson, "On the analysis of long-term voltage stability," *International Journal of Electrical Power & Energy Systems*, vol. 15, no. 4, pp. 229–237, 1993.
- [26] R. Rincón, A. Pavas, and E. Mojica-Nava, "Long-term voltage stability analysis and network topology in power systems," in *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 2016, pp. 1–6.
- [27] D. Wang, S. Parkinson, W. Miao, H. Jia, C. Crawford, and N. Djilali, "Online voltage security assessment considering comfort-constrained demand response control of distributed

- heat pump systems," *Applied Energy*, vol. 96, pp. 104–114, 2012.
- [28] K. D. Dharmapala, A. Rajapakse, K. Narendra, and Y. Zhang, "Machine Learning Based Real-Time Monitoring of Long-Term Voltage Stability Using Voltage Stability Indices," *IEEE Access*, vol. 8, pp. 222\,544–222\,555, 2020.
- [29] M. Amroune, "Machine Learning Techniques Applied to On-Line Voltage Stability Assessment: A Review," *Archives of Computational Methods in Engineering*, vol. 28, no. 2, pp. 273–287, 3 2021.
- [30] T. Van Cutsem, Y. Jacquemart, J.-N. Marquet, and P. Pruvot, "A comprehensive analysis of mid-term voltage stability," *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1173–1182, 1995.
- [31] T. Babnik, K. Gorner, and B. Mahkovec, "Wide area monitoring system," *Encyclopedia of Wireless Networks*, 2020.
- [32] X. Liu, X. Zeng, L. Yao, G. I. Rashed, and C. Deng, "Power system state estimation based on fusion of wams/scada measurements: A survey," in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, 2018, pp. 1–6.
- [33] M. A. Donolo, "Advantages of synchrophasor measurements over SCADA measurements for power system state estimation," *cdn.selinc.com*, 2006.
- [34] Y. Zhang, L. Wang, Y. Xiang, and C.-W. Ten, "Power system reliability evaluation with scada cybersecurity considerations," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1707–1721, 2015.
- [35] J. Cai, Z. Huang, J. Hauer, and K. Martin, "Current status and experience of wams implementation in north america," in *2005 IEEE/PES Transmission and Distribution Conference and Exposition: Asia and Pacific*, 2005, pp. 1–7.
- [36] M. Mynam, A. Harikrishna, and V. Singh, "Synchrophasors redefining scada systems," vol. 26, no. 2. International Association on Electricity Generation Transmission & Distribution, 2013, pp. 22–28.
- [37] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power system dynamics: stability and control*, 3rd ed. John Wiley & Sons, Ltd, 2011.

- [38] R. Chintakindi and A. Mitra, "WAMS challenges and limitations in load modeling, voltage stability improvement, and controlled island protection—A review," *Energy Reports*, vol. 8, pp. 699–709, 4 2022.
- [39] D. Karlsson, M. Hemmingsson, and S. Lindahl, "Wide area system monitoring and control," *IEEE Power and Energy Magazine*, vol. 2, no. 5, pp. 68–76, 9 2004.
- [40] Prabha Kundur, "Power System Stability and Control - Prabha Kundur - McGraw-Hill Education," p. 1176, 1994. [Online]. Available: [https://books.google.com/books/about/Power\\_System\\_Stability\\_and\\_Control.html?id=2cbvfyf8Ly4AC](https://books.google.com/books/about/Power_System_Stability_and_Control.html?id=2cbvfyf8Ly4AC) <https://www.mheducation.co.in/html/9780070635159.html>
- [41] D. Mondal, *Power system small signal stability analysis and control*, 1st ed. London: Elsevier, 2014.
- [42] G. Rogers, *Power system oscillations*. Springer Science & Business Media, 2012.
- [43] S. Chakrabarti and B. Jeyasurya, "Generation rescheduling using ANN-based computation of parameter sensitivities of the voltage stability margin," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 8, pp. 1164–1169, dec 2008. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0952197608000754>
- [44] Y. Li, M. Zhang, and C. Chen, "A deep-learning intelligent system incorporating data augmentation for short-term voltage stability assessment of power systems," *Applied Energy*, vol. 308, p. 118347, 2022.
- [45] X. Meng, P. Zhang, and D. Zhang, "Decision tree for online voltage stability margin assessment using c4. 5 and relief-f algorithms," *Energies*, vol. 13, no. 15, p. 3824, 2020.
- [46] P. Duraipandy and D. Devaraj, "On-line voltage stability assessment using least squares support vector machine with reduced input features," *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies, ICCICCT 2014*, pp. 1070–1074, 12 2014.
- [47] B. Gao, G. K. Morison, and P. Kundur, "Voltage Stability Evaluation Using Modal Analysis," *IEEE Transactions on Power Systems*, vol. 7, no. 4, pp. 1529–1542, 1992. [Online]. Available: <http://ieeexplore.ieee.org/document/207377/>

