# Modernized Power System Optimal Operation & Safety Protection through Mathematical and Artificial Intelligence Techniques

by

## Meng Zhao

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the proposal exam of

## Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Meng Zhao

It was defended on

April 5, 2023

and approved by

Zhi-hong Mao, PhD, Professor, Department of Electrical and Computer Engineering

Alexis Kwasinski, PhD, Professor, Department of Electrical and Computer Engineering

Murat Akcakaya, PhD, Professor, Department of Electrical and Computer Engineering

Oliver Hinder, PhD, Professor, Department of Industrial Engineering

Lloyd B. Gordon, PhD, Los Alamos National Laboratory

Dissertation Director: Masoud Barati, PhD, Professor, Department of Electrical and

Computer Engineering

# Modernized Power System Optimal Operation & Safety Protection through Mathematical and Artificial Intelligence Techniques

Meng Zhao, PhD

University of Pittsburgh, 2023

The electricity industry is facing significant expectations and requirements to optimize the energy operation, secure the electrical system and critical infrastructure that depend on electric power against cyber or physical attacks, and maintain system stability. Power and energy optimization problem is one of the most fundamental concerns that must be dealt with in electrical power system management. During the operation of electric power systems, the distribution lines may experience faults caused by unpredictable natural disasters or accidents. Also, Sub-Synchronous Oscillations (SSOs) have recently emerged as a critical concern in modern power systems primarily due to the rapid proliferation of grid-connected wind power plants and installation of series-compensated transmission lines.

To create an economical and optimal power system, this thesis proposes a novel algorithm by leveraging linear approximation and convex relaxation for the non-convex AC Optimal Power Flow (ACOPF) problems, making it effective to get the global near-optimal solutions. To improve the stability and resiliency of power systems, two deep learning models are developed for fault recognition and SSO detection: a real-time deep learning model based on convolution neural network (ConvNet) is developed to classify the fault types and localize the fault with the access to the measured data, and the single observability as its superiority guarantees the performance with limited data access or measurement devices; the second deep learning model integrates ConvNet and long short-term memory (LSTM) techniques for SSO detection and evaluation that harnesses the full potential in pseudo-continuous quadrature wavelet transform (PCQ-WT).

By utilizing mathematical and artificial intelligence techniques, this thesis proposes algorithms for system optimal operation and faulty status detection to guarantee an economical and resilient power system.

# Table of Contents

# List of Tables

# List of Figures

# 1.0 Introduction

The electrical power system consists of three major components: generation, a high-voltage transmission grid, and a distribution system. With the fast increasing trend of pursuing environmental benefit and vigorously enforced emission restriction recently, the electric power system is experiencing a rapid proliferation and integration of eco-friendly wind power generation. However, sub-synchronous oscillations (SSOs) have recently emerged as a critical concern in modern power systems primarily due to the rapid proliferation of grid-connected wind power plants and installation of series-compensated transmission lines. During the operation of the power system, the distribution lines may experience some undesirable faults caused by natural hazards. These abnormal conditions would bring security risks as well as large economical loss if they were not dealt with promptly. Therefore, it is urgent to recognize and repair these faulty conditions as soon as possible. Moreover, optimal power and energy allocation is the most concerned issue during the normal operation of power system, which focuses on solving the alternating current optimal power flow (ACOPF) problems. In this context, this dissertation proposes artificial intelligence techniques for the recognition of faulty conditions, and mathematical algorithms for the optimal operation of the power system to maintain an economic, stable modernized power system.

The present chapter expounds upon the underlying impetus of this dissertation, which is centered on the normal optimal operation of bulk power systems and analyzing the power grids against faults, incidents, and contingencies through the application of optimization modeling and artificial intelligence methodologies.

## 1.1 AC Optimal Power Flow Problems

Alternating current optimal power flow (ACOPF) is an optimization problem that considers the full AC power flow equations and it exists in multiple fields of power systems. One of the most common issues is the minimization of generation cost while satisfying physical

laws and engineering constraints. Generally, ACOPF problems are non-convex due to the nonlinear power flow equations which makes the OPF problem NP-hard [3]. Due to its non-linearity and non-convexity, numerous researches have been done on ACOPF. The common algorithms based on ACOPF problems can be classified into two types: linearization and convexification.

The linearized version of ACOPF is known as DC optimal power flow (DCOPF) which is very fast and scalable. It considers the power flows in a linearized form but ignores the ohmic losses and reactive power in the system.

Due to the inaccuracy and high level of approximation in DCOPF, ACOPF-based methods got more attention with convexified methods. Among these algorithms, semi-definite programming (SDP) [4, 5] and second-order cone programming (SOCP) [6, 7] were widely used and well-established. Based on these methods, some improved algorithms such as convex iteration have developed with more efficiency [8]. The reference [4] proposed solving the non-convex ACOPF problem by dealing with its Lagrangian dual problem which is a convex semi-definite programming and recovered a primal solution from a dual optimal solution. The [8] exploited sparsity technique and strengthened SOCP relaxation of ACOPF through enforcing submatrices corresponding to maximal cliques and minimal chordless cycles in the cycle basis positive semi-definite by using convex iteration in the proposed solver. This algorithm added virtual lines in minimal chordless cycles to decompose each of them into 3-node cycles and the resulting convex relaxation had a tight gap. A penalized perturbed SDP relation is suggested in [9] to increase the real part of some entries of Hermitian matrix to ensure the rank-1 solution of this matrix at optimality condition. In particular, the larger value of regularized coefficient leads to the higher penalty on the rank of Hermitian matrix. The existence of a rank-1 SDP solution is also studied in [10] for the recovery of a global solution of SCOPF with line contingencies. They developed a graph-theoretic convex program to identify the problematic lines of the network and incorporate the loss over those lines into the objective as a penalization (regularization) term, leading to a penalized SDP problem when the relaxation is not exact. A novel parabolic relaxation in the form of convex quadratically constrained quadratic programs (QCQPs) proposed in [11] with penalization technique which is compatible with the SDP, SOCP, and parabolic relaxations and guaran-

tees the recovery of feasible solutions for OPF, under certain assumptions. These methods have two main issues: 1) in many cases the value of penalized factor and term are predefined and set based on heuristics; 2) the regularized terms change the value of original objective cost function in original OPF as these additional terms do not disappear in global optimal solution.

However, these algorithms cannot guarantee global optimum [12] in ideal situation. Therefore, it is urgent and important to develop a globally optimal algorithm for ACOPF problems to tackle the aforementioned issues. Inspired by the penalized SDP algorithm, we add a penalty term in the objective function that push the rank of the Hermitian matrix $W$ to 1, which will be disappeared in the objective function at global optimum solution. Our penalized SDP formulation will be crafted as a canonical quadratic programming with linear constraints (LCQP) problem. Finally, we propose a successive linear programming (SLP) and SLPBB (Branch & Bound) algorithms to solve this LCQP problem find a $\epsilon$-approximate solution.

## 1.2  Fault Recognition

During the operation of electric power systems, the distribution lines may experience faults caused by unpredictable natural disasters or accidents. These faults can induce supply interruptions and fire hazards which are dangerous to people and may cause enormous economical losses. To avoid these dangers and losses, the protective relays and other control devices were widely installed to support circuit breakers to take appropriate actions, and minimize damage and length of interruption at the maximum desired resiliency level [13]. However, with the increased number of faults and incidents occurring in power networks due to natural hazards and technical system issues, the behavior and effect of faults in wildfires become more complex. As a result, the faulty assets may not only cause the incident energy and sparks but also lead to the misoperation and breakouts of the circuit breakers [14]. The faults in distribution systems can cause tremendous disasters to the power equipment which is a big loss to the companies. Also, there are potential safety hazards that existed
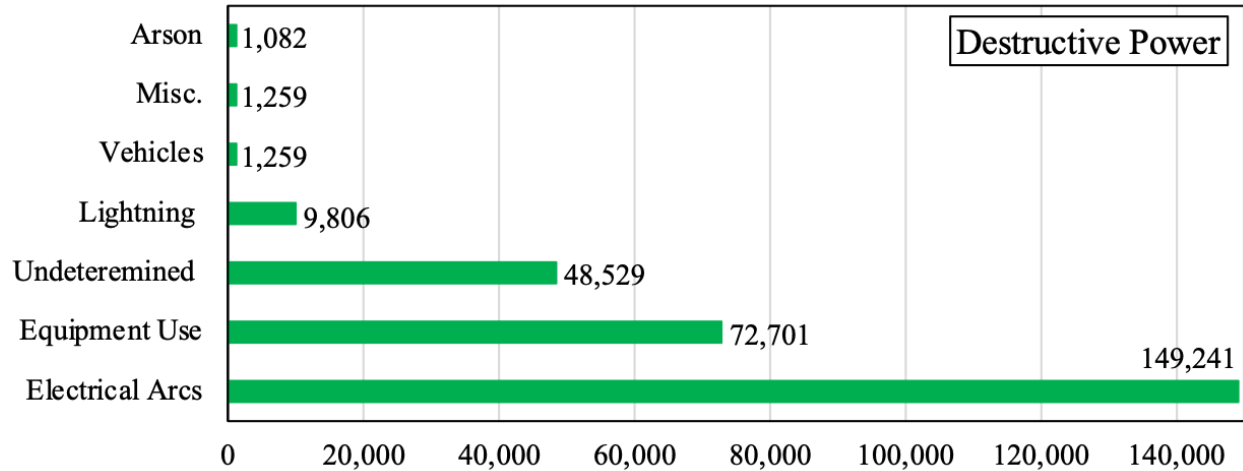
Figure 1: Wildfire sparked by power lines and electrical equipment burned the most acreage in California in 2015 [1].

in them such as fire hazards and explosions that can pose much risk to our society. The explosive failure of overhead lines and other electrical equipment has frequently rated among the top three singular origins of California wildfires for the last several years. Figure 1 shows the power-related wildfire reported data in 2015, electrical power problems sparked the burning of 149,241 acres which was greater than almost twice the amount from any other causes. Therefore, accurate fault localization and quick potential fire detection are urgent and necessary in power distribution grids.

Sparking and arc flash, as defined by [15], is "an electric arc event with thermal energy dissipated as radiant, convective, and conductive heat." The electric arc can form, due to equipment malfunction or failure, poor operation and maintenance, tools dropped while energized, tools left in equipment during de-energized maintenance, animals, and poor electrical connections that cause a fault to occur and allow current to flow through the air. Depending upon the severity, it can cause a wildfire and then injure or kill people. In the USA, the California Department of Forestry and Fire Protection reported that these sparking and arc flashes started a fire when trees and branches touched the overhead distribution lines. One such fire, the Redwood Fire, burned over 36,000 acres, destroyed 543 structures, and resulted
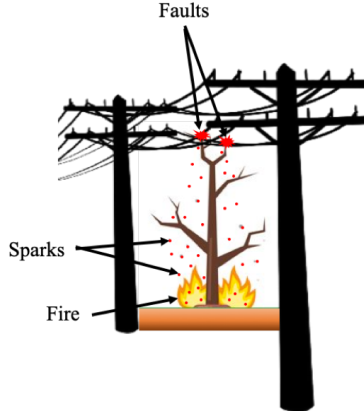
Figure 2: Wildfire created by fault incident energy.

in nine civilian deaths [16].

Figure 2 indicates a simple electrical fault caused by trees, their branches, and dry grass that can cause sparking, arc flash, and finally wildfire due to the released incident energy. Therefore, the wildfire is set by incident energy as a function of the electrical faults.

Generally, there are four main fault types that occur in distribution systems: single line-to-ground fault (LG), line-to-line fault (LL), double line-to-ground fault (LLG), and three-phase to-ground fault (LLLG). In recent years, lots of research and methods have been put forward to fault localization in power systems. These methods can be classified into two types: conventional techniques and artificial intelligent techniques.

Conventional techniques include the traveling wave-based method and impedance-based methods. The authors in [17] proposed a new low-cost device that is part of a system that should include several units installed in couples at the two ends of MV cable lines for fault location in the power distribution system. The system exploited a two-end traveling wave technique that depends on data synchronization. However, it is worth noting that the traveling wave-based method is sensitive to inaccuracies in electrical line parameters and it is costly because of the requirements of GPS devices and sensors. [18] introduced a new impedance-based fault location scheme for overhead unbalanced radial distribution networks using Clarke transformation of voltage and current measurements at one-end substation

only. Also, it calculated the distance to the fault for all fault types by using one quadratic formula and two groups of constants. However, this method needs modal transformation for the measured voltages and currents, and the fault distance $d$ is obtained by solving a quadratic formula which needs further analysis to delete one impossible solution. Although it is more economical to implement an impedance-based method, the accuracy of this method is vulnerable to the system characteristics, measurement errors and other factors.

Due to the complexity of distribution systems and various uncertainty factors that are difficult to address by using conventional techniques, a knowledge-based technique was applied for locating faults recently [19]. In general, the technique requires information such as feeder measurement, substation, feeder switch status, and some additional information provided by fault detection devices installed along the feeders and atmospheric conditions such as Digital Fault Recorder (DFR) and/or Remote Fault Indicator (RFI). This information can be then analyzed by using artificial intelligence methods [20, 21, 22, 23]. Several techniques have been applied in fault localization and detection in distribution systems such as, Decision Tree (DT) methodology [24, 25], Artificial Neural Network (ANN) [26], Support Vector Machine (SVM) [27], Fuzzy Logic [28], Convolutional Neural Network (ConvNet) [29] and so on. [24] proposed the DT approach in classifying and detecting fault signals which were extracted by the wavelet-decomposition method. In [25], the J48 DT algorithm was used to locate the single-phase faults that occurred in a 34-bus distribution system. But this paper located the fault based on the knowledge of only one fault type. [30] aimed at high impedance fault (HIF) detection and location in overhead distribution feeders by using the proposed enhanced FTU (feeder terminal unit). The enhanced FTU which is the retrofitting of the ready-made FTU with an embedded digital signal processor (DSP) and a high-frequency current sensor includes wavelet transform for extracting the features of the HIF signals and neural network for HIF detection. In [31], the authors proposed a faulted line localization method based on ConvNet classifier using bus voltages and it is able to localize the faulted line to a small neighborhood with high probability under low observability of the buses by using phasor measurement units (PMU) placement strategy. Compared with neural networks or ConvNet algorithms, DT algorithm is non-robust because of the non-parametric property: a small change in the training data can result in a large change in the tree and

6

consequently the final predictions. Also, the problem of learning an optimal decision tree is NP-complete. ConvNet is a transfer-learning algorithm that it learns the hidden relationship between the input features and output labels, resulting in a more robust and universal model. In fact, the papers mentioned above mostly focused on one side: locating the fault or fault type classification but ignored the special features of the fault type which are essential in fault localization and calculation of the released energy that can contribute to wildfire management.

In this research, a new fault recognition algorithm based on ConvNet technique was proposed, resulting in fault type classification and localization. The main contributions of this algorithm are summarized in four aspects: 1) feature vectors were extracted from the measured voltages and currents with real phases $(a, b, c)$ to determine the fault types by using ConvNet Model-C; 2) the positive, negative and zero asymmetrical component sequences were added in the feature vectors to locate the fault occurred in the system with ConvNet Model-L. These three sequences have many contributions to the unbalanced faults; 3) the presented intelligent models are established based on training and testing dataset which is a function of the fault location $d$. This functionality creates an extensive dataset that can lead to accurate fault localization; 4) in terms of the limited number of measurements (sensor) in the distribution system, our method possesses the Single Observability characteristic: we train and test the feature vectors extracted from each single measurement device. Therefore, we may do classification and localization by having the information of a single measurement device.

### 1.3   Sub-Synchronous Oscillation Detection

With the trend of pursuing environmental benefits and vigorously enforced emission restrictions in the modern world, the electric power system is experiencing a rapid proliferation and integration of renewable energy. As a consequence, the power system dynamics are affected greatly by the increasing renewable penetration. One of the undesirable consequences is a sub-synchronous oscillation (SSO), which is caused by the interaction between renewable

energy sources and power system components. The first sub-synchronous torsional vibration event that occurred in Square Butte in North Dakota (1977) has resulted from the interaction of High Voltage Direct Current (HVDC) controls with a turbine generator. Recently, more and more destructive incidents by SSOs were reported globally.

Generally, SSO can be classified into four categories: sub-synchronous resonance (SSR), device-dependent SSO (DDSSO), sub-synchronous control interaction (SSCI), and sub-synchronous torsional interaction (SSTI). Among them, SSR and SSCI are the most commonly occurring events. SSCI typically has a fast-growing nature but does not involve any kind of mechanical interactions nor fixed oscillation frequencies [32, 33]. Also, SSR raised by the poorly damped oscillations around the electrical resonant frequency, or the natural mechanical torsional oscillation can cause a slow but intensive energy exchange between generator and the power grid, which results in growing fatigue and critical failure on the generator turbine shaft. Therefore, an effective strategy for quickly detection and mitigating the SSCI and SSR phenomena is of significant importance.