- [48] H. Pinto, N. Martins, X. Vieira, A. Bianco, P. Gomes, and M. dos Santos, “Modal analysis for voltage stability: Application at base case and point of collapse,” *Proceedings of Bulk Power Systems Voltage Phenomena III: Voltage Stability Security and Control*, pp. 215–228, 1994.
- [49] A. J. Flueck and J. R. Dondeti, “A new continuation power flow tool for investigating the nonlinear effects of transmission branch parameter variations,” *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 223–227, 2 2000.
- [50] M. Chakravorty and S. Patra, “Voltage stability analysis using conventional methods,” *International Conference on Signal Processing, Communication, Power and Embedded System, SCOPES 2016 - Proceedings*, pp. 496–501, 6 2017.
- [51] J. P. Abbott, “An efficient algorithm for the determination of certain bifurcation points,” *Journal of Computational and Applied Mathematics*, vol. 4, no. 1, pp. 19–27, 1978.
- [52] H. B. Keller, “Numerical methods in bifurcation problems,” *Lectures on Mathematics and Physics, Tata Institute of Fundamental Research (Bombay)*, 1987, 1987.
- [53] W. C. Rheinboldt and J. V. Burkardt, “A Locally Parameterized Continuation Process,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 9, no. 2, pp. 215–235, 6 1983.
- [54] R. G. Melhem and W. C. Rheinboldt, “A comparison of methods for determining turning points of nonlinear equations,” *Computing*, vol. 29, no. 3, pp. 201–226, 9 1982.
- [55] G. Pönisch, H. S. Computing, and undefined 1981, “Computing turning points of curves implicitly defined by nonlinear equations depending on a parameter,” *Springer*, 1981. [Online]. Available: <https://link.springer.com/content/pdf/10.1007/BF02241778.pdf>
- [56] A. B. Neto and D. A. Alves, “Improved geometric parameterisation techniques for continuation power flow,” *IET Generation, Transmission and Distribution*, vol. 4, no. 12, pp. 1349–1359, 12 2010.
- [57] Y. Ju, W. Wu, B. Zhang, and H. Sun, “Continuation power flow based on a novel local geometric parameterisation approach,” *IET Generation, Transmission & Distribution*, vol. 8, no. 5, pp. 811–818, 2014.

- [58] K. Iba, H. Suzuki, M. Egawa, and T. Watanabe, “Calculation of critical loading condition with nose curve using homotopy continuation method,” *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 584–593, 1991.
- [59] A. B. Almeida, E. V. De Lorenci, R. C. Leme, A. C. Z. De Souza, B. I. L. Lopes, and K. Lo, “Probabilistic voltage stability assessment considering renewable sources with the help of the pv and qv curves,” *IET Renewable Power Generation*, vol. 7, no. 5, pp. 521–530, 2013.
- [60] M. Haque, “Determination of steady-state voltage stability limit using pq curve,” *IEEE Power Engineering Review*, vol. 22, no. 4, pp. 71–72, 2002.
- [61] Z. A. Kamaruzzaman and A. Mohamed, “Impact of grid-connected photovoltaic generator using P-V curve and improved voltage stability index,” *Conference Proceeding - 2014 IEEE International Conference on Power and Energy, PECon 2014*, pp. 196–200, 3 2014.
- [62] L. D. Arya and H. K. Verma, “A method for tracing P-Vcurve for voltage stability analysis with voltage dependent loads,” <http://dx.doi.org/10.1080/07313569608955696>, vol. 24, no. 6, pp. 583–596, 9 2007. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/07313569608955696>
- [63] K. Ramalingam, C. I. . J. International, and undefined 2008, “Determination of steady state voltage stability limit using PQ curves for voltage sensitive loads,” [ieeexplore.ieee.org](https://ieeexplore.ieee.org/abstract/document/4745152/), 2008. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/4745152/>
- [64] A. K. Sinha and D. Hazarika, “A comparative study of voltage stability indices in a power system,” *International Journal of Electrical Power & Energy Systems*, vol. 22, no. 8, pp. 589–596, 11 2000. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0142061500000144>
- [65] Y. H. Hong, “Fast calculation of a voltage stability index of power systems,” *IEEE Transactions on Power Systems*, vol. 12, no. 4, pp. 1555–1560, 1997.
- [66] F. Gubina and B. Strmcnik, “A simple approach to voltage stability assessment in radial networks,” *IEEE Transactions on Power Systems*, vol. 12, no. 3, pp. 1121–1128, 1997.
- [67] U. Eminoglu, M. H. . n. International, and undefined 2007, “A voltage stability index for radial distribution networks,” [ieeexplore.ieee.org](https://ieeexplore.ieee.org/abstract/document/4468982/), 2007. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/4468982/>

- [68] A. Berizzi, P. M. .... P. Bresesti, and undefined 1996, "System-area operating margin assessment and security enhancement against voltage collapse," *ieeexplore.ieee.org*, 1996. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/535686/>
- [69] X. Ancheng, L. Ruihuang, L. Mingkai, J. H. Chow, B. Tianshu, Y. Ting, and P. Tianjiao, "On-line voltage stability index based on the voltage equation of transmission lines," *IET Generation, Transmission and Distribution*, vol. 10, no. 14, pp. 3441–3448, 11 2016.
- [70] J. Modarresi, E. Gholipour, and A. Khodabakhshian, "A comprehensive review of the voltage stability indices," *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 1–12, 9 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S1364032116301204>
- [71] M. Moghavvemi and M. Faruque, "Technique for assessment of voltage stability in ill-conditioned radial distribution network," *IEEE Power Engineering Review*, vol. 21, no. 1, pp. 58–60, 2001.
- [72] S. Ratna, R. Tiwari, and K. R. Niazi, "Voltage stability assessment in power systems using line voltage stability index," *Computers & Electrical Engineering*, vol. 70, pp. 199–211, 2018.
- [73] I. Musiri and T. K. Abdul Rahman, "On-line voltage stability based contingency ranking using fast voltage stability index (FVSI)," *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, 2002.
- [74] I. Dobson, M. Parashar, and C. Carter, "Combining phasor measurements to monitor cutset angles," pp. 1–9, 2010.
- [75] S. Yari, H. Khoshkhoo, S. Yari, and H. Khoshkhoo, "Assessment of line stability indices in detection of voltage stability status," *ieeexplore.ieee.org*, 2017. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7977454/>
- [76] A. MOhamed, G. Jasmon, and S. Yusof PTI Asia Sdn Bt, "A static voltage collapse indicator using line stability factors," *Journal of Industrial Technology*, vol. 7, no. 1, pp. 73–85, 1998.
- [77] L. Neerugattu and G. Raju, "New criteria for voltage stability evaluation in interconnected power system," in *Nat. Power Syst. Conf.*, 2012.

- [78] P. Kessel and H. Glavitsch, "Estimating the voltage stability of a power system," *IEEE Transactions on Power Delivery*, vol. 1, no. 3, pp. 346–354, 1986.
- [79] G. Verbic and F. Gubina, "A new concept of voltage-collapse protection based on local phasors," *IEEE Transactions on Power Delivery*, vol. 19, no. 2, pp. 576–581, 2004.
- [80] C. D. Vournas, C. Lambrou, and M. Kanatas, "Application of local autonomous protection against voltage instability to ieee test system," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3300–3308, 2015.
- [81] C. D. Vournas, C. Lambrou, and P. Mandoulidis, "Voltage stability monitoring from a transmission bus pmu," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3266–3274, 2016.
- [82] V. Balamourougan, T. Sidhu, and M. Sachdev, "Technique for online prediction of voltage collapse," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 151, no. 4, pp. 453–460, 2004.
- [83] B. Mahesh, "Machine Learning Algorithms-A Review Machine Learning Algorithms-A Review," *International Journal of Science and Research*, 2018. [Online]. Available: <https://www.researchgate.net/publication/344717762>
- [84] S. Mahdi Miraftabzadeh, M. Longo, F. Foiadelli, M. Pasetti, R. Igual, E. Hossain, and A. Abu-Siada, "Advances in the application of machine learning techniques for power system analytics: A survey," *mdpi.com*, pp. 11–14, 2019.
- [85] A. Krogh, "What are artificial neural networks?" *Nature Biotechnology*, vol. 26, 2008. [Online]. Available: <http://www.r-project.org/>
- [86] D. Popović, D. Kukolj, and F. Kulić, "Monitoring and assessment of voltage stability margins using artificial neural networks with a reduced input set," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 145, no. 4, p. 355, 2002. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/ip-gtd{-.}19981977>
- [87] T. Jain, L. Srivastava, and S. N. Singh, "Fast Voltage Contingency Screening Using Radial Basis Function Neural Network," *IEEE Transactions on Power Systems*, vol. 18, no. 4, pp. 1359–1366, 11 2003. [Online]. Available: <http://ieeexplore.ieee.org/document/1245558/>

- [88] S. Hashemi and M. R. Aghamohammadi, "Wavelet based feature extraction of voltage profile for online voltage stability assessment using RBF neural network," *International Journal of Electrical Power and Energy Systems*, 2013.
- [89] D. Devaraj and J. Preetha Roselyn, "On-line voltage stability assessment using radial basis function network model with reduced input features," *International Journal of Electrical Power and Energy Systems*, 2011.
- [90] A. R. Bahmanyar and A. Karami, "Power system voltage stability monitoring using artificial neural networks with a reduced set of inputs," *International Journal of Electrical Power and Energy Systems*, vol. 58, pp. 246–256, 6 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0142061514000325>
- [91] S. M. Ashraf, A. Gupta, D. K. Choudhary, and S. Chakrabarti, "Voltage stability monitoring of power systems using reduced network and artificial neural network," *International Journal of Electrical Power and Energy Systems*, vol. 87, pp. 43–51, 6 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.ijepes.2016.11.008>
- [92] V. Malbasa, C. Zheng, P. C. Chen, T. Popovic, and M. Kezunovic, "Voltage Stability Prediction Using Active Machine Learning," *IEEE Transactions on Smart Grid*, 2017.
- [93] T. Peter and R. P. Sajith, "Voltage stability assessment in power systems using Artificial Neural Networks," in *2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives, AICERA/iCMMD 2014 - Proceedings*, 2014.
- [94] W. M. Villa-Acevedo, J. M. López-Lezama, D. G. C. Energies, and undefined 2020, "Voltage Stability Margin Index Estimation Using a Hybrid Kernel Extreme Learning Machine Approach," *Energies*, 2020. [Online]. Available: <https://www.mdpi.com/642262>
- [95] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," *School of EECS, Washington State University*, 2006.
- [96] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine," *Journal of Theoretical and Applied Information Technology*, 2010. [Online]. Available: [www.jatit.org](http://www.jatit.org)
- [97] K. Sajan, V. Kumar, and B. Tyagi, "Genetic algorithm based support vector machine for on-line voltage stability monitoring," *International Journal of Electrical Power & Energy*