To address the SSCI and/or SSR detection problem, [34] applied Discrete Time Fourier Transform (DTFT) to the sampled waveform data to directly extract the frequency and amplitude of SSCI. However, the trade-off between the high frequency resolution and the long-time delay of the data observation window will limit the performance of this approach. Also, this DTFT method was not tested in a real-time operation condition. In [35], an SSR detection method through comparing the half-cycle time of positive or negative voltage signal with the damping and trip thresholds was proposed. [36] proposed a Kalman Filter (KF) based SSR detection and mitigation scheme, where KF was used for sub-synchronous components state estimation and a KF-damping controller was developed to select and suppress the frequency sub-synchronous components. An SSR detection and characterization algorithm was proposed in [37], where, the voltage and current waveforms captured from two ends of the series capacitor compensated transmission line were used to identify the SSR current magnitude, frequency, an event and its characteristics. However, two waveform measurement units at both ends were required which means it cannot work on waveforms captured from single end. In [38], a dynamic frequency scanning method considering the wind turbine generator nonlinearities was derived, and then it was applied for evaluating and analyzing

the wind turbine SSCI phenomenon. At the same time, the electrical damping analysis was developed through approximating the electromagnetic torque at low frequencies. A wide area protection scheme against unstable SSCI was proposed in [39], where distributed protection relays measured the local wind farms' real-time sub-synchronous impedances. Then, a centralized protection coordinator collected those impedances and determined the tripping action for each wind farm. However, the SSCI detection time was 0.5s which is relatively slow and it will take another 0.1s to trip. A time-series analysis approach through non-linear dynamics originated method was proposed in [40] for detecting sub-synchronous oscillation in power systems. In this method, nonlinear oscillation was evaluated through the maximum Lyapunov exponent without filtering. This approach is capable of deciding whether the SSCI phenomenon tends to damp or not, but it still relies on a threshold setting and has to wait for more than one second to provide the detection results.

As discussed above, many approaches have been proposed to mitigate the SSCI or SSR phenomenons. However, the actuation of any migration actions relies on the accurate detection of the existence of any type of Sub-synchronous Oscillations (SSOs). Also, the escalation or decline of SSO amplitude can also be used as an indicator to instruct the trip action. Therefore, the ultimate problem of SSO detection can be melted down into the identification of the SSO occurrence and evaluation of SSO amplitude trend. Inspired by the concept of waveform pattern extraction techniques [41, 42, 43], the event detection and prediction using artificial intelligence [44], we proposed a rapid SSO detection which provides:

- Fast SSO detection compared with the conventional approaches.
- Accurate SSO amplitude trend evaluation and prediction of the growth of SSO
- Isolation command to protective relays for preventing wind turbines and other energy resources from SSO damage.

Consequently, the trip action of wind turbines under SSO will not rely on a fixed threshold, and then, the development of fatigue in the wind turbines' mechanical and the risk of damaging electrical components can be potentially reduced. This proposed scheme gathers the original waveforms locally at the node of interest in the power grid and achieves an online surveillance mechanism with fast and accurate event detection and prediction.

The rest of this thesis is organized as follows: In Chapter 2, the ACOPF, fault recognition and SSO detection problems are introduced; Chapter 3 illustrates the proposed mathematical and intelligent models for these problems; The simulation results are shown in Chapter 4 and conclusions are summarized in Chapter 5;

## 2.0 Problem Formulation

Based on the discussions in Chapter 1, this chapter will discuss about the problem formulations of ACOPF, fault recognition in distribution system and SSO detection which include: the classical ACOPF problem considered in this research, as well as feature extractions for these two faulty conditions.

## 2.1 ACOPF Problem

To solve ACOPF problems, one of the most common formulations: minimizing the generation cost while satisfying physical laws and engineering constraints was considered. In this case, the formulation of an $n_b$-bus power system is as follows:

$$
\begin{aligned}
\min \quad & \sum_{k \in \mathcal{G}} f_k(P_{Gk}) \\
s.t. \quad & P_{G_k} - P_{D_k} = V_{dk} \sum_{i=1}^{n_b} (G_{ik}V_{di} - B_{ik}V_{qi}) \\
& \qquad\qquad + V_{qk} \sum_{i=1}^{n_b} (B_{ik}V_{di} + G_{ik}V_{qi}), \\
& Q_{G_k} - Q_{D_k} = V_{dk} \sum_{i=1}^{n_b} (-B_{ik}V_{di} - G_{ik}V_{qi}) \\
& \qquad\qquad + V_{qk} \sum_{i=1}^{n_b} (G_{ik}V_{di} - B_{ik}V_{qi}), \\
& P_k^{min} \leq P_{G_k} \leq P_k^{max}, Q_k^{min} \leq Q_{G_k} \leq Q_k^{max}, \\
& (V_k^{min})^2 \leq V_{dk}^2 + V_{qk}^2 \leq (V_k^{max})^2, k \in \mathcal{N}.
\end{aligned}
\tag{2-1}
$$

where $\mathcal{G}$ is the set of generator buses, $\mathcal{N} = \{1, 2, ..., n_b\}$ is the set of all buses in the system. $P_{Gk}$, $P_{Dk}$, $Q_{Gk}$, $Q_{Dk}$ are the active and reactive power of generators and demands, respectively. $V_{dk}$ and $V_{qk}$ are the real parts and imaginary parts of the complex voltages $V_k$, $G_{ik}$ and $B_{ik}$ are the real parts and imaginary parts of the admittance matrix Y.

In (2-1), the quadratic voltage forms in the active and reactive power flow equations make the ACOPF problems non-convex and NP-hard. To solve this non-convex and nonlinear ACOPF problem, we adopted the penalized semi-definite programming (SDP) algorithm from [9] which is introduced in the next section.

## 2.2 Fault Recognition Problem - Feature Vector

An $n$-bus power distribution grid is considered in this research, where a single line may face one of the common faults: LG, LL, LLG, and LLLG. We focus on real-time fault and possible fire localization in the power distribution system. Therefore, the data from measurement devices such as PMU (phasor measurement units), DFR, and RFI will be used before and during the fault occurrence to extract the feature vectors, and Deep CNN (**ConvNet**) algorithm will be applied on them. This section will introduce the extraction of the feature vectors which are used for the CNN algorithm.

Suppose that the fault occurs at the line between bus $i$ and bus $j$, as shown in Figure 3 $(a)$. It is clear that the impedance between these two buses will change when there is a fault occurred between them. Meanwhile, the voltage and current at these two buses will also be affected. Based on Ohm's Law, the bus voltage $\boldsymbol{V}^{abc-pref} \in \boldsymbol{C}^n$ and bus current $\boldsymbol{I}^{abc-pref} \in \boldsymbol{C}^n$ in the steady state has the following relationship,

$$\boldsymbol{I}^{abc-pref} = \boldsymbol{Y}^{abc-pref}\boldsymbol{V}^{abc-pref} \tag{2-2}$$

where $\boldsymbol{Y}^{abc-pref} \in \boldsymbol{C}^{n \times n}$ is the admittance matrix of the system and can be expressed as,

$$\boldsymbol{Y}^{abc-pref} = \begin{bmatrix} Y_{11}^{abc} & \dots & Y_{1i}^{abc} & \dots & Y_{1j}^{abc} & \dots & Y_{1n}^{abc} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{i1}^{abc} & \dots & Y_{ii}^{abc} & \dots & Y_{ij}^{abc} & \dots & Y_{in}^{abc} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{j1}^{abc} & \dots & Y_{ji}^{abc} & \dots & Y_{jj}^{abc} & \dots & Y_{jn}^{abc} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ Y_{n1}^{abc} & \dots & Y_{ni}^{abc} & \dots & Y_{nj}^{abc} & \dots & Y_{nn}^{abc} \end{bmatrix}. \tag{2-3}$$

12
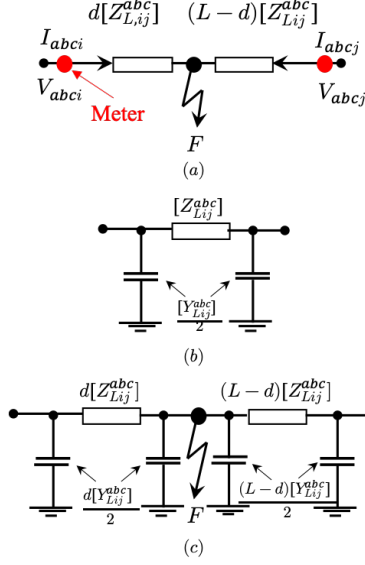
Figure 3: Faulted feeder model with fault at distance $d$: $(a)$ Location of measurement devices in faulty feeder, $(b)$ Pre-fault nominal-$\pi$ feeder model of affected section, $(c)$ Faulted nominal-$\pi$ feeder model of affected section.

When the fault occurs at the line between bus $i$ and bus $j$, the new admittance matrix $\boldsymbol{Y}^{abc-f}$ has the following format,

$$
\left[
\begin{array}{ccccccc|c}
Y_{11}^{abc} & \dots & Y_{1i}^{abc} & \dots & Y_{1j}^{abc} & \dots & Y_{1n}^{abc} & \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \\
Y_{i1}^{abc} & \dots & Y_{ii}'^{abc} & \dots & Y_{ij}'^{abc} & \dots & Y_{in}^{abc} & \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \boldsymbol{y}_f^{abc} \\
Y_{j1}^{abc} & \dots & Y_{ji}'^{abc} & \dots & Y_{jj}'^{abc} & \dots & Y_{jn}^{abc} & \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \\
Y_{n1}^{abc} & \dots & Y_{ni}'^{abc} & \dots & Y_{nj}'^{abc} & \dots & Y_{nn}^{abc} & \\
\hline
\multicolumn{7}{c|}{\boldsymbol{y}_f^{abc\ T}} & Y_{ff}^{abc}
\end{array}
\right]
$$

where,

$$
\boldsymbol{y}_f^{abc} = [0 \ \dots \ Y_{i(n+1)}'^{abc} \ 0 \ \dots \ Y_{j(n+1)}'^{abc} \ 0 \ \dots \ 0]^T \in \boldsymbol{C}^{n \times 1} \tag{2-4}
$$

which is the admittance between the faulted point and the $n$ buses. $Y_{ff}^{abc}$ is the self-admittance of the faulted point, and the changed admittance can be represented by,

$$
\begin{aligned}
Y'^{abc}_{ij} &= Y'^{abc}_{ji} = 0, \\
Y'^{abc}_{ii} &= Y^{abc}_{ii} - \frac{Y^{abc}_{Lij}}{2} - \frac{1}{Z^{abc}_{Lij}} + \frac{d \cdot Y^{abc}_{Lij}}{2} + \frac{1}{d \cdot Z^{abc}_{Lij}}, \\
Y'^{abc}_{jj} &= Y^{abc}_{jj} - \frac{Y^{abc}_{Lij}}{2} - \frac{1}{Z^{abc}_{Lij}} + \frac{(1-d) \cdot Y^{abc}_{Lij}}{2} + \frac{1}{(1-d) \cdot Z^{abc}_{Lij}}, \\
Y'^{abc}_{i(n+1)} &= Y'^{abc}_{(n+1)i} = -\frac{1}{d \cdot Z^{abc}_{Lij}}, \\
Y'^{abc}_{j(n+1)} &= Y'^{abc}_{(n+1)j} = -\frac{1}{(1-d) \cdot Z^{abc}_{Lij}}, \\
Y^{abc}_{ff} &= \frac{Y^{abc}_{Lij}}{2} + \frac{1}{d \cdot Z^{abc}_{Lij}} + \frac{1}{(1-d) \cdot Z^{abc}_{Lij}},
\end{aligned}
\tag{2-5}
$$

where $d$ is the *fault distance* between the faulted point and bus $i$, and also the length between bus $i$ and bus $j$ is supposed to be one per-unit $L = 1(pu)$. Therefore, the difference between the pre-fault voltage, current and during-fault voltage, current of each phase at bus $i$ and bus $j$ can be written as a function of impedance matrix's elements in $\mathbf{Z}^f(d) = \mathbf{Y}^{f^{-1}}(d)$ and $\mathbf{Y}^f(d)$,

$$
\begin{aligned}
\Delta V_{i_{abc}} &= V^f_{i_{abc}} - V^{pre-f}_{i_{abc}} = Z'^{abc}_{ii}(d)I'_i - Z^{abc}_{ii}(d)I_i - Z^{abc}_{ij}(d)I_j, \\
\Delta V_{j_{abc}} &= V^f_{j_{abc}} - V^{pre-f}_{j_{abc}} = Z'^{abc}_{jj}(d)I'_j - Z^{abc}_{jj}(d)I_j - Z^{abc}_{ij}(d)I_i, \\
\Delta I_{i_{abc}} &= I^f_{i_{abc}} - I^{pre-f}_{i_{abc}} = Y'^{abc}_{ii}(d)V'_i - Y^{abc}_{ii}(d)V_i - Y^{abc}_{ij}(d)V_j, \\
\Delta I_{j_{abc}} &= I^f_{j_{abc}} - I^{pre-f}_{j_{abc}} = Y'^{abc}_{jj}(d)V'_j - Y^{abc}_{jj}(d)V_j - Y^{abc}_{ij}(d)V_i.
\end{aligned}
\tag{2-6}
$$

### 2.2.1 Feature Vector for Fault Classification

The voltage and current differences before and during the fault in (2-6) have contribution to the fault type classification and fault localization based on the admittance differences. Consequently, the feature vector for the fault type classification is defined as,

$$
\varphi_t = \{\Delta \mathbf{V}(d), \Delta \mathbf{I}(d)\}
\tag{2-7}
$$

14

where, $\boldsymbol{\varphi}_t$ is a complex variable, both of its real and imaginary parts have effect on the fault type and localization. Therefore, we redefine the feature vector for fault type classification with the magnitude voltage difference, magnitude current difference and their phase angles' difference of each phase in (2-8),

$$\boldsymbol{\varphi}_t = \{|\Delta \boldsymbol{V}^{abc}(d)|, |\Delta \boldsymbol{I}^{abc}(d)|, \Delta \boldsymbol{\theta}_V^{abc}, \Delta \boldsymbol{\theta}_I^{abc}\}. \tag{2-8}$$

### 2.2.2 Feature Vector for Fault Localization

Except the feature vector $\boldsymbol{\varphi}_t$ for the fault type classification, we also realized that the zero $(Z)$, negative $(N)$ and positive $(P)$ sequences $(Z, N, P)$ contain valuable information that can improve the accuracy of fault localization. The main reason is that the behavior of the unbalanced faults, LG, LL, and LLG in $(Z, N, P)$ domain in terms of special magnitude changes and phase angle rotations can provide an additional knowledge about the fault position. The computation of these three sequences on $\chi = V$ or $I$ (phase voltages and phase currents) is,

$$\begin{aligned}
\chi_{a0} &= \frac{\chi_a + \chi_b + \chi_c}{3} \\
\chi_{a1} &= \frac{\chi_a + a\chi_b + a^2\chi_c}{3} \\
\chi_{a2} &= \frac{\chi_a + a^2\chi_b + a\chi_c}{3}
\end{aligned} \tag{2-9}$$

where $a = 1\angle 120° = -0.5 + j0.866$, and $a^2 = 1\angle 240° = -0.5 - j0.866$. Therefore, the *augmented feature vector* for fault localization is defined as,

$$\boldsymbol{\varphi}_l = \{|\Delta \boldsymbol{V}^{abc}|, |\Delta \boldsymbol{I}^{abc}|, \Delta \boldsymbol{\theta}_V^{abc}, \Delta \boldsymbol{\theta}_I^{abc}, \boldsymbol{V}^{ZNP}, \boldsymbol{I}^{ZNP}\}. \tag{2-10}$$

## 2.3   Sub-Synchronous Oscillation: Waveform Modeling

The electrical power waveform (voltages and currents) captured at the point of measurement carries valuable information of the undergoing events in the power grid. Based on the study in [37], the magnitudes of fundamental frequency components generally remain constant during the SSO phenomenon establishes, and the voltage in each phase is mainly determined by the grid voltage. Therefore, current waveforms can be used to examine the existence of sub-synchronous current which reflects the intensity of the SSR. The sampled current waveform through analog to digital conversion can be written as

$$i(t) = I_1 cos\left(\omega_1 t + \phi_1\right) + I_{ss} cos\left(\omega_{ss} t + \phi_{ss}\right) \tag{2-11}$$

where, $I_1$, $\omega_1$, and $\phi_1$ are respectively the fundamental frequency component's magnitude, frequency, and phase angle. And $I_{ss}$, $\omega_{ss}$, and $\phi_{ss}$ are respectively the sub-synchronous component's magnitude, frequency, and phase angle respectively. One can see, if the second term (sub-synchronous current) on the right hand side of (2-11) is obtained, the SSO detection can be easily solved. However, in practice, only $i(t)$ is available. Therefore, the main target in the proposed analytics will be focusing on the SSO pattern extraction process.

### 2.3.1   Wavelet Transform

Short-time Fourier transform (STFT) via a slide window can be used to calculate the sub-synchronous component in real time if the time efficiency requirement is low. However, it would be time-consuming since a higher STFT resolution requires larger buffer length under the same sampling rate which compromises the time resolution. A higher sampling rate can be achieved but at the cost of a high computational complexity [45, 46], which still slows down the processing speed. On the contrary, if the frequency measurement is not required, wavelet transform can be used for waveform feature extractions with a low requirement of frequency resolutions, and meanwhile, it requires a relatively lower computing complexity [47].