- Systems*, vol. 73, pp. 200–208, 2015.
- [98] A. Shukla, K. Verma, and R. Kumar, “Online voltage stability monitoring of distribution system using optimized Support Vector Machine,” *2016 IEEE 6th International Conference on Power Systems, ICPS 2016*, 10 2016.
- [99] A. H. Poursaeed and F. Namdari, “Online Voltage Stability Monitoring and Prediction by Using Support Vector Machine Considering Overcurrent Protection for Transmission Lines,” *Iranian Journal of Electrical and Electronic Engineering*, vol. 16, no. 3, pp. 325–335, 2020.
- [100] ——, “Real-time voltage stability monitoring using weighted least square support vector machine considering overcurrent protection,” *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107690, 2022.
- [101] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, “Classification and regression trees,” *Classification and Regression Trees*, pp. 1–358, 1 2017.
- [102] R. Siciliano and F. Mola, “Multivariate data analysis and modeling through classification and regression trees,” *Computational Statistics & Data Analysis*, vol. 32, no. 3-4, pp. 285–301, 1 2000.
- [103] C. Zheng, V. Malbasa, and M. Kezunovic, “A fast stability assessment scheme based on classification and regression tree,” ., pp. 1–6, 2012.
- [104] Z. Li and W. Wu, “Phasor measurements-aided decision trees for power system security assessment,” in *2009 2nd International Conference on Information and Computing Science, ICIC 2009*, vol. 1, 2009.
- [105] V. Krishnan and J. D. McCalley, “Progressive entropy based contingency grouping for deriving decision trees for multiple contingencies,” *International Journal of Electrical Power & Energy Systems*, vol. 45, no. 1, pp. 35–41, 2 2013.
- [106] H. Mohammadi and M. Dehghani, “PMU based voltage security assessment of power systems exploiting principal component analysis and decision trees,” *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 655–663, 1 2015.

- [107] H. Mohammadi, G. Khademi, D. Simon, and M. Dehghani, “Multi-objective optimization of decision trees for power system voltage security assessment,” *10th Annual International Systems Conference, SysCon 2016 - Proceedings*, 6 2016.
- [108] H. I. H. Z. Abidin, K. L. Lo, and Z. F. Hussein, “Multiple attribute dynamic fuzzy decision tree approach for voltage collapse evaluation,” *National Power Engineering Conference, PECon 2003 - Proceedings*, pp. 62–65, 2003.
- [109] X. Meng, P. Zhang, Y. Xu, and H. Xie, “Construction of decision tree based on C4.5 algorithm for online voltage stability assessment,” *International Journal of Electrical Power & Energy Systems*, vol. 118, pp. 1–8, 6 2020.
- [110] L. Vanfretti and V. S. Arava, “Decision tree based classification of multiple operating conditions for power system voltage stability assessment,” *International Journal of Electrical Power and Energy Systems*, vol. 123, pp. 1–10, 12 2020.
- [111] V. S. Narasimham Arava and L. Vanfretti, “Decision trees for voltage stability assessment,” *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2020*, 11 2020.
- [112] L. Wehenkel and M. Pavella, “Decision tree approach to power systems security assessment,” *International Journal of Electrical Power & Energy Systems*, vol. 15, no. 1, pp. 13–36, 1993.
- [113] Z.-H. Zhou, “Ensemble Learning,” *Encyclopedia of Biometrics*, pp. 270–273, 2009. [Online]. Available: [https://link.springer.com/referenceworkentry/10.1007/978-0-387-73003-5{\\_}293](https://link.springer.com/referenceworkentry/10.1007/978-0-387-73003-5{_}293)
- [114] S. S. Maaji, G. Cosma, A. Taherkhani, A. A. Alani, and T. M. McGinnity, “On-line voltage stability monitoring using an Ensemble AdaBoost classifier,” *2018 4th International Conference on Information Management, ICIM 2018*, pp. 253–259, 6 2018.
- [115] R. Zhang, Y. Xu, Z. Y. Dong, P. Zhang, and K. P. Wong, “Voltage stability margin prediction by ensemble based extreme learning machine,” in *IEEE Power and Energy Society General Meeting*, 2013.
- [116] K. S. Sajan, B. Tyagi, and V. Kumar, “Genetic algorithm based artificial neural network model for voltage stability monitoring,” *2014 18th National Power Systems Conference*,

*NPSC 2014*, 5 2015.