In Figure 4, a comparison of the spectrogram from STFT and scalogram from Morlet WT is presented. Obviously, from the spectrum in Figure 4a, one can find that, the STFT can accurately extract the frequency components within the waveform, i.e the yellow bar is the fundamental frequency and the light blue bars are the harmonic components. Also, the scalogram in Figure 4b can tell the existence of the harmonic components, however, the frequency measurement accuracy is compromised in WT. In the SSO detection application, the existence of the sub-synchronous component, which ranges from 5Hz to 55Hz in a 60Hz system, is of much interest. Therefore, the scalogram of current signal can be used to detect the SSO.



(a) STFT                    (b) Morlet
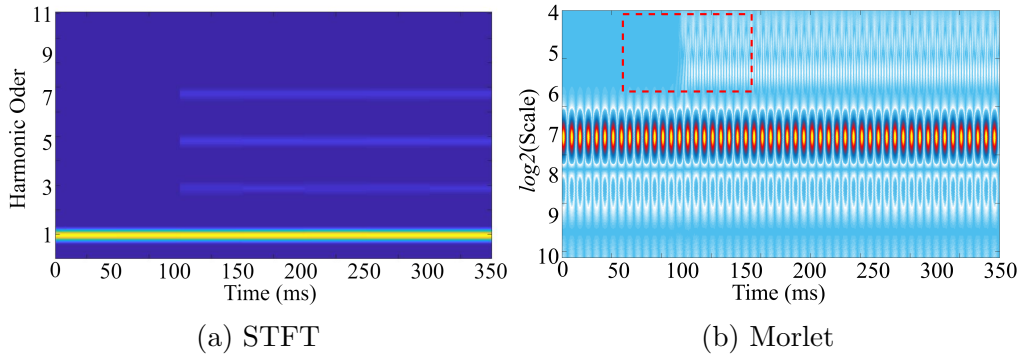
Figure 4: STFT vs Morlet WT under harmonic distortion condition (start at t=100ms) with harmonic order 3, 5, 7.

The wavelet transform can be seen as a linear transformation of the signal of interest $x(t)$ with a series of basis functions, where $x(t)$ is the measured single-phase waveform. The WT calculation is defined as follows:

$$X(\omega|a,b) = \int_{-\infty}^{\infty} x(t)\Psi^*(\frac{t-b}{a})dt \tag{2-12}$$

where $a$, $b$ are the scaling factor and time shift; $\Psi(t)$ is the scaled and shifted basis function, i.e., the selected wavelet. When $a = 1$ and $b = 0$, $\Psi(t)$ is called mother wavelet. And "*" denotes the complex conjugate operator. With different selected values of $a$ and $b$, $\Psi(\frac{t-b}{a})$ becomes the "daughter wavelets" of $\Psi(t)$ [48, 49], and $a$ determines a range of pseudo

17

frequency. Typically in digital computing devices, discrete WT is used and (2-12) can be written as following:

$$X[\omega|a_k, b_k] = \sum_{n=0}^{W-1} x[n]\Psi^*[\frac{nT_s - b_k}{a_k}]$$

(2-13)

where $T_s$ is the sampling time, and $W$ stands for sampled waveform data in the buffer. A buffer with larger size will bring more computing burden and buffer size will directly affect the feature extraction performance. Therefore, with a fixed-length buffer, selecting a proper type of wavelet is of great importance. Tuning the parameters correctly also matters, since they will affect the feature extraction performance and the computing speed. Meanwhile, one should notice that the real-time feature extraction and event detection require un-congested real-time data streaming and processing.

## 3.0    Proposed Computational Techniques

To create a stable and robust power system, a mathematical algorithm is proposed to solve the non-convex ACOPF problems. The solutions of ACOPF problem represent the steady states of the system which can be used to recognize the faulty conditions: an intelligent model based on convolutional neural network is developed to classify and localize the faults occured in distribution system; combined the convolutional neural network with long short-term memory, an SSO detection model is developed to detect and predict SSO events which usually occur in generation & transmission systems.

## 3.1    Penalized SDP Relaxation for ACOPF Problems

To solve the non-convex and nonlinear ACOPF problems, we first relax the general ACOPF problem (2-1) by using semi-definite programming (SDP) that linearize the constraints, transferring the solution goal to get the rank-1 solution of the new-generated positive semi-definite $W$ matrix. Then some penalized terms are added in the objective function that aim to push getting the rank-1 solution. As a result, the original ACOPF problem is transferred to a quadratic programming with linear constraints (LCQP). For solving SDP problems, [50] proposed a nonlinear programming via low-rank factorization. [51] developed a framework to compute structured low-rank approximations of a matrix from a random linear image with lower dimension than the original matrix. These algorithms proposed good theoretical improvements and proofs but they are not practical to be used in power system problems since they rather ignore the constraints or need to solve some sub-problems. For solving LCQP problems, [52, 53] developed branch&bound (B&B) approach for QP using SDP. [54] further adopted the so-called completely co-positive program to improve the B&B approach for LCQP. As reported in [54], these B&B-based global solvers can solve many small-scale LCQP instances, but may run out of time/memory and return only a sub-optimal solution when the problem size is large.

The penalty terms in the objective function make the problem to be nonlinear again and the quadratic terms may have some negative eigenvalues. Therefore, Difference Convex and linear programming are utilized to transfer the objective into a bi-convex format. To solve this bi-convex problem, we propose a new successive linear programming (SLP) algorithm which is different from [55] (where the nonlinear terms must be differentiable and the gradients are used for linearization). The proposed SLP algorithm cannot find a global optimum when the number of negative eigenvalues is larger than 1. To tackle this issue, we propose the SLPBB algorithm which combines the SLP with a new B&B method to get the optimal solutions of the LCQP problem.

### 3.1.1 SDP Relaxation

To start with, we first define our notations: $e_k$ is the $k$-th standard basis vector, $v = [V_{d1}, V_{d2}, ..., V_{dn_b}, V_{q1}, V_{q2}, ..., V_{qn_b}]^T$ is the voltage vector of an $n_b$-bus power system. Let $Y_k = e_k e_k^T Y$, $W = vv^T$ which is a $2n_b \times 2n_b$ symmetric matrix, then (2-1) can be reformulated as,

$$
\begin{aligned}
\min \quad & \sum_{k \in \mathcal{G}} f_k(P_{Gk}) \\
s.t. \quad & P_{G_k} - P_{D_k} = tr(\mathrm{Y}_k W), Q_{G_k} - Q_{D_k} = tr(\overline{\mathrm{Y}}_k W) \\
& P_k^{min} \le P_{G_k} \le P_k^{max}, Q_k^{min} \le Q_{G_k} \le Q_k^{max}, \\
& (V_k^{min})^2 \le W_{kk} + W_{(k+n_b)(k+n_b)} \le (V_k^{max})^2, \\
& rank(W) = 1, W \succeq 0, k \in \mathcal{N}.
\end{aligned} \tag{3-1}
$$

where $tr$ indicates the matrix trace operator (i.e., sum of the diagonal elements), $\succeq$ implies that $W$ is a positive semi-definite matrix. $W_{kk}$ is the $k$-th diagonal term in $W$ matrix that satisfy $W_{kk} = V_{dk}^2$, $W_{(k+n_b)(k+n_b)} = V_{qk}^2$ and

$$
\mathrm{Y}_k = \frac{1}{2} \begin{bmatrix} Re(Y_k + Y_k^T) & Im(Y_k^T - Y_k) \\ Im(Y_k^T - Y_k) & Re(Y_k + Y_k^T) \end{bmatrix},
$$

$$\overline{Y}_k = -\frac{1}{2} \begin{bmatrix} Im(Y_k + Y_k^T) & Re(Y_k^T - Y_k) \\ Re(Y_k^T - Y_k) & Im(Y_k + Y_k^T) \end{bmatrix}.$$

In this case, the solutions that satisfy the $rank - 1$ standard of symmetric $W$ matrix can be identified as the optimal solutions of the ACOPF problems.

**Proposition 3.1.1.** *The $rank-1$ solutions of $W$ matrix can be recovered by the second-order determinants of $W$, since we have,*

1. *If the rank of $W$ matrix is 1, it implies that the ranks of all second-order determinants of $W$ also equal to 1.*

2. *Also, the ohmic losses constraints can guarantee that the second-order terms are non-negative.*

*Proof.* Statement 1 follows easily with the definition of the matrix rank.

About Statement 2, first consider the line loss between bus $i$ and $j$ which can be denoted as,

$$\begin{aligned} S_{loss,ij} &= P_{loss,ij} + jQ_{loss,ij} \\ &= g_{ij}(W_{ii} + W_{jj} - 2W_{ij}) \\ &\quad - jb_{ij}(W_{(i+n)(i+n)} + W_{(j+n)(j+n)} \\ &\quad - 2W_{(i+n)(j+n)}), \end{aligned}$$

where $g_{ij} > 0$ and $b_{ij} < 0$ can be derived from the admittance $Y_{ij} = -\frac{1}{z_{ij}} = -g_{ij} - b_{ij}$.

Since the ohmic loss $P_{loss,ij}$ and reactive loss $Q_{loss,ij}$ in the system are always non-negative, it implies that,

$$W_{ii} + W_{jj} - 2W_{ij} \geq 0,$$

$$W_{(i+n)(i+n)} + W_{(j+n)(j+n)} - 2W_{(i+n)(j+n)} \geq 0,$$

where $W_{ii}$, $W_{jj}$, $W_{(i+n)(i+n)}$ and $W_{(j+n)(j+n)}$ are all non-negative because they are quadratic to the real or imaginary voltages. $\square$

If the rank of $W$ matrix is 1, it implies that the ranks of all second-order determinants of $W$ also equal to 1. For example, suppose $W$ is a $4 \times 4$ symmetric matrix,

$$
W = \begin{bmatrix} W_{11} & W_{12} & W_{13} & W_{14} \\ W_{21} & W_{22} & W_{23} & W_{24} \\ W_{31} & W_{32} & W_{33} & W_{34} \\ W_{41} & W_{42} & W_{43} & W_{44} \end{bmatrix}
$$

Since $W$ is a symmetric matrix, there are $W_{12} = W_{21}$, $W_{13} = W_{31}$, $W_{23} = W_{32}$, and etc. If the rank of $W$ is 1, it implies that $W_{11} \times W_{22} = W_{12}^2$, $W_{11} \times W_{33} = W_{13}^2$, and vice versa. With this property, we proposed the penalized SDP relaxation by adding these quadratic constraints into (3-1) to push the rank of $W$ matrix reach to 1.

### 3.1.2 Penalized SDP Relaxation

The penalized SDP relaxation of (3-1) can be written as,

$$
\begin{aligned}
\min \quad & \sum_{k \in \mathcal{G}} f_k(P_{Gk}) + \sum \lambda_i (W_{ii} W_{jj} - W_{ij}^2) \\
s.t. \quad & P_{G_k} - P_{D_k} = tr(Y_k W), Q_{G_k} - Q_{D_k} = tr(\overline{Y}_k W) \\
& P_k^{min} \leq P_{G_k} \leq P_k^{max}, Q_k^{min} \leq Q_{G_k} \leq Q_k^{max}, \\
& (V_k^{min})^2 \leq W_{kk} + W_{(k+n_b)(k+n_b)} \leq (V_k^{max})^2, \\
& W \succeq 0, \ k \in \mathcal{N}.
\end{aligned}
\tag{3-2}
$$

The penalty terms $W_{ii} W_{jj} - W_{ij}^2$ are from the second-order determinants of the $W$ matrix and $\lambda_i$ is adjustable.

In (3-2), the objective function with linear constraints is quadratic and has some negative eigenvalues, leading to a typical quadratic programming problem with linear constraints (LCQP). In the next subsection, we will illustrate the successive linear programming (SLP) and SLPBB algorithms for the LCQP problems.

### 3.1.3   Successive Linear Programming

To begin with, we first show the canonical quadratic format of problem (3-2) to make it easier to express,

$$
\begin{aligned}
\min \quad & f(w) = w^T Q w + q^T w \\
\text{s.t.} \quad & w \in \mathcal{F}, \ \mathcal{F} = \{w \in \mathbb{R}^n : Aw \leq b\}
\end{aligned}
\tag{3-3}
$$

where $w \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{m \times n}$, $q \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. It should be noted that the $w$ in (3-3) is a vector form of $W$ matrix in (3-1) and (3-2) by considering the symmetrical property of the $W$ matrix with the new dimension of $n = n_b(1 + 2n_b)$.

Since $Q$ has some negative eigenvalues, we can rewrite the quadratic term in the objective function via the so-called Difference Convex (D.C.) program [56],

$$
\begin{aligned}
w^T Q w &= w^T Q_+ w - w^T Q_- w \\
&= w^T Q_+ w - \|Cw\|^2,
\end{aligned}
\tag{3-4}
$$

where both $Q_+$ and $Q_-$ are symmetric positive semi-definite matrices derived from the singular value decomposition of $Q$, $C \in \mathfrak{R}^{r \times n}$ and $r$ is the number of negative eigenvalues of $Q$ (the definition of matrix $C$ is given in 3.1.4.1). Then, problem (3-3) can be recast as the following lifted optimization problem,

$$
\min_{(w,t) \in \mathcal{F}_t} w^T Q_+ w + q^T w - \|t\|^2,
\tag{3-5}
$$

$$
\mathcal{F}_t = \{(w,t) | \ Aw \leq b, \ Cw - t = 0, \ t_l \leq t \leq t_u\},
$$

and $t_l, t_u \in \mathfrak{R}^r$ are lower and upper bounds via solving the following linear optimization problems,

$$
t_l^i = \min_{w \in \mathcal{F}} c_i^T w, \quad t_u^i = \max_{w \in \mathcal{F}} c_i^T w, \quad i = 1, ..., r,
\tag{3-6}
$$

where $c_i^T$ is the $i$-th row vector of matrix $C$. Next, we consider a special Lagrangian method for problem (3-5) as follows,

$$
L(w, t, \lambda) = w^T Q_+ w + q^T w - \|t\|^2 + \lambda^T (Cw - t),
\tag{3-7}
$$

where the Lagrangian multipliers are applied to the constraint $Cw - t = 0$. By examining the Karush–Kuhn–Tucker (KKT) conditions of the above Lagrangian function with respect to $t$, we have $\frac{\partial L(w,t,\lambda)}{\partial t} = -2t - \lambda = 0$, which implies

$$\lambda = -2t. \tag{3-8}$$

Recall that the optimal solution of problem (3-3) must be a KKT point of the above Lagrangian function [57], then problem (3-3) can be expressed as the following bi-convex optimization problem,

$$
\begin{aligned}
\min \quad & w^T Q_+ w + q^T w - 2t^T Cw + \|t\|^2 \\
s.t. \quad & w \in \mathcal{F}
\end{aligned}
\tag{3-9}
$$

We should mention that the idea of recasting non-convex QP problems as bi-convex optimization has been widely used in the literature [58, 59, 60]. The novelty of model (3-9) lies in the fact that there are no constraints in $t$ while other existing bi-convex models for non-convex QP problems usually involve certain constraints on $t$. It is easy to see that for any given $t \in [t_l, t_u]$, problem (3-9) is feasible and well-defined. Therefore, we have:

**Proposition 3.1.2.** *If $\bar{w}$ is an optimal solution of problem (3-9) with $t = C\bar{w}$, then $\bar{w}$ is a KKT point of problem.*

Let $S(t)$ denote the optimal solution set of problem (3-9) and $w(t) \in S(t)$. The following proposition summarizes the properties of $Cw(t)$ as a map of $t$.

**Proposition 3.1.3.** *Let $w(t) \in S(t)$. We have*

1. *If $w \in \mathcal{F}$ satisfies the following inequality*

$$\|Cw - t\| \leq \|Cw(t) - t\|,$$

   *then it must hold*

$$f(w) \geq f(w(t))$$

2. *Let $w^*$ be the global optimal solution of problem. Then*

$$\|Cw(t) - t\| \leq \|Cw^* - t\|, \tag{3-10}$$

24

3. *The map $Cw(t)$ is monotone with respect to $t$ in the sense that*

$$(t_1 - t_2)^T [Cw(t_1) - Cw(t_2)] \geq 0 \qquad (3\text{-}11)$$

   *where*

$$t_1 \neq t_2 \in [t_l, t_u], \ w(t_1) \in S(t_1), \ w(t_2) \in S(t_2)$$

*Proof.* Statement 1 follows easily from the fact that $w(t)$ is the optimal solution of problem (3-9) and that $\|Cw - t\| \leq \|Cw(t) - t\|$.

Now, we turn to Statement 2. Since $w^*$ is the global solution of problem and $w(t) \in \mathcal{F}$, we have

$$f(w(t)) \geq f(h^*). \qquad (3\text{-}12)$$

Recall that $w^*$ is a feasible solution of problem (3-9). Since $w(t) \in S(t)$, we have

$$f(w(t)) + \|Cw(t) - t\|^2 \leq f(w^*) + \|Cw^* - t\|^2. \qquad (3\text{-}13)$$

It follows from (3-12) and (3-13) that

$$\|Cw(t) - t\| \leq \|Cw^* - t\|.$$

This proofs Statement 2.

Next, we consider about Statement 3. Let $t_1 \neq t_2 \in [t_l, t_u]$, $w(t_1) \in S(t_1)$, and $w(t_2) \in S(t_2)$. From the optimal conditions of problem (3-9), we obtain

$$f(w(t_1)) + \|Cw(t_1) - t_1\|^2 \leq f(w(t_2)) + \|Cw(t_2) - t_1\|^2,$$
$$f(w(t_2)) + \|Cw(t_2) - t_2\|^2 \leq f(w(t_1)) + \|Cw(t_1) - t_2\|^2.$$

Adding the above two inequalities together, we derive

$$t_2^T Cw(t_2) + t_1^T Cw(t_1) \geq t_2^T Cw(t_1) + t_1^T Cw(t_2)$$
$$\Longleftrightarrow (t_1 - t_2)^T C(w(t_1) - w(t_2)) \geq 0,$$

which yields relation (3-11). This completes the proof. $\qquad \square$

Based on (3-9), the successive linear programming (SLP) method for LCQP problems can be illustrated as follows,

---

**Algorithm 1** Successive Linear Programming $\mathrm{SLP}(w^0, \epsilon)$

**Input:** Initial parameter $w^0$ and stop criteria $\epsilon > 0$;

**Step 0** Set $k = 0$, $t^k = Cw^k$;

**Step 1** Solve (3-9) with $t = t^k$ for optimal solution $w(t^k)$. Set $w^{k+1} = w(t^k)$, $t^{k+1} = Cw^{k+1}$;

**Step 2** If $\|t^{k+1} - t^k > \sqrt{\epsilon}\|$, then set $k = k + 1$ and go back to step 1; Otherwise stop, and output $(w^{k+1}, t^{k+1})$ as the final solution.

---

In the sequel, we need the following assumption:

**Assumption 3.1.1.** *The constrained set $\mathcal{F}$ is bounded.*

**Lemma 3.1.1.** *Let the two sequences $\{w^k\}$ and $\{t^k\}$ be generated by the SLP algorithm and $w^*$ be the global solution of (3-3). We have,*

1. *The sequence $\{t^k\}$ is monotone in the sense that,*

$$(t^k - t^{k-1})^T(t^{k+1} - t^k) \geq 0;$$

2. $f(w^{k-1}) - f(w^k) \geq \|t^{k-1} - t^k\|^2$, *for all $k$;*
3. $f(w^*) \leq f(w^k) \leq f(w^*) + \|t^* - t^{k-1}\|^2$ *for all $k$.*

*Proof.* Statement 1 follows from Statement 3) of Proposition 3.1.3 and the SLP algorithm. Next, we consider about Statement 2.

From Step 1 in the SLP algorithm, since $w^k$ is the optimal solution of (3-9) with $t = t^{k-1}$ and $w^{k-1} \in \mathcal{F}$, we can deduce,

$$f(w^k) + \|Cw^k - t^{k-1}\|^2 \leq f(w^{k-1}) + \|Cw^{k-1} - t^{k-1}\|^2,$$

which, together with the fact that $Cw^{k-1} = t^{k-1}$, yields

$$f(w^{k-1}) - f(w^k) \geq \|t^k - t^{k-1}\|^2, \ \forall k \geq 1.$$

About Statement 3, since $w^*$ is the global solution of problem (3-3) and $w^k \in \mathcal{F}$, we have

$$f(w^k) \geq f(w^*), \ for \ all \ k. \tag{3-14}$$

Also, $w^k$ is the optimal solution of problem (3-9) with $t = t^k$ and $w^* \in \mathcal{F}$, we have

$$f(w^k) + \|Cw^k - t^{k-1}\|^2 \leq f(w^*) + \|Cw^* - t^{k-1}\|^2. \tag{3-15}$$

It then follows from (3-14) and (3-15) that

$$f(w^*) \leq f(w^k) \leq f(w^*) + \|Cw^* - t^{k-1}\|^2.$$

This completes the proof. $\qquad\square$

Next, we characterize the properties of the accumulation point of a sequence generated from the SLP algorithm. For such a purpose, we introduce the so-called $\epsilon$-local minimizer as follows.

**Definition 3.1.1.** *$w^* \in \mathcal{F}$ is called an $\epsilon$-local minimizer of problem (3-3) with respect to a neighborhood,*

$$\mathcal{N}(w^*, \delta) = \{w : \|w - w^*\| \leq \delta\},$$

*if for every $w \in \mathcal{N}(w^*, \delta)$, the following relation*

$$f(w^*) \leq f(w) + \epsilon$$

*holds.*

The above definition can be viewed as an extension of the so-called $\epsilon$-solution introduced in [61]. Then we have,

**Theorem 3.1.1.** *Suppose that $w(t)$ is the optimal solution to (3-9) for some $t \in [t_l, t_u]$. If $t = Cw(t)$, then $w(t)$ is an $\epsilon$-local solution of the function $f(w)$ in a suitable neighborhood.*

*Proof.* Since $w(t)$ is the optimal solution to problem (3-9) and $t = Cw(t)$, we have

$$f(w) + \|Cw(t) - Cw\|^2 \geq f(w(t)), \ \forall w \in \mathcal{F}. \tag{3-16}$$

For a given $\epsilon$, define

$$\delta = \frac{\sqrt{\epsilon}}{\|C\|}.$$

It follows that

$$\|Cw - Cw(t)\|^2 \leq \|C\|^2 \|w - w(t)\|^2 \leq \epsilon,$$

$$\forall \|w - w(t)\| \leq \delta,$$

which, together with (3-16), yields

$$f(w) \geq f(w(t)) - \epsilon.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Based on Theorem 3.1.1 and Lemma 3.1.1, we can obtain the convergence results of the SLP algorithm to an $\epsilon$-local minimizer of problem (3-3) as follows,

**Theorem 3.1.2.** *Let $\{(w^k, t^k)\}$ be a sequence generated from the SLP algorithm with $\epsilon = 0$ and $(\bar{w}, \bar{t})$ be an accumulation point of the sequence. Then $\bar{w}$ is both a KKT point and an $\epsilon$-local minimizer of problem.*

*Proof.* Statement 2 in Lemma 3.1.1 implies that $\{f(w^k)\}$ is a non-increasing sequence. Since $\mathcal{F}$ is compact and $\{w^k\} \subseteq \mathcal{F}$, $\{f(w^k)\}$ is bounded, $\{f(w^k)\}$ converges and consequently $f(w^{k-1}) - f(w^k) \to 0$. It then follows that,

$$\lim_{k \to \infty} \|t^{k-1} - t^k\| = 0. \tag{3-17}$$

Let $(\bar{w}, \bar{t})$ be an accumulation point of $\{(w^k, t^k)\}$. Then, there exists a subsequence $\{(w^{k_j}, t^{k_j})\}$ such that,

$$w^{k_j} \to \bar{w}, \ t^{k_j} \to \bar{t}, \ where \ j \to \infty$$

By (3-17), we can obtain $t^{k_j-1} \to \bar{t}$ as $j \to \infty$. Since $t^k = Cw^k$ for all $k$, we have $\bar{t} = C\bar{w}$. The closeness of $\mathcal{F}$ implies $\bar{h} \in \mathcal{F}$. Since $w^{k_j}$ is an optimal solution of problem (3-9) with $t = t^{k_j-1}$, we have,

$$w^T Q_+ w + q^T w - 2(t^{k_j-1})^T Cw + \|t^{k_j-1}\|^2$$
$$\geq (w^{k_j})^T Q_+ w^{k_j} + q^T w^{k_j} - 2(t^{k_j-1})^T Cw^{k_j} + \|t^{k_j-1}\|^2,$$

$$\forall w \in \mathcal{F}.$$

By taking the limit on both sides of the inequality as $j \to \infty$, we can obtain,

$$w^T Q_+ w + q^T w - 2(\bar{t})^T Cw + \|\bar{t}\|^2$$
$$\geq \bar{w}^T Q_+ \bar{w} + q^T \bar{w} - 2\bar{t}^T C\bar{w} + \|\bar{t}\|^2, \ \forall h \in \mathcal{F}.$$

This implies that $\bar{w}$ is an optimal solution of (3-9) with $t = \bar{t} = C\bar{w}$. Therefore, it illustrates that $\bar{w}$ is both an $\epsilon$-local solution and a KKT point of problem. This completes the proof. $\square$

However, when $r > 1$, SLP Algorithm may fail to find the optimal solutions to problem (3-3) via searching the convex hull constructed from the accumulation points provided by the SLP algorithm. Luckily, the SLP algorithm can locate a very good feasible solution which can be used as an upper bound in the new Branch & Bound (B&B) approach.

### 3.1.4   SLPBB Algorithm

In this section, we develop a new global algorithm that can find a globally optimal solution to problem (2-1) within a pre-specified $\epsilon$-tolerance by combining the SLO algorithm, branch-and-bound (B&B) framework and convex relaxation techniques. The convergence of the algorithm and its complexity are also established.

29

### 3.1.4.1 Convex Relaxation of the Lifted Problem

We start by considering a restricted version of the lifted problem (3-5) where the variable $t$ is in a subbox $[l, u]$ with $l, u \in \Re^r$, $[l, u] \subseteq [t_l, t_u]$,

$$
\begin{aligned}
\min \quad & f(w, t) = w^T Q_+ w + q^T w - \|t\|^2, \\
\text{s.t.} \quad & Cw - t = 0, \\
& w \in \mathcal{F}, \ t \in [l, u].
\end{aligned}
\tag{3-18}
$$

Let $c_i^T$ denote the $i$-th row vector of matrix $C$. Let $\lambda_i, i = 1, ..., r$ be the absolute value of the negative eigenvalues of $Q$, $p_i$ be the corresponding orthogonal unit eigenvector, and $s_i = t_i^2$. By using the singular value decomposition (SVD) of $Q$, $t_i = c_i^T w = \sqrt{\lambda_i} p_i^T w$. Since $\sum_{i=1}^n p_i p_i^T = I$, we have,

$$
\sum_{i=1}^r \frac{s_i}{\lambda_i} = \sum_{i=1}^r w^T (p_i p_i^T) w
$$

$$
\leq w^T (\sum_{i=1}^r p_i p_i^T) w = w^T w \leq \bar{u}^T w, \quad \forall w \in \mathcal{F}.
$$

where $\bar{u} \in \mathbb{R}^n$ is the upper bound of $w$ over $\mathcal{F}$. On the other hand, when $t \in [l, u]$, we have

$$
s_i = t_i^2 \leq (l_i + u_i) t_i - l_i u_i, \ i = 1, ..., r.
$$

Therefore, we can derive the following convex relaxation for problem (3-18):

$$
\begin{aligned}
\min \quad & w^T Q_+ w + q^T w - \sum_{i=1}^r s_i, \\
\text{s.t.} \quad & Cw - t = 0, \\
& t_i^2 \leq s_i, \ s_i \leq (l_i + u_i) t_i - l_i u_i, \ i = 1, ..., r, \\
& \sum_{i=1}^r \frac{s_i}{\lambda_i} \leq \bar{u}^T w, \ w \in \mathcal{F}, \ t \in [l, u].
\end{aligned}
\tag{3-19}
$$

Next, we compare the objective values at the optimal solutions to problem (3-18) and its relaxation (3-19). We have,

**Theorem 3.1.3.** *Let $f_{[l,u]}^*$ and $v_{[l,u]}$ be the optimal values of (3-18) and its relaxation (3-19), respectively. Let $(\bar{w}, \bar{s}, \bar{t})$ be the optimal solution to (3-19). Then we have,*

$$0 \leq f_{[l,u]}^* - v_{[l,u]} \leq f(\bar{w}, \bar{t}) - v_{[l,u]} \leq \frac{\|u - 1\|^2}{4}.$$

*Proof.* Clearly, $(\bar{w}, \bar{t})$ is a feasible solution to eqrefres. Since (3-19) is a convex relaxation of (3-18), it follows from the choice of $f_{[l,u]}^*$ that

$$
\begin{aligned}
0 \leq f_{[l,u]}^* - v_{[l,u]} &\leq f(\bar{w}, \bar{t}) - v_{[l,u]} \\
&= \sum_{i=1}^{r} (\bar{s}_i - \bar{t}_i^2) \\
&\leq \sum_{i=1}^{r} [-\bar{t}_i^2 + (l_i + u_i)\bar{t}_i - l_i u_i] \\
&\leq \frac{1}{4} \sum_{i=1}^{r} (u_i - l_i)^2 = \frac{1}{4}\|u - l\|^2,
\end{aligned}
\tag{3-20}
$$

where the second inequality follows from the constraints on $s_i$ in (3-19). This completes the proof. $\qquad\square$

From the proof of Theorem 3.1.3, we have,

**Proposition 3.1.4.** *Let $(\bar{w}, \bar{s}, \bar{t})$ be the optimal solution to (3-19) and $\epsilon > 0$. If $\sum_{i=1}^{r}[\bar{s}_i - \bar{t}_i^2] \leq \epsilon$, then $(\bar{w}, \bar{t})$ is an $\epsilon$-optimal solution of (3-18).*

*Proof.* Note that $(\bar{w}, \bar{t})$ is a feasible solution to problem (3-18). It follows from (3-20) that,

$$0 \leq f(\bar{w}, \bar{t}) - f_{[l,u]}^* \leq f(\bar{w}, \bar{t}) - v_{l,u} \leq \sum_{i=1}^{r} (\bar{s}_i - \bar{t}_i^2) \leq \epsilon.$$

Then, $f(\bar{w}, \bar{t}) \leq f_{[l,u]}^* + \epsilon$ which implies that $(\bar{w}, \bar{t})$ is an $\epsilon$-optimal solution of (3-18). This completes the proof. $\qquad\square$

Note that, for a given feasible point $w^*$, we can carefully choose a neighborhood $\mathcal{N}(w^*, \delta)$ that meets the following constraint,

$$t_i \in \left[l_i = (Cw^*)_i - \sqrt{\epsilon/r}, \ u_i = (Cw^*)_i + \sqrt{\epsilon/r}\right]. \tag{3-21}$$

From Theorem 3.1.3, we can obtain the following Theorem.

31

**Theorem 3.1.4.** *Given $w^* \in \mathcal{F}$ and a neighborhood $\mathcal{N}(w^*, \delta)$ satisfying all the constraints in (3-21). Let $(\bar{w}, \bar{s}, \bar{t})$ be the optimal solution to the relaxed problem (3-19). Then, $(\bar{h}, \bar{t})$ is an $\epsilon$-local solution to (3-18) with respect to $\mathcal{N}(w^*, \delta)$.*

*Proof.* Let $f^*_{[l,u]}$ and $v_{[l,u]}$ be the optimal values of (3-18) and its relaxation (3-19), respectively. By Theorem 3.1.3, we have,

$$f(\bar{w}, \bar{t}) - f^*_{[l,u]} \leq f(\bar{w}, \bar{t}) - v_{l,u} \leq \|u - l\|^2 / 4.$$

Together with (3-21), we can obtain,

$$f(\bar{w}, \bar{t}) \leq f^*_{[l,u]} + |u - l\|^2 / 4 \leq \min_{w \in \mathcal{N}(w^*, \delta), t \in [l,u]} f(w, t) + \epsilon.$$

Therefore, $(\bar{w}, \bar{t})$ is an $\epsilon$-local solution to (3-18) with respect to $\mathcal{N}(w^*, \delta)$. This completes the proof. $\qquad \square$

Based on Theorem 3.1.4, we can divide each interval $[t^i_l, t^i_u]$ into $\left\lceil \frac{\sqrt{r}(t^i_u - t^i_l)}{2\sqrt{\epsilon}} \right\rceil$ sub-intervals $[l_i, u_i]$, where $t^i_l, t^i_u$ are defined by . Correspondingly, the feasible region $\mathcal{F}$ can be divided into many sub-regions where each sub-region is bounded with box constraints like $t_i \in [l_i, u_i]$, $u_i - l_i \leq 2\sqrt{\epsilon/r}$. Then, (3-19) can be solved in all the sub-regions to obtain an $\epsilon$-local approximation in each sub-region. Among these $\epsilon$-local solutions, we choose the one with the minimal objective function value as the final solution, which is clearly a global $\epsilon$-approximate solution to the original problem (3-3). Therefore, such partitioning procedure can provide an $\epsilon$-approximation to (3-3). For convenience, we call this procedure the brutal force algorithm. The following Theorem implies the above discussion.

**Theorem 3.1.5.** *The brutal force algorithm can find a global $\epsilon$-approximate solution to (3-3) in $\mathcal{O}\left( \prod_{i=1}^{r} \left\lceil \frac{\sqrt{r}(t^i_u - t^i_l)}{2\sqrt{\epsilon}} \right\rceil N \right)$ time, where $N$ is the complexity to solve (3-19).*

### 3.1.4.2 Proposed SLPBB Algorithm

Based on (3-19), the proposed SLPBB algorithm for LCQP problems is described as follows,

---

**Algorithm 2** The SLPBB Algorithm

**Input:** $Q, q, A, b$, and stop criteria $\epsilon > 0$.

**Step 0 (Initialization)**

(i) Set $k := 0$, $\Delta_k := [l^k, u^k]$ with $l^0 = t_l$, $u^0 = t_u$, where $t_l$ and $t_u$ are computed by (3-6), $\underline{v}^k := -\infty$, $\Omega := \{(\Delta^k, \underline{v}^k)\}$.