- [117] T. Van Cutsem and C. Vournas, *Voltage stability of electric power systems*. Boston, MA: Springer US, 2008. [Online]. Available: <http://link.springer.com/10.1007/978-0-387-75536-6>
- [118] D. Q. Zhou, U. D. Annakkage, and A. D. Rajapakse, “Online monitoring of voltage stability margin using an artificial neural network,” *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1566–1574, 8 2010. [Online]. Available: <http://ieeexplore.ieee.org/document/5405083/>
- [119] A. C. Zambroni De Souza, F. W. Mohn, I. F. Borges, and T. R. Ocariz, “Using PV and QV curves with the meaning of static contingency screening and planning,” *Electric Power Systems Research*, 2011.
- [120] P. Aristidou, D. Fabozzi, and T. Van Cutsem, “Dynamic simulation of large-scale power systems using a parallel schur-complement-based decomposition method,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 10, pp. 2561–2570, 2014. [Online]. Available: <https://core.ac.uk/download/pdf/46664880.pdf>
- [121] D. Rangaprakash, T. Odemuyiwa, D. Narayana Dutt, and G. Deshpande, “Density-based clustering of static and dynamic functional mri connectivity features obtained from subjects with cognitive impairment,” *Brain informatics*, vol. 7, no. 1, pp. 1–13, 2020.
- [122] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, “Ensemble feature selection: Homogeneous and heterogeneous approaches,” *Knowledge-Based Systems*, vol. 118, pp. 124–139, 2 2017.
- [123] O. Sagi and L. Rokach, “Ensemble learning: A survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.
- [124] G. Chandrashekhar and F. Sahin, “A survey on feature selection methods,” *Computers and Electrical Engineering*, 2014.
- [125] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.
- [126] J. Lv, M. Pawlak, and U. D. Annakkage, “Prediction of the transient stability boundary using the lasso,” *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 281–288, 2013.

- [127] V. Brandwajn, “Efficient bounding method for linear contingency analysis,” *IEEE Transactions on Power Systems*, vol. 3, no. 1, 1988.
- [128] L. Yang and A. Shami, “On hyperparameter optimization of machine learning algorithms: Theory and practice,” *Neurocomputing*, 2020.
- [129] M. Feurer, J. T. Springenberg, and F. Hutter, “Initializing Bayesian hyperparameter optimization via meta-learning,” in *Proceedings of the National Conference on Artificial Intelligence*, 2015.
- [130] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient Global Optimization of Expensive Black-Box Functions,” *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.
- [131] J. Kim, S. Kim, and S. Choi, “Learning to Warm-Start Bayesian Hyperparameter Optimization,” ., 2018.
- [132] P. Aristidou, S. Lebeau, L. Loud, and T. Van Cutsem, “Prospects of a new dynamic simulation software for real-time applications on the Hydro-Québec system,” in *2015 CIGRÉ Canada Conference*, 2015.
- [133] IEEE PES Task Force, “Test Systems for Voltage Stability Analysis and Security Assessment.” [Online]. Available: <https://resourcecenter.ieee-pes.org/technical-publications/technical-reports/PESTR19.html>
- [134] <Https://www.nationalgridus.com/Upstate-NY-Business/Supply-Costs/Load-Profiles>, “Load Profiles National Grid,” 2018. [Online]. Available: <Https://www.nationalgridus.com/Upstate-NY-Business/Supply-Costs/Load-Profiles>
- [135] T. Van Cutsem, M. Glavic, W. Rosehart, C. Canizares, M. Kanatas, L. Lima, F. Milano, L. Papangelis, R. A. Ramos, J. A. D. Santos, B. Tamimi, G. Taranto, and C. Vournas, “Test Systems for Voltage Stability Studies,” *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 4078–4087, 9 2020.
- [136] D. J. Trudnowski, J. W. Pierre, N. Zhou, J. F. Hauer, and M. Parashar, “Performance of three mode-meter block-processing algorithms for automated dynamic stability assessment,” *IEEE Transactions on Power Systems*, 2008.
- [137] D. N. Kosterev and C. W. Taylor, “WSCC System Outage,” *IEEE Transactions on Power Systems*, vol. 14, no. 3, 1999.

- [138] G. Rogers, “Power system structure and oscillations,” in *Power system oscillations*. Springer, 2000, pp. 101–119.
- [139] M. Adamiak, Z. Zhang, and I. Voloh, “Multi-range signal oscillation detection - Concepts and applications,” in *71st Annual Conference for Protective Relay Engineers, CPRE 2018*, vol. 2018-Janua, 2018, pp. 1–7. [Online]. Available: <http://prorelay.tamu.edu/wp-content/uploads/sites/3/2018/03/Multi-Range-Signal-Oscillation-Detection.pdf>
- [140] Z. Rafique, H. M. Khalid, S. Muyeen, and I. Kamwa, “Bibliographic review on power system oscillations damping: An era of conventional grids and renewable energy integration,” *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107556, 3 2022. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0142061521007912>
- [141] M. Ghorbaniparvar, “Survey on forced oscillations in power system,” *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 5, pp. 671–682, 2017.
- [142] R. B. Myers and D. J. Trudnowski, “Effects of forced oscillations on spectral-based mode-shape estimation,” in *2013 IEEE Power and Energy Society General Meeting*, 2013, pp. 1–6.
- [143] R. Xie and D. Trudnowski, “Distinguishing features of natural and forced oscillations,” in *2015 IEEE Power and Energy Society General Meeting*, 2015, pp. 1–5.
- [144] L. G. Meegahapola, S. Bu, D. P. Wadduwage, C. Y. Chung, and X. Yu, “Review on oscillatory stability in power grids with renewable energy sources: Monitoring, analysis, and control using synchrophasor technology,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 519–531, 2021.
- [145] C. Chen, W. Du, H. Wang, and T. Littler, “Sub-synchronous oscillations in power systems caused by grid-connected wind farms — a survey of mechanism studies,” *CSEE Journal of Power and Energy Systems*, vol. 4, no. 4, pp. 495–503, 2018.
- [146] V. Mohale and T. R. Chelliah, “Sub synchronous oscillation in asynchronous generators serving to wind and hydro power systems – a review,” in *2021 IEEE Industry Applications Society Annual Meeting (IAS)*, 2021, pp. 1–7.