(ii) For $j \in 1, ..., 2^r$, solve problem $\min_{(w,t) \in \mathcal{F}_t} u_j^T t$ with $u_j \in (-1, 1)^r$ to get the optimal solution $(w_j, t_j)$, respectively.

**Step 1 (SLP)** Find $\epsilon$-local solutions $w_j^k$ of (3-3) by running SLP$(t^0, \epsilon)$ with $t^0 = t_j$. Set $w^* := argmin\{f(w_j^k)\}$ and $v^* = f(w^*)$.

**Step 2 (Termination)** If $\Omega \neq \emptyset$, go to Step 3; Else, $w^*$ is an approximate optimal solution to (3-3), stop.

**Step 3 (Node Selection)** Choose the node $(\Delta^k, \underline{v}^k)$ from $\Omega$ with the smallest $\underline{v}^k$ and remove it from $\Omega$. Solve (3-19) over $\Delta^k$. If it is infeasible, go to Step 2; else, denote $(w^k, s^k, t^k)$ and $v^k$ as its optimal solution and optimal value, respectively.

**Step 4** If $v^k \geq v^* - \epsilon$, go to Step 2; else if $\sum_{i=1}^r [s_i^k - (t_i^k)^2] \leq \epsilon$, then $w^k$ is an $\epsilon$-solution to (3-18) over $\Delta^k$, update $v^*$ and delete all the nodes $(\Delta^j, \underline{v}^j)$ in $\Omega$ with $\underline{v}^j \geq v^* - \epsilon$, go to Step 2; else, go to Step 5.

**Step 5 (Restart SLP)** If $f(w^k) < v^*$, then find an $\epsilon$-local solution $\overline{w}^k$ of (3-3) by running SLP$(t^0, \epsilon)$ with $t^0 = Cw^k$, update $v^*$ and delete all the nodes $(\Delta^j, \underline{v}^j)$ in $\Omega$ with $\underline{v}^j \geq v^* - \epsilon$.

**Step 6 (Partition)** Choose $i^*$ maximizing $s_i^k - (t_i^k)^2$. Set $\omega_{i^*} = \frac{l_{i^*}^k + l_{u^*}^k}{2}$ and partition $\Delta^k$ into two sub-rectangles $\Delta_1^k$ and $\Delta_2^k$ along the edge $[l_{i^*}^k, u_{i^*}^k]$. Set $\underline{v}_1^k := \underline{v}_2^k := v^k$, $\Omega := \Omega \cup \{(\Delta_1^k, \underline{v}_1^k), (\Delta_2^k, \underline{v}_2^k)\}$.

**Step 7** Set $k = k + 1$ and go to Step 3.

---

We have the following remarks regarding the steps of the SLPBB algorithm.

1. In Step 1, the SLP algorithm is applied with different feasible points of the lifted problem as starting point to find a good $\epsilon$-local solution of (3-3).

2. In Step 5, the SLP algorithm is restarted to find a better $\epsilon$-local solution if the objective value at the feasible point derived from the solution of the relaxation problem is less than the current upper bound.

3. In Step 6, the optimal solution of the relaxation problem is cut off after each iteration so that the lower bound will be improved.

We next establish the convergence of the SLPBB Algorithm. Let $\bar{\Delta} = [\bar{l}, \bar{u}]$ be a sub-rectangular satisfying,

$$\bar{u}_i - \bar{l}_i \leq 2\sqrt{\epsilon/r}, \quad i = 1, ..., r. \tag{3-22}$$

Let $(\bar{w}, \bar{s}, \bar{t})$ be an optimal solution of (3-19) over $\bar{\Delta}$. From (3-20), there is,

$$\bar{s}_i - \bar{t}_i^2 \leq \frac{1}{4}(\bar{u}_i - \bar{l}_i)^2, i = 1, ..., r, \tag{3-23}$$

which implies that,

$$\sum_{i=1}^{r}[\bar{s}_i - \bar{t}_i)^2] \leq \epsilon. \tag{3-24}$$

By Proposition 3.1.4, $\bar{w}$ is an $\epsilon$-optimal solution to (3-18) over $\bar{\Delta}$. According to Step 3 and Step 4 of the SLPBB algorithm, this sub-rectangle $\bar{\Delta}$ will be permanently deleted from $\Omega$. In other words, a sub-rectangle $\Delta$ satisfying (3-22) will never be divided further in the algorithm. Then for a current sub-problem to be considered further, if its optimal solution $(\bar{w}, \bar{s}, \bar{t})$ does not satisfy (3-24), then there exists $i^* \in 1, ..., r$ such that $\bar{s}_{i^*} - \bar{t}_{i^*}^2 > \epsilon/r$, which in turn by (3-23) implies that $\bar{u}_{i^*} - \bar{l}_{i^*} > 2\sqrt{\epsilon/r}$. That is, the $i^*$-th edge of the corresponding sub-rectangular must be longer than $2\sqrt{\epsilon/r}$. Based on Step 6, it will be divided at the point $\bar{t}_{i^*}$ or the midpoint of this edge. Furthermore, all the edges of sub-rectangle corresponding to a sub-problem required to be solved will never be shorter than $2\sqrt{\epsilon/r}$. Therefore, every edge of the initial rectangle will be divided into at most $\lceil \frac{\sqrt{r}(t_u^i - t_l^i)}{2\sqrt{\epsilon}} \rceil$ sub-intervals. That is, to get an $\epsilon$-optimal solution to (3-3), the number of the relaxed sub-problems required to be solved with the SLPBB algorithm is bounded by,

$$\prod_{i=1}^{r} \lceil \frac{\sqrt{r}(t_u^i - t_l^i)}{2\sqrt{\epsilon}} \rceil.$$

From the discussion above, we have,

**Theorem 3.1.6.** *The SLPBB Algorithm can find a global $\epsilon$-approximate solution to problem in at most $O(\prod_{i=1}^{r}\lceil \frac{\sqrt{r}(t_u^i - t_l^i)}{2\sqrt{\epsilon}} \rceil N)$ time, where $N$ is the complexity to solve problem (3-19).*

## 3.2 Fault Recognition Algorithm

As a category of deep neural networks, convolutional neural network (CNN, or ConvNet) has been applied in multiple fields: image and video recognition, recommender systems, image classification, natural language processing, et al [62, 63]. Based on the shared-weights architecture and translation invariance characteristics [64], ConvNet utilizes the hierarchical pattern in data and assembles complex patterns with smaller and simpler patterns.

To classify and localize the four faults occurred in distribution systems, we develop Model-C (fault classification) and Model-L (fault localization) based on ConvNet with the input feature vectors extracted from Chapter 2.2.

### 3.2.1 ConvNet Algorithm

Generally, there are four primary operations in ConvNet: convolution, non linearity, pooling and classification. After importing the feature vector $\varphi_l$ or $\varphi_t$ extracted in Section II as the input data and label vector $\boldsymbol{y}$ which represents the fault types or fault localization, the networks will learn the input data by forward propagation and the parameters of the networks will be optimized by backpropogation with the errors between the output and labels [65].

In convolution, the main propose of it is to extract features from the input data and preserve the spatial relationship between them. Suppose the input of the $l-th$ convolutional layer is $\boldsymbol{X}_l$, and the kernels of this layer are $\boldsymbol{W}_l$. Then The output of the convolution computation between $\boldsymbol{X}_l$ and $\boldsymbol{W}_l$ with defined stride size and padding size is,

$$\boldsymbol{C}_l = \boldsymbol{X}_l \otimes \boldsymbol{W}_l. \tag{3-25}$$

The second layer after convolution layer is usually the nonlinear rectified linear unit (ReLU) layer which keeps the items to be non-negative by changing the negative items to zeros. The output of this layer can be expressed by,

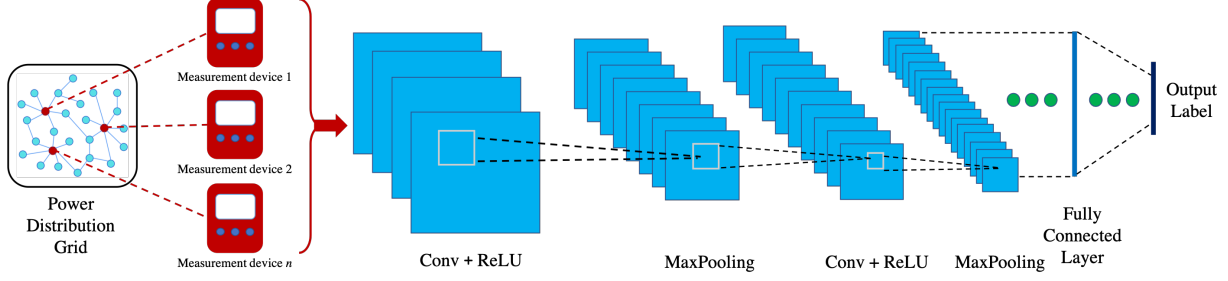$$\boldsymbol{R}_l = \mathbf{max}(\boldsymbol{C}_l, 0). \tag{3-26}$$

Figure 5: The structure of ConvNet algorithm.

The Pooling layer is linked after the ReLU layer which picks the maximum or average value within a defined size of neighborhood. Also, users can define the stride and padding size of this layer which are similar to the convolution layer. In this paper, we use **MaxPooling** that can be shown as,

$$\boldsymbol{P}_l = \textbf{MaxPooling}(\boldsymbol{R}_l). \tag{3-27}$$

In practice, these layers can be combined with each other multiple times and a fully connected layer is followed after the last layer of them to compute the probabilities for each label by using the softmax function,

$$p_{j[n \times K]} = \frac{[e^{f_j}]_{[n \times K]}}{[\sum_{k=1}^{K} e^{f_k}]_{[n \times 1]}}, j = 1, ..., K \tag{3-28}$$

where $f_k$ is the $k - th$ item of the last fully connected layer and the total summation of $p_j$ is 1. The position of the largest $p_j$ will be the output label which is compared with the input label. Figure 5 shows the structure of the ConvNet algorithm with all used operators.

### 3.2.2 Fault Recognition (Fault-RE) Algorithm

In order to quickly and accurately determine the fault type and location occurred in the system, an algorithm named Fault-RE is proposed and is shown below:

---
**Algorithm 3** The Fault-RE Algorithm
---
**Input:** Voltages $V_{abc}^{pref}$, $V_{abc}^{f}$, currents $I_{abc}^{pref}$, $I_{abc}^{f}$ and corresponding phase angles $\theta_{(V/I)abc}^{pref}$, $\theta_{(V/I)abc}^{f}$ before and during the fault occurred in the power system

36

**Step 1 (Feature vector $\varphi_t$)** Calculate the voltage differences $\Delta V_{abc} = V_{abc}^f - V_{abc}^{pref}$, current differences $\Delta I_{abc} = I_{abc}^f - I_{abc}^{pref}$, phase angle differences $\Delta \theta_{(V/I)abc} = \theta_{(V/I)abc}^f - \theta_{(V/I)abc}^{pref}$. Create the feature vector $\varphi_t$ defined in (2-8).

**Step 2 (Fault classification)** Insert the feature vector $\varphi_t$ and fault type labels as the inputs of the ConveNet fault type model (Model-C), obtain the information of the fault types.

**Step 3 (Feature vector $\varphi_l$)** Calculate the zero, negative and positive $(Z, N, P)$ sequences of voltages and currents for each phase as defined in (2-9), then augment them into $\varphi_t$ as the feature vector $\varphi_l$ (defined in (2-10)).

**Step 4 (Fault localization)** Insert the feature vector $\varphi_l$ and fault location labels as the inputs of the ConveNet fault localization model (Model-L), obtain the information of the fault locations.

By using this algorithm, we can get the information of the fault types and locations.

### 3.3    Sub-Synchronous Oscillation Detection Model

To get a better and faster feature-extraction performance, we propose the Psudo-Continuous Quadrature WT (PCQWT) which can keep both of the fundamental and SSO features. Inspired by event detection & prediction in video processing, an SSO detection model is developed based on CNN and LSTM techniques which aim to detect and predict the SSO events.

### 3.3.1    Feature Extraction through Pseudo-Continuous Quadrature WT

As the frequency range of SSO is known, choosing a set of the scaling factors $a_k$ in (2-13) and ensuring a better coverage of the SSO frequency range can provide sufficient information about the existence of the SSO and its development trend. Also, using less number of scaling factors $a_k$ while maintaining the similar feature extraction performance is desired.

Fig. 6 illustrates a performance comparison of three different wavelet transforms on the current waveform obtained from three different SSO events. In an SSO escalating event (Fig.

6(a)), the oscillation of active power starts to intensify. The extracted features from single phase current through different wavelet transforms are shown in Fig. 6(b)(c)(d). One can see that the SSO current features become more obvious and this phenomenon match the active power oscillation. When in a constant SSO event (Fig. 6(e)), the SSO features in Fig. 6(f)(g)(h) are different with each other, the intensities still remain constant. In an SSO damping event (Fig. 6(i)), the intensities of features damp in Fig. 6(j)(k)(l) as the active power oscillation damps too.

It is very obvious that, the SSO features obtained from Morlet are the finest, and therefore, Morlet can be the candidate for SSO detection. However, the features appear periodically in twice of the fundamental frequency, which matches the waveform's peak and valley. To remove the intermittent pattern and ease the event detection in the next stage, a complex Morlet is the most desirable wavelet, it is written as:

$$\Psi_{\text{CMor}}(t) = \frac{1}{\sqrt{\pi\beta}} e^{-\frac{t^2}{2\beta}} e^{j\omega_c t} \tag{3-29}$$

where, $\beta$ controls the shape of Gaussian window and $\omega_c$ is the center frequency. We can see that, WT using complex Morlet is Gaussian-windowed sinusoids for time-frequency decomposition in *quadrature* plane, because the other complex pair is neglected. Consequently, the feature related to the rotating phasor is extracted instead of the waveform amplitude. And the proof is as follows: quadrature WT of a complex Morlet on a sinusoidal waveform with magnitude $A$, frequency $\omega$ and phase $\phi$ with considering scaling factor $a$ and time shift $b$ can be written as:

$$
\begin{aligned}
X(\omega|a, b) &= \int_{-\infty}^{\infty} A\cos(\omega t + \phi)\Psi_{\text{CMor}}^*(\frac{t-b}{a})dt \\
&= \int_{-\infty}^{\infty} A\cos(\omega t + \phi)\frac{1}{\sqrt{\pi\beta}} e^{\left(-\frac{(t-b)^2}{2\beta a^2} - j\omega_c(\frac{t-b}{a})\right)} dt
\end{aligned}
\tag{3-30}
$$

By applying the Hubbard-Stratonovich transformation [66]:

$$\sqrt{2\pi\alpha} \cdot e^{-\frac{\alpha}{2}x^2} = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\alpha^2} - jxy}dy, \tag{3-31}$$

(3-30) can be simplified as:

$$X(\omega|a, b) = \frac{aA}{\sqrt{2}}[e^{j\omega(\phi+b)} e^{\frac{-\beta(\omega_c - a\omega)^2}{2}} + \underbrace{e^{-j\omega(\phi+b)} e^{\frac{-\beta(\omega_c + a\omega)^2}{2}}}_{\approx 0}] \tag{3-32}$$

(a) Case: SSO Escalating

(b) Haar

(c) Daubechies 4

(d) Morlet

(e) Case: Constant SSO

(f) Haar

(g) Daubechies 4

(h) Morlet

(i) Case: SSO Damping

(j) Haar

(k) Daubechies 4

(l) Morlet
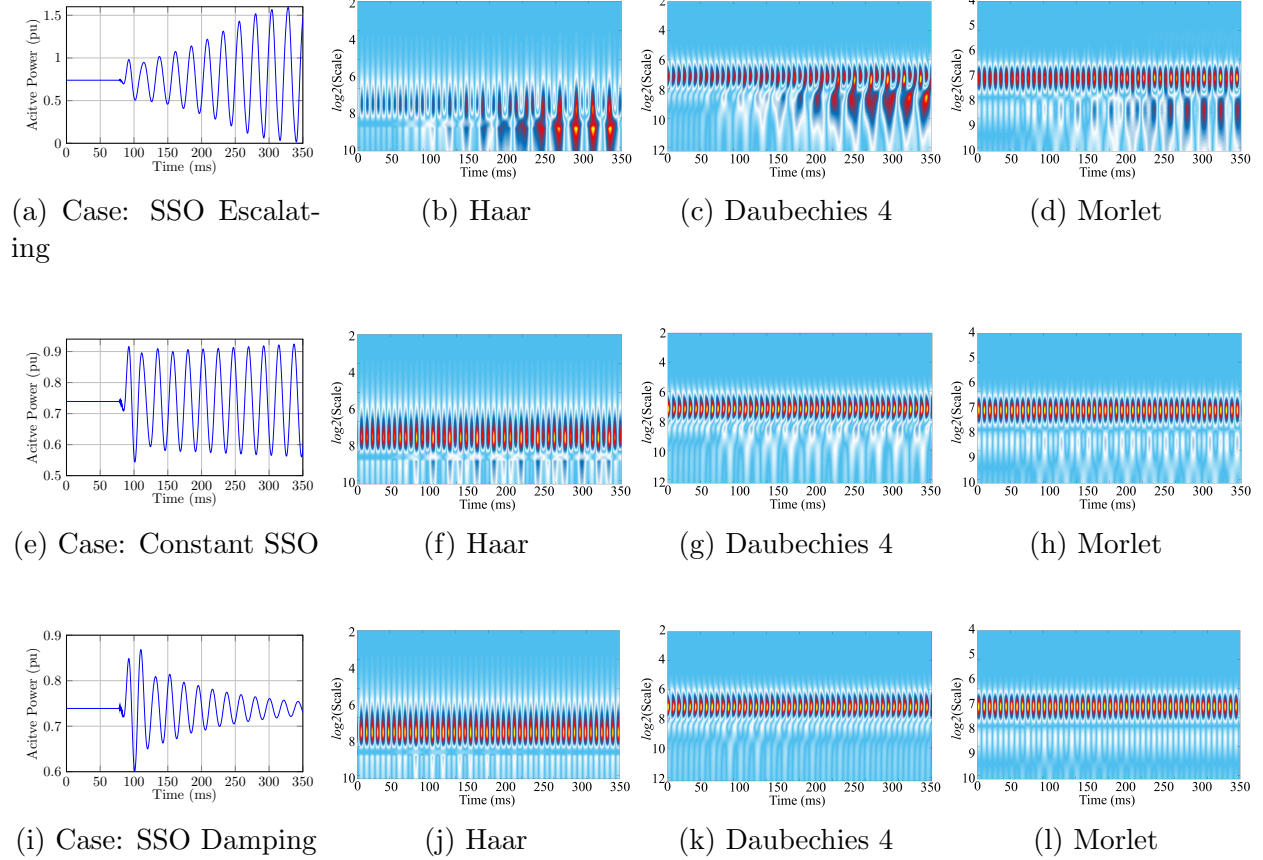
Figure 6: Demonstration of different wavelet transforms on current signals from simulated SSO events (start at t=75ms). (a)(e)(i): Active powers from three different SSO events. (b)-(d): extracted current waveform features for (a). (f)-(h): extracted current waveform features for (e). (j)-(l): extracted current waveform features for (i).

39

For simplicity, set $b$ as a constant value, and select different values of $a$ with a length of $K$. Then the implemented wavelet bank can be written as

$$\mathbf{\Psi}^{K \times W} = \left[ \Psi\left[\frac{nT_s - b_1}{a_1}\right], \cdots, \Psi\left[\frac{nT_s - b_k}{a_k}\right], \cdots, \Psi\left[\frac{nT_s - b_K}{a_K}\right] \right]^T. \tag{3-33}$$

And the extracted features using complex Morlet WT at time instant $n$ are expressed as:

$$|X(\omega|a_k, b)| \approx \frac{a_k A}{\sqrt{2}} e^{\frac{-\beta(\omega_c - a_k\omega)^2}{2}} \tag{3-34}$$

Now, the SSO features can be extracted by selecting the set of $a$ since the extracted features are determined by the value of $a_k$ in (3-34) and the SSO appears at the certain frequency and the feature reaches peak when $\omega_c / a_k$. Also, one should note that low frequency component is amplified by the scaling factor, which is also desired. Typically, a set of discrete scaling factors (in forms of $a^i$), wherein, $i$ is an integer, which is used in discrete WT. In order to achieve high-wealth waveform features and have redundant pseudo frequency coverage in the scalogram, a set of linearly-increasing real numbers are assigned to $i$ in the proposed *Pseudo-Continuous-Quadrature*-WT. Similar to the discrete WT, the proposed PCQ-WT can be directly implemented by using (2-13).

### 3.3.2 Long Short-term Memory (LSTM) Technique

As an improvement of recurrent neural networks (RNN), long short-term memory (LSTM) networks are capable of learning long-term dependencies and avoid the vanishing gradient problems existed in the traditional RNN by incorporating gating functions into their state dynamics [67]. An LSTM network consists of the input layer, hidden layer and output layer. The main characteristics of LSTM are the memory cells which include memory blocks in the hidden layers [68, 2]. Each block contains one or more self-connected memory cells and three multiplicative units—the input, output and forget gates—that provide continuous analogues of write, read and reset operations for the cells [69]. Fig. 7 shows the diagram of the LSTM network at time $t$.
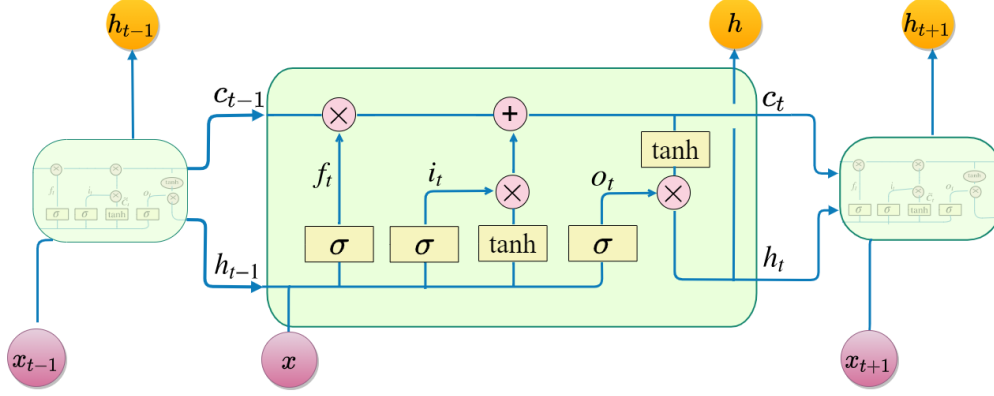
Figure 7: LSTM network at time $t$ [2].

Given an input sequence vector $\boldsymbol{x} = (x_1, ..., x_T)$, the LSTM network calculates the hidden vector and output vector by using the following unit activation equations [68, 69],

$$
\begin{aligned}
i_t &= \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i), \\
f_t &= \sigma(W_{fx}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \\
o_t &= \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_t + b_o), \\
c_t &= f_t \odot c_{t-1} + i_t \odot tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c), \\
h_t &= o_t \odot tanh(c_t), \\
y_t &= W_{yh}h_{t-1} + b_y, \\
t &= 1, ..., T.
\end{aligned}
\tag{3-35}
$$

where $W$ are the weight matrices, i.e., $W_{ix}$ represents the weight matrix from the input to the input gate, $\sigma$ is the logistic sigmoid function. $i$, $f$, $o$ and $c$ denote the activation vectors of input gate, forget gate, output gate and cell, respectively. Since the values of all the gating vectors vary for each time step, the model can learn how to represent information over multiple time steps.

41

### 3.3.3 Event Detection and Prediction Through CNN-LSTM

The SSO event detection and amplitude trend evaluation is inspired by the concept of event detection and prediction in the video processing, as show in Fig. 8,

- Before the player hit the ball, we can use the tennis ball's and the racket's relative positions from previous observations (Frame B1–B4) to decide the likelihood of a successful hit. And that the ball being hit can be determined immediately if the observation in B5 is made.

- After the ball was hit by the player, the trajectory and possible location of the tennis ball in H6 can be predicted using the previous observations in Frame H1–H5. The tennis ball's previous true locations are marked by red dots in Frame H6.

The same concept can be applied in the SSO detection and prediction by using a series of extracted features as the input. However, in real-time application, a fast detection and prediction are required while maintaining accuracy. The scalograms in Fig. 6 cannot be used as each of them requires to wait at least 350ms to generate. Therefore, a sequence of short scalogram segments is desired, because (i) the newest scalogram segment can be processed with minimum delay, if SSO is observed, the positive detection is reported, (ii) a sequence of small segments can be stored in memory to predict the trend of SSO amplitude.

B1       B2       B3       B4       B5

(a) Before Hitting Ball

H1     H2     H3     H4     H5     H6

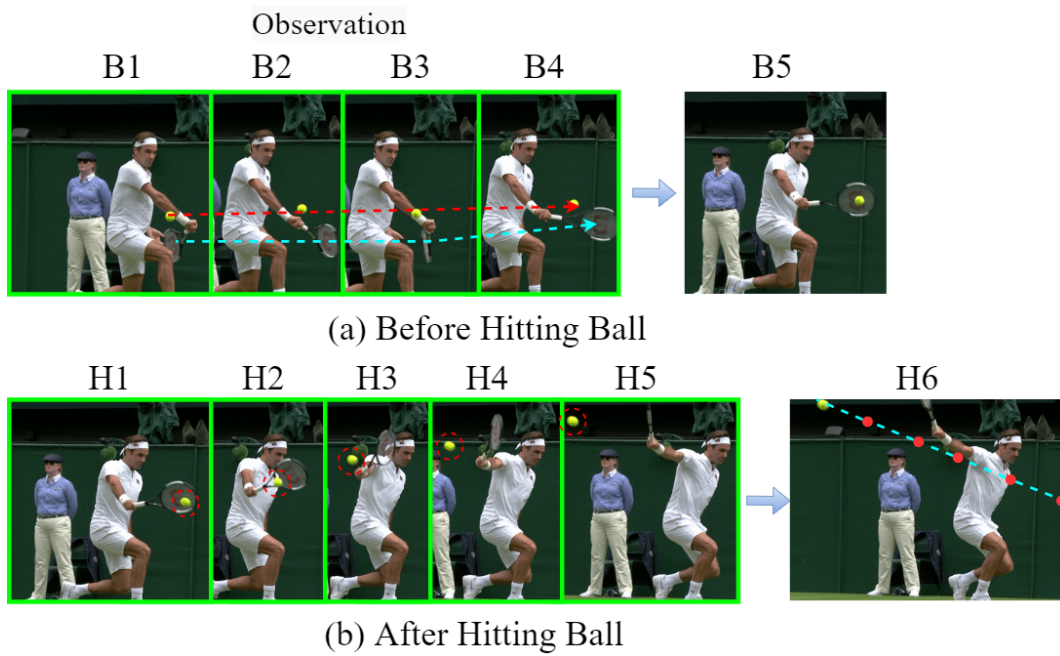(b) After Hitting Ball

Figure 8: An example of event detection and prediction using video frames.

<h1 style="text-align:center">4.0   Numerical Simulations</h1>

To validate the performances of the proposed optimization algorithms for ACOPF problem and the intelligent models for fault recognition as well as SSO detection, this chapter will discuss about the simulation results of these three topics.

<h2 style="text-align:center">4.1   Simulation Results of ACOPF Problems</h2>

In this section, we tested two small and one large scale power systems to examine the efficiency of our proposed algorithm for ACOPF problems. We also compared the results of the performance of SLP and SLPBB algorithms for small systems with the associated local results of in [70] and large system with [71].

### 4.1.1   Case 1: Two-bus Test System

Figure 9 shows the two-bus system, where $bus-1$ is the power generator as a slack bus, and $bus-2$ is the load bus with (352.5 MW, 358 MVars) active and reactive power demand.
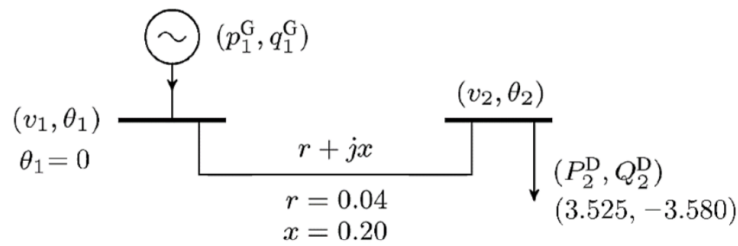


Figure 9: Two-bus power system.

Based on Figure. 9, the admittance matrix of this system is,

$$Y_{2-bus} = \begin{bmatrix} 0.961 - j4.807 & -0.961 + j4.807 \\ -0.961 + j4.807 & 0.961 - j4.807 \end{bmatrix}$$

The penalized SDP relaxation formulation of this two-bus power system can be written as follows,

$$\min_{W \in \Re^{4 \times 4}} 2P_{G1} + \lambda_1(W_{11}W_{22} - W_{12}^2) + \lambda_2(W_{33}W_{44} - W_{34}^2)$$

$$s.t. \quad 0 \leq P_{G_1} = 0.961W_{11} + 0.961W_{33} - 0.961W_{12}$$

$$- 0.961W_{34} - 4.807W_{14} + 4.807W_{23} \leq 5,$$

$$- 4 \leq Q_{G_1} = 4.807W_{11} + 4.807W_{33} - 4.807W_{12}$$

$$- 4.807W_{34} + 0.961W_{14} - 0.961W_{23} \leq 4,$$

$$- P_{D_2} = 0.961W_{22} + 0.961W_{44} - 0.961W_{12} \tag{4-1}$$

$$+ 4.807W_{14} - 4.807W_{23} - 0.961W_{34},$$

$$- Q_{D_2} = 4.807W_{22} + 4.807W_{44} - 4.807W_{12}$$

$$- 4.807W_{34} - 0.961W_{14} + 0.961W_{23},$$

$$(0.95)^2 \leq W_{11} + W_{33} \leq (1.05)^2,$$

$$(0.95)^2 \leq W_{22} + W_{44} \leq (1.05)^2.$$

Utilizing the presented SLP and SLPBB algorithms to solve problem (4-1), the results are shown in Table 1 when ($\lambda_1 = \lambda_2 = 1$),

Table 1: Simulation Results of Case 1.

| Solution Type | $P_{G1}(MW)$ | $Q_{G1}(MVar)$ | $V_1 \angle \theta_1$ | $V_2 \angle \theta_2$ |
|---|---|---|---|---|
| SLP(BB) Results | 438.44 | 92.39 | $0.95\angle 0°$ | $1.05\angle -45°$ |
| Results [70] | 452.86 | 164.32 | $0.95\angle 0°$ | $0.98\angle -64.94°$ |

From Table 1, we can see that the power generation of the SLP algorithm is less than the local result which means that there is less loss in the system and the results are closer to the optimal ones.

45

### 4.1.2 Case 2: Three-bus Test System

In this case, we did simulations for a three-bus power system, which is shown in Figure 10. In this system, $bus-1$ is the power generator which was identified as the slack bus, $bus-2$ and $bus-3$ are the demand buses in which the active power demands are 120 MW and 68 MW, the reactive power demands are 86 MVAr and 50 MVAr, respectively.
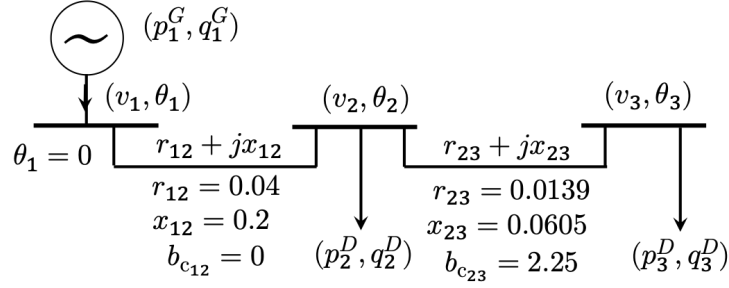


Figure 10: Three-bus power system.

Based on Figure. 10, the admittance matrix $Y_{3-bus}$ of this system is,

$$
\begin{bmatrix}
0.961 - j4.807 & -0.961 + j4.807 & 0 \\
-0.961 + j4.807 & 4.571 - j19.396 & -3.61 + j15.714 \\
0 & -3.61 + j15.714 & 3.61 - j14.589
\end{bmatrix}
$$

In order to run the simulation with limited number of penalized terms, we tried different combinations of them and found that the variables which also existed in the constraints had better effect since they were used in two ways: satisfy the constraints as well as get the $rank-1$ solutions. Hence, the objective function of this three-bus system is as follows $(\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_8 = 1)$,

$$
\begin{aligned}
\min_{W \in \Re^{6\times 6}} \ & 2P_{G_1} + \lambda_1(W_{12}^2 - W_{11}W_{22}) + \lambda_2(W_{15}^2 - W_{11}W_{55}) \\
& + \lambda_3(W_{24}^2 - W_{22}W_{44}) + \lambda_4(W_{45}^2 - W_{44}W_{55}) \\
& + \lambda_5(W_{23}^2 - W_{22}W_{33}) + \lambda_6(W_{26}^2 - W_{22}W_{66}) \\
& + \lambda_7(W_{35}^2 - W_{33}W_{55}) + \lambda_8(W_{56}^2 - W_{55}W_{66}).
\end{aligned}
\tag{4-2}
$$

The simulation results of Case 2 are shown in Table 2. It's noticeable that the power generation of the SLP or SLPBB algorithm is also less than the local result and the results are more optimal and much closer to the optimal ones.