- [147] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W. A. Mittelstadt, “Electromechanical mode online estimation using regularized robust RLS methods,” *IEEE Transactions on Power Systems*, 2008.
- [148] J. F. Hauer, C. Demeure, and L. Scharf, “Initial results in prony analysis of power system response signals,” *IEEE Transactions on power systems*, vol. 5, no. 1, pp. 80–89, 1990.
- [149] J. W. Ballance, B. Bhargava, and G. D. Rodriguez, “Monitoring power system dynamics using phasor measurement technology for power system dynamic security assessment,” in *2003 IEEE Bologna PowerTech - Conference Proceedings*, vol. 3, 2003, pp. 683–689. [Online]. Available: <https://www.researchgate.net/publication/4077954>
- [150] X. Yang and A. Fellachi, “Stabilization of inter area oscillation modes through excitation systems,” *IEEE Transactions on Power Systems*, 1994.
- [151] S. Alghamdi, N. Smith, P. Aristidou, and J. Schiffer, “Delay-Robust Distributed Secondary Frequency Control: A Case Study,” in *2019 IEEE Milan PowerTech, PowerTech 2019*, 2019.
- [152] A. R. Messina, V. Vittal, G. T. Heydt, and T. J. Browne, “Nonstationary approaches to trend identification and denoising of measured power system oscillations,” *IEEE Transactions on Power Systems*, 2009.
- [153] J. P. Amezquita-Sanchez and H. Adeli, “A new music-empirical wavelet transform methodology for time-frequency analysis of noisy nonlinear and non-stationary signals,” *Digital Signal Processing: A Review Journal*, 2015.
- [154] K. C. Lee and K. P. Poon, “Analysis of power system dynamic oscillations with beat phenomenon by fourier transformation,” *IEEE Transactions on Power Systems*, 1990.
- [155] F. Shiri and B. Mohammadi-Ivatloo, “Identification of inter-area oscillations using wavelet transform and phasor measurement unit data,” *International Transactions on Electrical Energy Systems*, vol. 25, no. 11, pp. 2831–2846, 11 2015. [Online]. Available: <http://doi.wiley.com/10.1002/etep.1994>
- [156] J. L. Rueda, C. A. Juárez, and I. Erlich, “Wavelet-based analysis of power system low-frequency electromechanical oscillations,” *IEEE Transactions on Power*

- Systems*, vol. 26, no. 3, pp. 1733–1743, 8 2011. [Online]. Available: <http://ieeexplore.ieee.org/document/5708194/>
- [157] J. Xiao, X. Xie, Y. Han, and J. Wu, “Dynamic tracking of low-frequency oscillations with improved prony method in wide-area measurement system,” in *2004 IEEE Power Engineering Society General Meeting*, 2004.
- [158] C. J. Demeure and L. L. Scharf, “Initial results in prony analysis of power system response signals,” *IEEE Transactions on Power Systems*, 1990.
- [159] J. F. Hauer, “Application of prony analysis to the determination of modal content and equivalent models for measured power system response,” *IEEE Transactions on Power Systems*, 1991.
- [160] T. Becejac and T. Overbye, “Impact of PMU data errors on modal extraction using matrix pencil method,” in *2019 International Conference on Smart Grid Synchronized Measurements and Analytics, SGSMA 2019*, 2019.
- [161] M. L. Crow and A. Singh, “The matrix pencil for power system modal extraction,” *IEEE Transactions on Power Systems*, 2005.
- [162] P. Korba, “Real-time monitoring of electromechanical oscillations in power systems: First findings,” *IET Generation, Transmission and Distribution*, 2007.
- [163] R. A. Wiltshire, G. Ledwich, and P. O’Shea, “A Kalman filtering approach to rapidly detecting modal changes in power systems,” *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1698–1706, 11 2007.
- [164] J. M. Papy, L. De Lathauwer, and S. Van Huffel, “Common pole estimation in multi-channel exponential data modeling,” *Signal Processing*, 2006.
- [165] A. M. A. Hamdan, “Investigation of the significance of singular value decomposition in power system dynamics,” *International Journal of Electrical Power and Energy System*, 1999.
- [166] N. E. Huang, “An adaptive data analysis method for nonlinear and nonstationary time series: the empirical mode decomposition and hilbert spectral analysis,” in *Hilbert-Huang Transform And Its Applications, 2nd Edition*. Springer, 2006, pp. 363–376.