Table 2: Simulation Results of Case 2.

| Solution Type | $P_{G1}(MW)$ | $Q_{G1}(MVar)$ | $V_1 \angle \theta_1$ | $V_2 \angle \theta_2$ | $V_3 \angle \theta_3$ |
|---|---|---|---|---|---|
| SLP(BB) Results | 196.97 | 63.75 | $0.95 \angle 0°$ | $1.022 \angle -12.04°$ | $1.05 \angle -71.81°$ |
| Results [70] | 208.62 | 8.93 | $0.95 \angle 0°$ | $0.95 \angle -27.25°$ | $0.98 \angle -30.36°$ |

### 4.1.3 Case 3: 30-bus Test System

Consider the IEEE 30-bus system with the linear cost function $\sum_{k \in \mathcal{G}} c_k P_{Gk}$, where the cost function coeffcinet of six generation units are $c_1 = 1$, $c_2 = 10$, $c_{13} = 1$, $c_{33} = 10$, $c_{23} = 100$, $c_{27} = 1$ [71] with $\lambda_1 = ... = \lambda_{1830} = 1$. The results of our proposed algorithms compared with the penalized factor $\epsilon = 0$ which is shown the better optimal results when it satisfied KKT conditions in 98.5% with comparing to the 96.6% in [71].
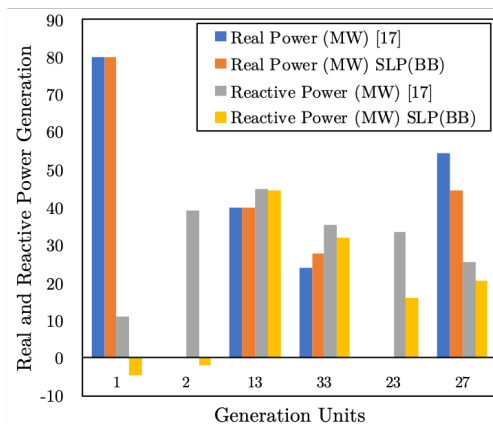


Figure 11: 30-bus test system results.

## 4.2    Simulation Results of Fault Recognition Problems

The studied network is a real power distribution feeder of a Rural Electricity Distribution Company with a length of 17 km main feeder, which includes a main route and some substations. This feeder is powered by a distribution substation that has two 30 MVA transformers with an open tie-switch under normal conditions. The impedance of the above distribution transformer is equal to 12.5% and the short circuit capacity of 63 kV upstream network is 3,450 MVA. The substation is equipped with grounding transformer (with solid ground). It has a zero impedance of 40 *ohms* for each phase. Figure 12 shows a single-line diagram of the studied feeder. The main feeder contains 51 pole mounted distribution transformers, and 476 poles, with 6.43 MW and 3.7 MVar total real and reactive power demand respectively.
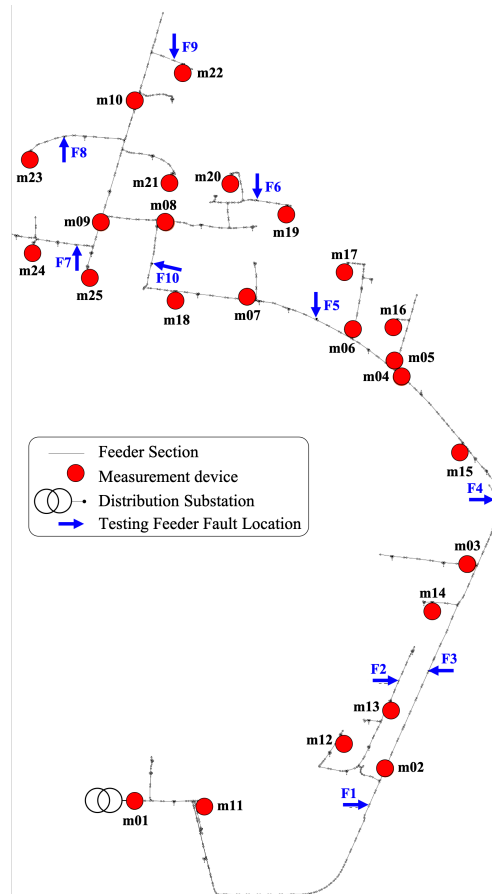


Figure 12: The real rural power distribution system.

In the studied network, 25 micro-PMUs as the meter (with phasor measurement capability) are installed in the indicated locations. Meters $m1$ to $m10$ are located in the main feeder path or feeder T-offs that measure the current flow of a branch. Meters $m11$ to $m25$ measure the operating parameters of a distribution transformers at the secondary side. In order to create the training and testing data sets, 10 fault zones are considered with corresponding letters $F1$ to $F10$ which are labeled in Fig. 12. The impact of the fault distance $d = 0.1$ (km) is also incorporated to support the accuracy of the fault localization.

### 4.2.1   Fault Type Classification Model-C

We aim to classify the fault type by using a ConvNet model called "Model-C". There are 5000 dataset for the fault type classification with 80% and 20% budgets for the training and validation processes, respectively. The fault type classification is evaluated by the fault type classification accuracy (FTCA) $\eta_t$,

$$\eta_t = \frac{the\ number\ of\ faults\ correctly\ categorised}{total\ number\ of\ faults}. \tag{4-3}$$

The proposed ConvNet structure of Model-C is shown in Table 3. There are totally four layers in this model: The input size is $1 \times 12$ and we use two Maxpooling operators, one fully connected layer and one output layer. In each convolution operator, the kernel size is 1 by 3 which was set based on the input data size with the ones padding size and 4 kernels are used which are denoted as $[1\ 3] \otimes 4$. In Maxpooling layer, the kernel size is 1 by 2 and the stride size is 2 which is denoted as $[1\ 2]@2$. The output of this model is 4 by 1 that represents the four fault types.

Table 3: The ConvNet structure of Model-C

| Layer NO. | Layer Type | Kernel Size | Output |
|---|---|---|---|
| the 1st layer | Convolution | $[1\ 3] \otimes 4$ | $[1\ 12] \otimes 4$ |
| | Maxpooling | $[1\ 2]@2$ | $[1\ 6] \otimes 4$ |
| the 2nd layer | Convolution | $[1\ 3] \otimes 4$ | $[1\ 6] \otimes 4$ |
| the 3rd layer | Convolution | $[1\ 3] \otimes 4$ | $[1\ 6] \otimes 4$ |
| the 4th layer | Convolution | $[1\ 3] \otimes 4$ | $[1\ 6] \otimes 4$ |
| | Maxpooling | $[1\ 2]@2$ | $[1\ 3] \otimes 4$ |
| fully connected layer | vectorize | ... | 10 |
| Output | Softmax | ... | 4 |

During the training process of this model, "Adam" optimizer with the decay parameter $\alpha = 0.9$ was utilized after comparing it with Rmsprop and stochastic gradient descent (SGD) optimizers. The initial learning rate was 0.001 which is the default value for "Adam" solver and it will drop every 20 period. $\lambda = 0.01$ which is the $L2$ regularization. The simulation results of Model-C are shown in Figure 13 and Table 4 illustrates the simulation summary of this model. Figure 13 shows that the Model-C with 3 layers reaches to the accuracy rate 100% quickly and smoothly with the best performance among the four models.

Table 4: The simulation summary of Model-C

| CNN layer number | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Elapsed time(s) | 12 | 15 | 19 | 20 |
| Epoch | 10 | 10 | 10 | 10 |
| Learning Rate | 0.001 | 0.001 | 0.001 | 0.001 |
| Validation accuracy | 99.8% | 99.5% | 100% | 100% |

(a) One layer        (b) Two layers
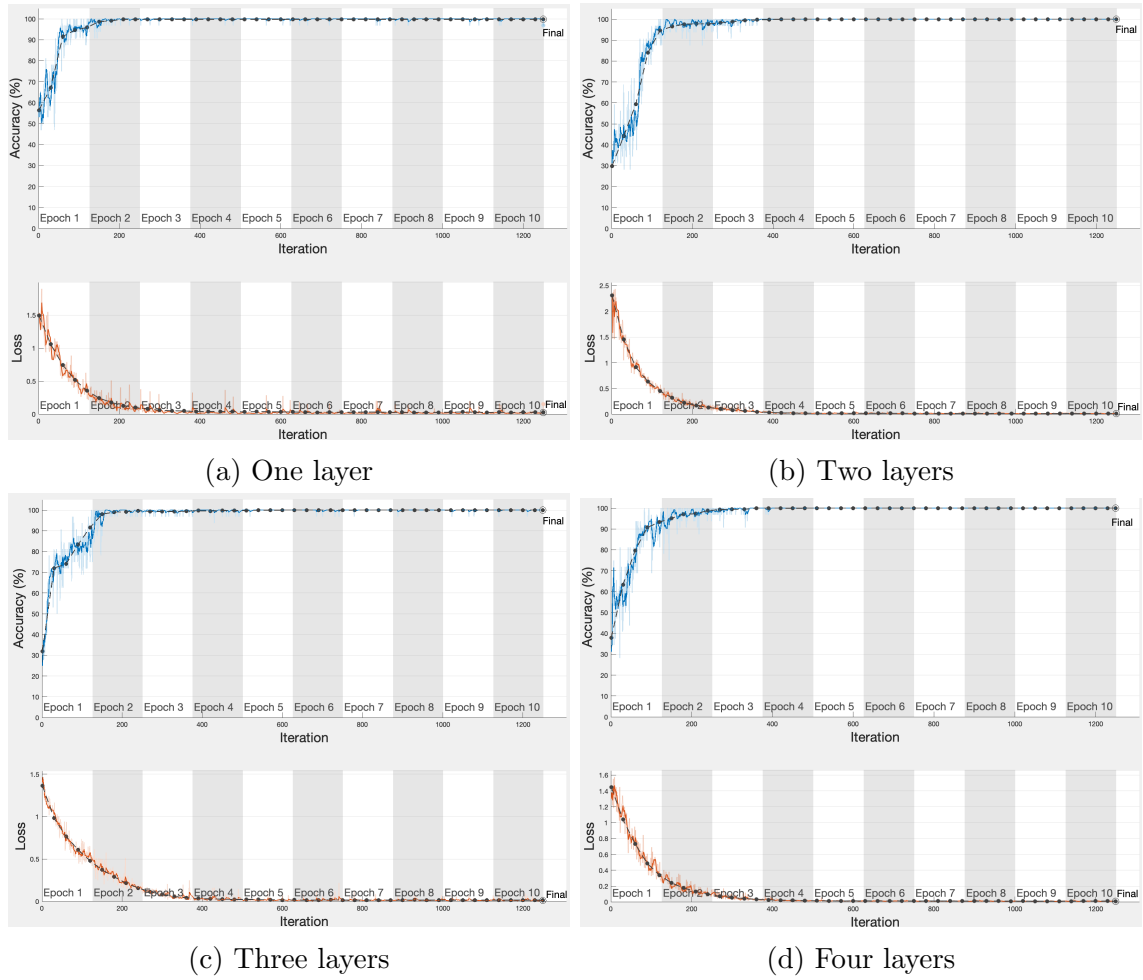
(c) Three layers        (d) Four layers

Figure 13: The Simulation Results of Model-C with Different Number of Layers.

### 4.2.2 Fault Localization Model-L

To localize the fault for the four fault types, the ConvNet model named "Model-L" is developed. Since LLLG is symmetric, only the zero sequence was added in the feature vector of it. There are 1250 dataset for LG and LLG, 1000 dataset for LL and LLLG, respectively. The fault localization performance is evaluated by the localization accuracy (LA) $\eta_l$,

$$\eta_l = \frac{the\ number\ of\ faults\ correctly\ located}{total\ number\ of\ faults} \tag{4-4}$$

Table 5: The ConvNet Structure of Model-L

| Layer NO. | Layer Type | Kernel Size | Output |
|---|---|---|---|
| the 1st layer | Convolution | $[1\ 3] \otimes 24$ | $[1\ 24] \otimes 24$ |
| the 2nd layer | Convolution | $[1\ 3] \otimes 36$ | $[1\ 24] \otimes 36$ |
| | Maxpooling | $[1\ 2]@2$ | $[1\ 12] \otimes 36$ |
| the 3rd layer | Convolution | $[1\ 3] \otimes 48$ | $[1\ 12] \otimes 48$ |
| the 4th layer | Convolution | $[1\ 3] \otimes 48$ | $[1\ 12] \otimes 48$ |
| the 5th layer | Convolution | $[1\ 3] \otimes 60$ | $[1\ 12] \otimes 60$ |
| the 6th layer | Convolution | $[1\ 3] \otimes 60$ | $[1\ 12] \otimes 60$ |
| the 7th layer | Convolution | $[1\ 3] \otimes 72$ | $[1\ 12] \otimes 72$ |
| the 8th layer | Convolution | $[1\ 3] \otimes 72$ | $[1\ 12] \otimes 72$ |
| the 9th layer | Convolution | $[1\ 3] \otimes 84$ | $[112] \otimes 84$ |
| | Maxpooling | $[1\ 2]@2$ | $[1\ 6] \otimes 84$ |
| the 10th layer | Convolution | $[1\ 3] \otimes 96$ | $[1\ 6] \otimes 96$ |
| | Maxpooling | $[1\ 2]@2$ | $[1\ 3] \otimes 96$ |
| fully connected layer | vectorize | ... | 144 |
| fully connected layer | vectorize | ... | 72 |
| fully connected layer | vectorize | ... | 10 |
| Output | Softmax | ... | 10 |

The ConvNet structure of Model-L is shown in Table 5. The kernel size of the convolution operator in each layer is 1 by 3 with the same paddings but they use different number of kernels in each layer. The Maxpooling operation has the same property as the one in Model-C, kernel size of 1 by 2 with the stride size of 2.

During the training process, "Adam" optimizer with the same properties was used. The initial learning rate for Model-L was set at 0.001 with the same drop rate 0.5 and period 10 as Model-C. The simulation results of these faults are presented in Figure 14 and Table 6. The results show that the unbalanced faults with zero sequence current and voltage have better accuracy rate, because the zero sequences of current and voltage provide more information in augmented feature vectors.

Table 6: The simulation summary of Model-L

| Fault Type | LG | LLG | LLLG | LL |
|---|---|---|---|---|
| Elapsed time(s) | 177 | 172 | 172 | 174 |
| Epoch | 50 | 100 | 100 | 100 |
| Accuracy | 85.2% | 98.4% | 77% | 78.5% |
| Starting learning rate | 0.001 | 0.001 | 0.001 | 0.001 |

Meanwhile, we also tested the model with these fault types under different fault resistances. Table 7, 8, 9, 10 show the test results of LG, LL, LLG and LLLG under different fault resistances separately.

Table 7: Test results of LG with different fault resistance

| Fault resistance | $R = 20$ | $R = 15$ | $R = 10$ | $R = 5$ | $R = 0$ |
|---|---|---|---|---|---|
| Accuracy | 94.8% | 88.8% | 91.6% | 93.2% | 94.4% |
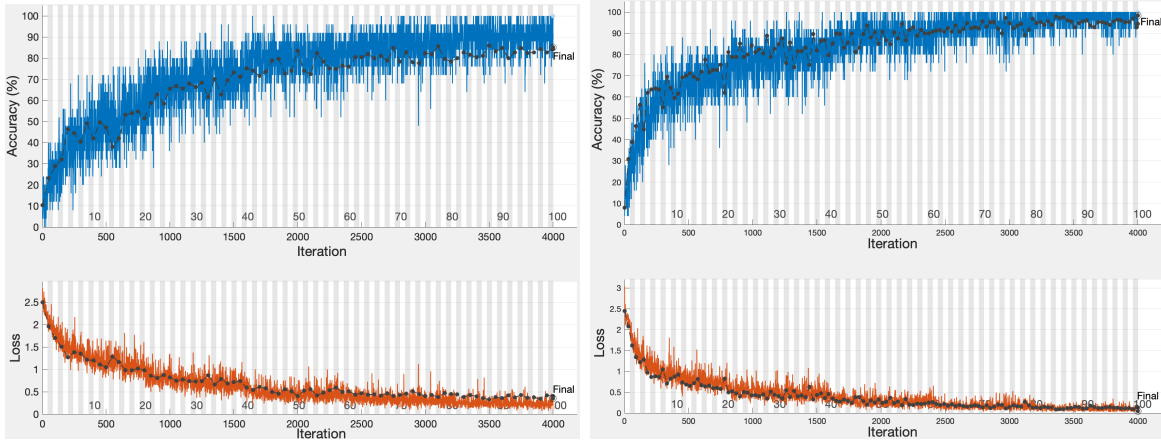
(a) Line to ground fault

(b) Double line to ground fault
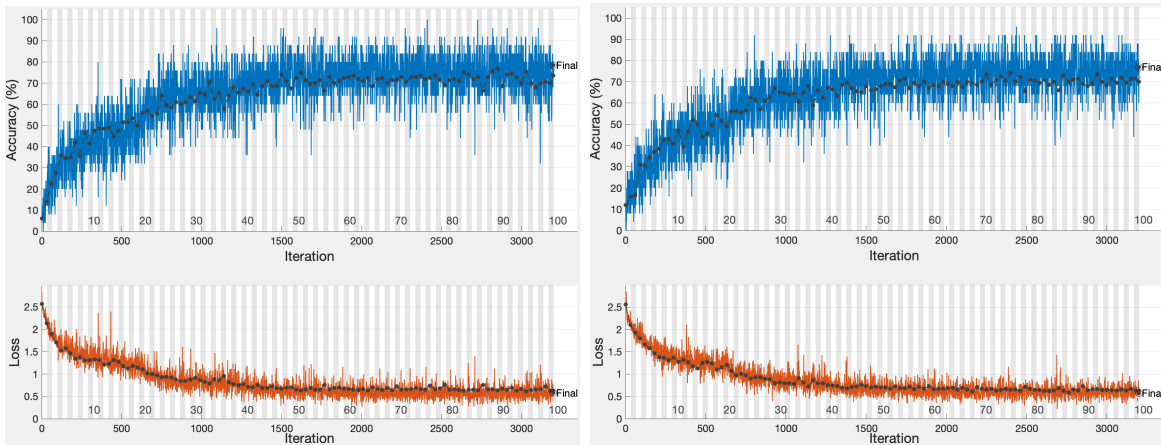
(c) Line to line fault

(d) Three-phase to ground fault

Figure 14: The simulation results of Model-L on different fault types.