- [167] N. E. Huang and S. S. Shen, *Hilbert-Huang Transform and its Applications*, 2nd Edition, ser. Interdisciplinary Mathematical Sciences. World Scientific Publishing Co., 1 2014, vol. 16. [Online]. Available: <https://www.worldscientific.com/worldscibooks/10.1142/8804>
- [168] Z. Feng, D. Zhang, and M. J. Zuo, “Adaptive Mode Decomposition Methods and Their Applications in Signal Analysis for Machinery Fault Diagnosis: A Review with Examples,” *IEEE Access*, 2017.
- [169] Z. Wu and N. E. Huang, “Ensemble empirical mode decomposition: A noise-assisted data analysis method,” *Advances in Adaptive Data Analysis*, 2009.
- [170] J. S. Smith, “The local mean decomposition and its application to EEG perception data,” *royalsocietypublishing.org*, vol. 2, no. 5, pp. 443–454, 2005. [Online]. Available: <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2005.0058>
- [171] J. Gilles, “Empirical wavelet transform,” *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3999–4010, 2013.
- [172] W. Liu, S. Cao, and Y. Chen, “Seismic time-frequency analysis via empirical wavelet transform,” *IEEE Geoscience and Remote Sensing Letters*, 2016.
- [173] S. Dong, M. Yuan, Q. Wang, and Z. Liang, “A modified empirical wavelet transform for acoustic emission signal decomposition in structural health monitoring,” *Sensors (Switzerland)*, 2018.
- [174] D. N. Kosterev, C. W. Taylor, and W. A. Mittelstadt, “Model validation for the august 10, 1996 wscc system outage,” *IEEE Transactions on Power Systems*, 1999.
- [175] Y. Y. Hsu, S. W. Shyue, and C. C. Su, “Low Frequency Oscillations in Longitudinal Power Systems: Experience with Dynamic Stability of Taiwan Power System,” *IEEE Transactions on Power Systems*, vol. PWRS-2, no. 1, pp. 92–100, 1987.
- [176] U. Bachmann, I. Erlich, and E. Grebe, “Analysis of interarea oscillations in the European electric power system in synchronous parallel operation with the Central-European networks,” *International Conference on Electric Power Engineering, PowerTech Budapest 1999*, p. 49, 1999.

- [177] P. K. Bajaria, S. R. Wagh, and N. M. Singh, “Interarea Oscillations & Chimera in Power Systems,” *undefined*, 2019.
- [178] A. Monticelli, “Electric power system state estimation,” *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.
- [179] M. Rihan, M. Ahmad, and M. S. Beg, “Vulnerability Analysis of Wide Area Measurement System in the Smart Grid,” *Smart Grid and Renewable Energy*, vol. 04, no. 06, pp. 1–7, 2013.
- [180] Y. Wang, K. Sun, X. Yuan, Y. Cao, L. Li, and H. N. Koivo, “A Novel Sliding Window PCA-IPF Based Steady-State Detection Framework and Its Industrial Application,” *IEEE Access*, 2018.
- [181] J. G. Desteeese, “A Tutorial on Detection and Characterization of Special Behavior in Large Electric Power Systems,” *Components*, 2004.
- [182] R. Bro and A. K. Smilde, “Principal component analysis,” 2014.
- [183] Y. Xu, K. Zhang, C. Ma, X. Li, and J. Zhang, “An improved empirical wavelet transform and its applications in rolling bearing fault diagnosis,” *Applied Sciences*, vol. 8, no. 12, 2018. [Online]. Available: <https://www.mdpi.com/2076-3417/8/12/2352>
- [184] R. I. Ramírez and L. A. Montejo, “On the identification of damping from non-stationary free decay signals using modern signal processing techniques,” *International Journal of Advanced Structural Engineering*, 2015.
- [185] D. S. Laila, A. R. Messina, and B. C. Pal, “A refined Hilbert-Huang transform with applications to interarea oscillation monitoring,” *IEEE Transactions on Power Systems*, 2009.
- [186] A. Shamisa, B. Majidi, and J. C. Patra, “Sliding-window-based real-time model order reduction for stability prediction in smart grid,” *IEEE Transactions on Power Systems*, 2019.
- [187] E. Foruzan, H. Sangrody, J. Lin, and D. Deepak Sharma, “Fast sliding detrended fluctuation analysis for online frequency-event detection in modern power systems,” in *2017 North American Power Symposium, NAPS 2017*, 2017.

- [188] T. Cutsem and L. Papangelis, “Description, Modeling and Simulation Results of a Test System for Voltage Stability Analysis,” *undefined*, 2013.
- [189] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, feb 2014.
- [190] J. S. Smith, “The local mean decomposition and its application to eeg perception data,” *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.

## **Appendix A**

### **appendix-A: Nordic Test System**

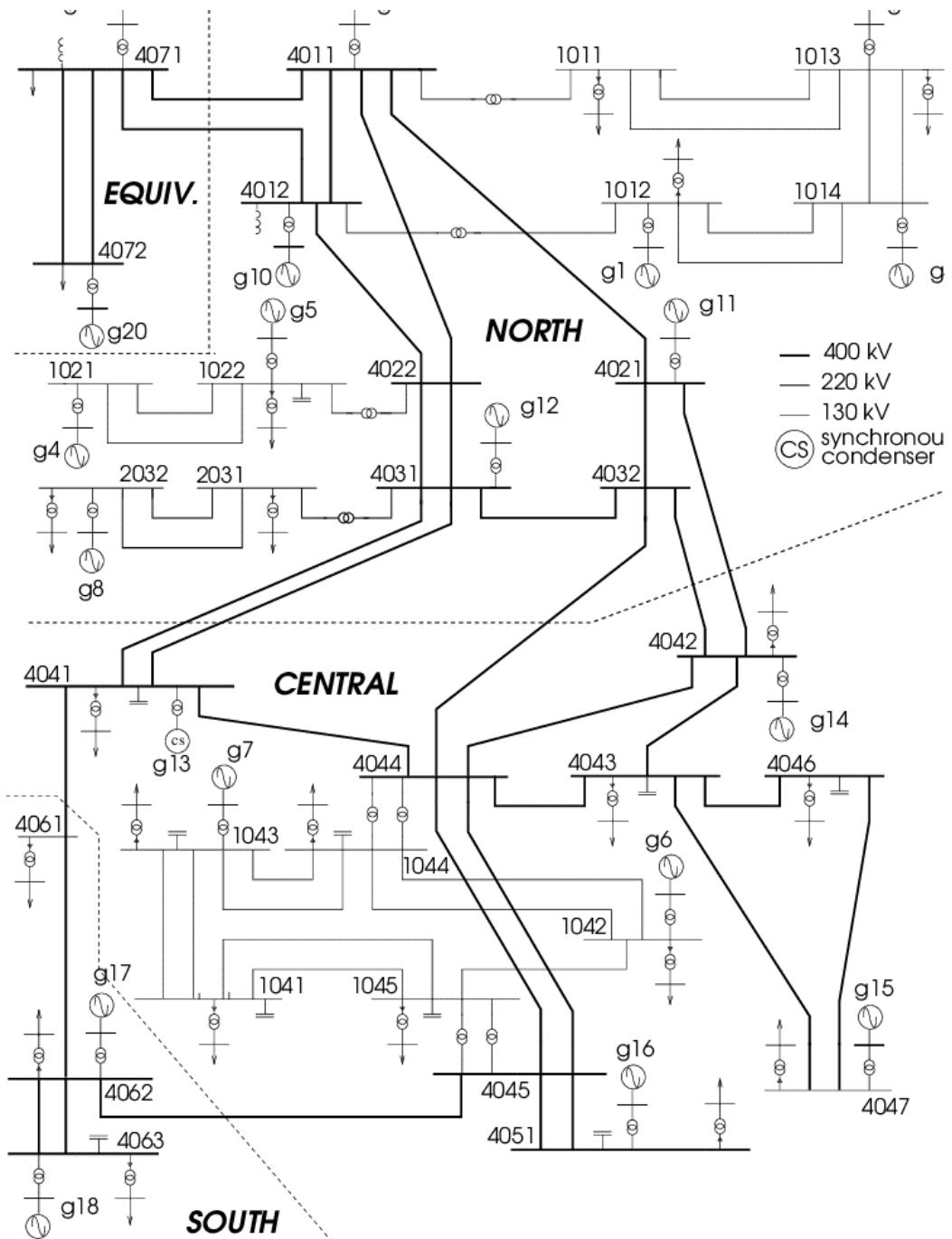


Figure A.1: Single line diagram of Nordic Test System: *Source* [3]

## **Appendix B**

### **Appendix-B: Parameter setting for machine learning model**

Feature Selector	Classifier	Parameter	Value
<b>F-Regression</b>	DT	max_features	11
<b>F-Regression</b>	DT	max_depth	16
<b>F-Regression</b>	DT	min_samples_split	5
<b>F-Regression</b>	DT	min_samples_leaf	7
<b>F-Regression</b>	KNN	neighbors	5
<b>F-Regression</b>	SVR	C	38.82
<b>VAR</b>	KNN	Threshold	0.0006
<b>VAR</b>	KNN	neighbors	7
<b>VAR</b>	SVR	Threshold	0.0002
<b>VAR</b>	SVR	C	1
<b>VAR</b>	DT	Threshold	0.00006
<b>VAR</b>	DT	max_features	11
<b>VAR</b>	DT	max_depth	11
<b>VAR</b>	DT	min_samples_split	3
<b>VAR</b>	DT	min_samples_leaf	11
<b>lasso</b>	KNN	Alpha	0.00002
<b>lasso</b>	KNN	neighbors	8
<b>lasso</b>	SVR	Alpha	0.00004
<b>lasso</b>	SVR	C	10
<b>lasso</b>	DT	Alpha	0.00008
<b>lasso</b>	DT	max_features	5
<b>lasso</b>	DT	max_depth	23
<b>lasso</b>	DT	min_samples_split	10
<b>lasso</b>	DT	min_samples_leaf	5
<b>MI-Regression</b>	DT	max_features	2
<b>MI-Regression</b>	DT	max_depth	25
<b>MI-Regression</b>	DT	min_samples_split	10
<b>MI-Regression</b>	DT	min_samples_leaf	3
<b>MI-Regression</b>	KNN	neighbors	6
<b>MI-Regression</b>	SVR	C	24.96
<b>Ridge</b>	KNN	Alpha	0.00001
<b>Ridge</b>	KNN	neighbors	7
<b>Ridge</b>	SVR	Alpha	0.00001
<b>Ridge</b>	SVR	C	10
<b>Ridge</b>	DT	Alpha	0.00009
<b>Ridge</b>	DT	max_features	12
<b>Ridge</b>	DT	max_depth	9
<b>Ridge</b>	DT	min_samples_split	7
<b>Ridge</b>	DT	min_samples_leaf	4

Table B.1: Best score for hyperparameter obtained through Bayesian optimization  
195