Table 8: Test results of LLG with different fault resistance

| Fault resistance | $R_l = 3$ $R_G = 10$ | $R_l = 2$ $R_G = 5$ | $R_l = 5$ $R_G = 0$ | $R_l = 0$ $R_G = 5$ | $R_l = 0$ $R_G = 0$ |
|---|---|---|---|---|---|
| Accuracy | 92.0% | 95.6% | 95.2% | 93.6% | 95.2% |

Table 9: Test results of LLLG with different fault resistance

| Fault resistance | $R = 20$ | $R = 15$ | $R = 10$ | $R = 5$ | $R = 0$ |
|---|---|---|---|---|---|
| Accuracy | 72.5% | 80.4% | 79.0% | 73.6% | 74.8% |

Table 10: Test results of LL with different fault resistance

| Fault resistance | $R = 20$ | $R = 15$ | $R = 10$ | $R = 5$ | $R = 0$ |
|---|---|---|---|---|---|
| Accuracy | 79.8% | 83.5% | 85.6% | 83.5% | 84.4% |

### 4.2.3   Single Observability Validation

During the training process of the ConvNet models, we trained the feature vector of each meter separately which makes the model possess the single observability characteristic. As a result, the models perform well even if there are only partial available meters in the system. To validate this performance of Model-C and Model-L, we take tests on the system with partial meters under 20%, 30%, 50%, 80% observability.

#### 4.2.3.1   Validation of Model-C

We test the validation of Model-C with 20%, 30%, 50% and 80% meters. The results confirm the 100% accuracy rates. It is clear the single observability of Model-C is robust to the system with partial meters.

### 4.2.3.2 Validation of Model-L

Along with testing the single observability of Model-C, we also test it for Model-L with the same percentage of accessible meters. Figure 15 presents the result summary of this model and validates the robustness of the single observability within Model-L.
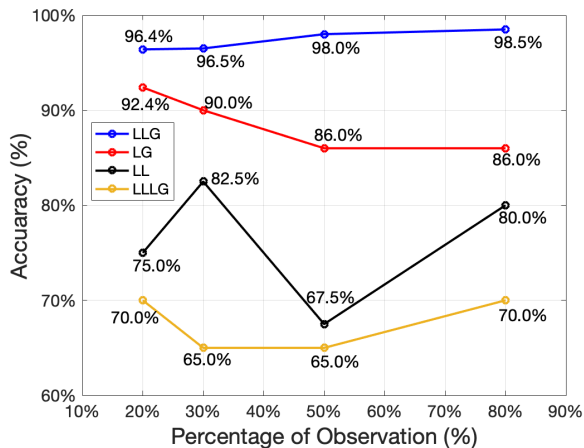


Figure 15: Test results of Model-L with partial observability.

### 4.2.4 Implication of Distributed Generation

Distributed generation (DG) is electrical generation and storage performed by a variety of small, grid-connected or distribution system-connected devices referred to as distributed energy resources (DER). When a DG source is connected to the distribution feeder, the system will be affected and it may perform differently when a fault occurs. To learn about these effects, we measured the current values when different faults occurred with different positions as well as under different lengths between substation and the DG. We assumed the DG has the wye-grounded stator winding configuration. Figure 16 shows the fault current ratio ($I_{f,DG}/I_f$) between the original system when DG is disconnected ($I_f$), and the fault current when DG is connected to the system ($I_{f,DG}$) when the fault distance is varying and length of the main feeder is fixed.

From Figure 16, it is clear that the current in the system is indeed affected when a DG is connected, especially when the feeder length between the substation and DG is longer, the

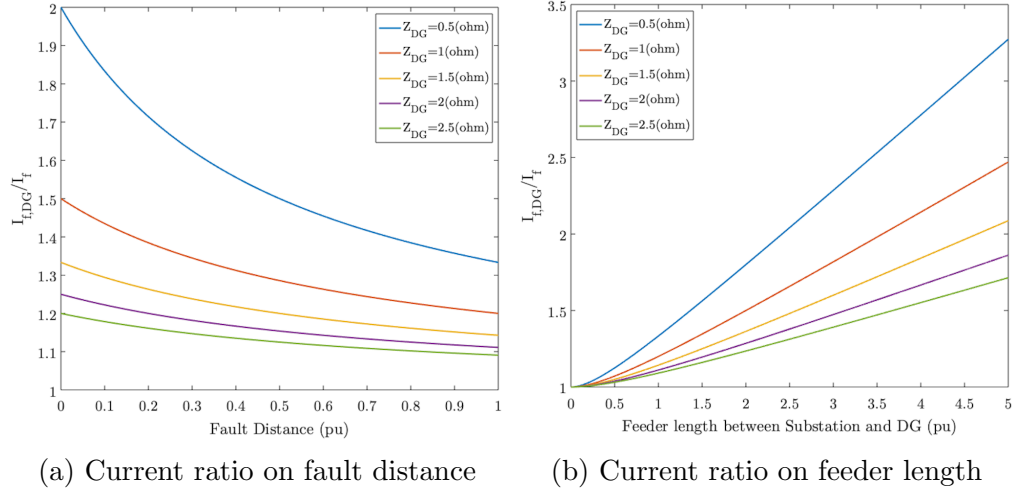(a) Current ratio on fault distance   (b) Current ratio on feeder length

Figure 16: Current ratio on fault distance and feeder length.

fault current ratio is also getting larger, drastically with smaller series impedance behind the voltage of the DG. However, when the fault distance is getting longer (end-far fault location along the feeder), the current ratio decreases. Figure 17 shows the 3-D figure of these aspects that the DG has decreased the short circuit level of the system when the size of the DG is considerable with smaller DG's series impedance. The accuracy of the training and testing processes have also changed with 0.86 percentage error when DG are disconnected.
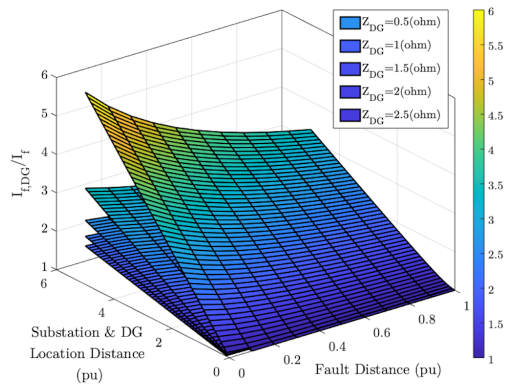


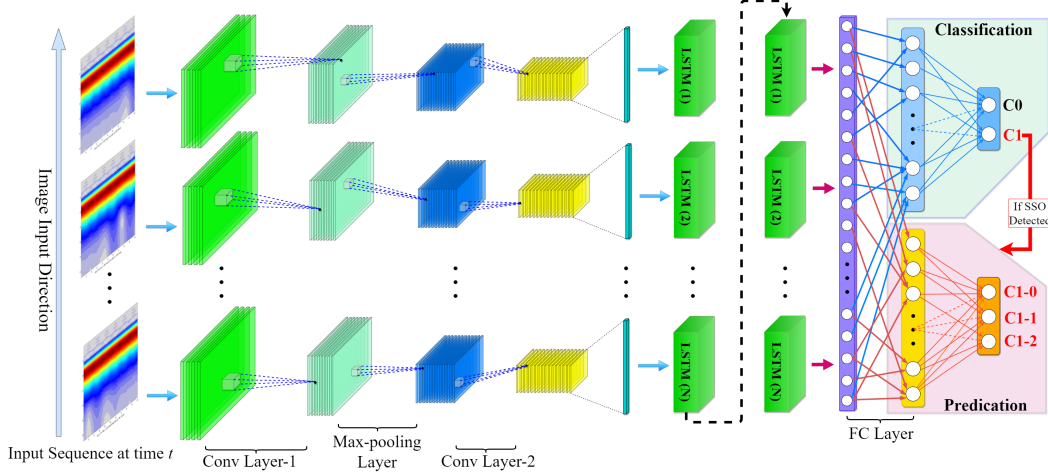Figure 17: 3-D interpretation under these two atmospheres.

Figure 18: Proposed SSO detection and prediction CNN-LSTM architecture

## 4.3 Simulation Results of SSO Detection

### 4.3.1 PCQ-WT and CNN-LSTM Parameter Setting

In this thesis, a sampling rate of $F_s = 2$kHz is used, the complex Morlet PCQ-WT wavelet bank in (3-33) uses 32 scaling factors defined as $a = 2^i$, where $i$ is uniformly selected from 1 to 8. Each wavelet filter has a length of 60 samples (30ms) and shares the same waveform buffer which caches 90 samples (45ms) of the sampled input waveform. For simplicity, set all the time-shift $b_k = 30$ (15ms) for $k = 1, 2, ..., 32$. The central frequency is set as $\omega_m = 2\pi \times 500$rad, and the window shape constant value $\beta$ is chosen to be 200 in (3-29). The PCQ-WT is performed every 5ms, therefore, to cover the feature of a 5Hz SSO's full cycle, a sequence containing 32 feature images is needed. Consequently, 32 small CNN models are used in the proposed CNN-LSTM network, which is shown in Fig. 18. The input of Fig. 18 is an example of extracted SSO waveform feature images via PCQ-WT. We can see that the feature of fundamental component (dark red area) is constant, and peculiarities—the features of SSO—are observed below the fundamental component's features, and they are intermittent since current waveform anomalies caused by SSO are varied slowly along time.

Table 11: Network Configuration of the CNN-LSTM Model

| Layer NO. | Layer Type | Kernel Size | Output |
|---|---|---|---|
| 1st layer | Convolution | [4 8]⊗16 | [15 42]⊗16 |
| 2nd layer | MaxPooling | [2 2]@2 | [8 21]⊗16 |
| 3nd layer | Convolution | [3 3]⊗32 | [3 10]⊗32 |
| 4th layer | LSTM | 50⊗2 | ... |
| FC layer | vectorize | ... | 50 |
| Output1-SSO detection | Softmax | ... | 2 |
| Output2-SSO prediction | Softmax | ... | 4 |

The network configuration of the proposed CNN-LSTM model is shown in Table 11. There are seven layers in the model: convolution layer-1 includes 16 kernels and the kernel size is 4 by 8 which is denoted as $[4\ 8] \otimes 16$; in the MaxPooling layer, the kernel size is 2 by 2 and the stride size is 2 which is denoted as $[2\ 2]@2$; convolution layer-2 contains 32 kernels with the kernel size 3 by 3; in the LSTM layer, the hidden size is set to 50 and 2 recurrent layers are used which are represent as $50 \otimes 2$; the fully-connected (FC) layer is linked with the LSTM layer and it vectorizes the output of the LSTM; after the FC layer, two output layers are set to output the results of the SSO detection and prediction.

The parameters of the test bed system are based on a real case study in radial system. The turbine model is a 1.5MW generic detailed DFIG model for transient studies. And a wind farm of total capacity of 200WM is formed using this model. The line configuration data is from [72].
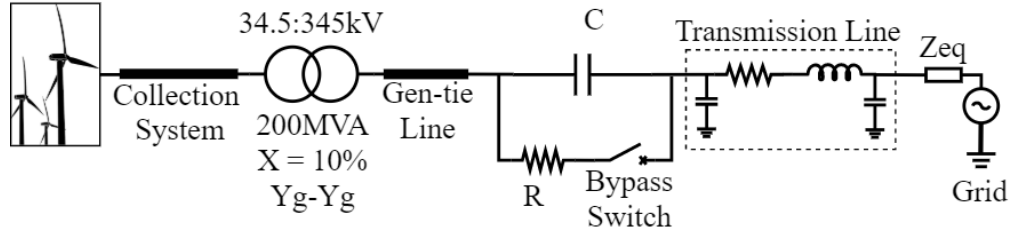
Figure 19: The test radial system with a 200WM wind farm.

### 4.3.2   Test Scenario Specifications and Data Preparation

To generate comprehensive test scenarios for training the proposed framework and make a frequency scan for all possible SSO frequencies of the wind farm, the maximum and minimum parameter specifications such as compensation level of the series capacitor variation, wind speed, number of wind turbine in service, and the equivalent power network condition, are listed in Table 12. When generating each test scenario, all the parameters are uniformly selected in the designated range. In total, 20,000 test scenarios are generated. For each test scenario, the generated current waveform is cut by a sliding window (waveform buffer) of 45ms with overlap 40ms, which matches the PCQ-WT execution rate, i.e., every 5ms. Then the waveform feature in each sliding window is extracted, therefore, in each test scenario, an extremely long feature sequence which consists of many scalograms sequentially in forms of images is generated. As mentioned before, a sequence only containing 32 feature images is used as the input of the proposed CNN-LSTM framework. To avoid confusion, here, we name this 32-feature-image sequence as "input feature sequence". Since each test scenario can generate thousands of input feature sequences, in total, millions of possible input feature sequences could be obtained. To make the detection and prediction actions as well as the proposed CNN-LSTM training process performed in a rationale way, a labeling process is conducted as follows:

- Each extracted *feature image* in the feature sequence is labeled individually according to Label-1 in Table 13.

60

Table 12: Test system parameter specification

| Parameter Category | | Min Value | Max Value |
|---|---|---|---|
| Compensation Level | | 10% | 70% |
| Wind Speed | | 6 | 30 |
| Wind Turbine in Service | | 10% | 100% |
| Equivalent | Positive | 1.48+j0.026 | 5.82+j0.104 |
| Network Impedance | Negative | 8.28+j0.090 | 32.58+j0.356 |

- If the input feature sequence contains feature images involving SSO event, then this input *feature sequence* is labeled by Label-2 in Table 13.

In such way, (i) if the newest input feature image contains SSO-event and all the previous feature images do not, then, an SSO occurrence event is reported, as no more previous information regarding SSO exists, prediction operation is not executed. (ii) If the newest input feature image contains non-SSO event, then a damped SSO event (non-SSO event) can be reported, prediction operation is no longer needed.

The input feature sequence used as the training data for C0, C1-0, C1-1, and C1-2 are randomly selected and balanced. Each class has 21,520 samples, thus, the ratio of data sample non-SSO vs SSO event is 1:3. And the training, validation and testing data ratio is 8:1:1. The workstation used for training the proposed CNN-LSTM model is equiped with Intel Core i7-8700K CPU and Nvidia GeForce GTX 2080Ti GPU.

After the training process is completed, two test cases are used to verify the propopsed scheme:

- Case 1: Additional three SSO events (damping, slow damping and rapid) are generated from the radial system (Fig. 19) and used for online tests. And these three events are absent in the training process.
- Case 2: A portion of the ERCOT grid with high penetration of wind generation which exports power through a series-compensated line from [33] is used to generate SSO events
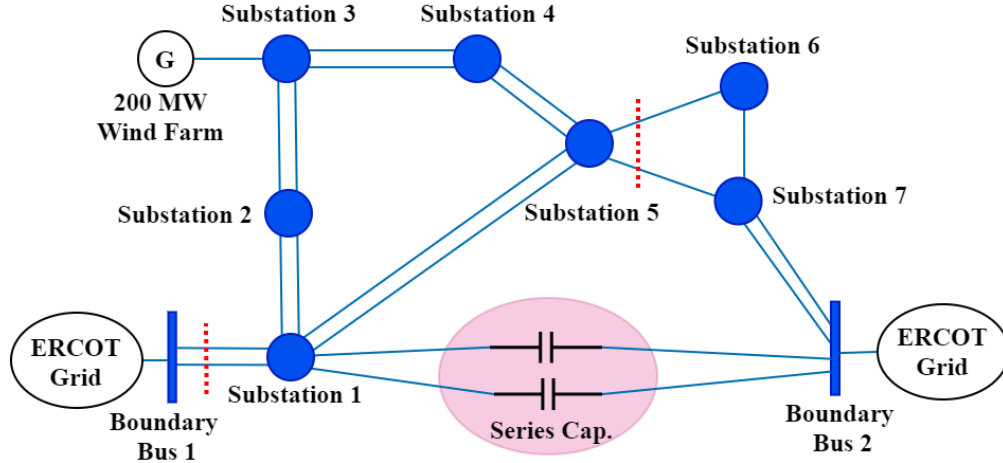
Figure 20: ERCOT grid portion with series-compensated line.

(see Fig. 20). N-4 is the minimum contingency rank that can lead to SSO risk and radial condition in this test bed according to [33]. The definition of this contingency is marked by red dashed line in Fig. 20. Under 100% turbines and 100% dispatch operating conditions, two non-damp SSO events are generated during transmission system with all switch shunts in service and out of service, respectively. Also, the method proposed from [33], which uses Prony and Eigensystem Realization Algorithm (ERA), and a moving FFT, is used for comparison. Here, the method in [33] is named as "*Hybrid*".

### 4.3.3 Experimental Results and Analysis: Case 1

The best detection and prediction accuracy obtained after training is 98.95% and 98.22%, respectively. And the online test results of Case 1 are shown in Fig. 21, Fig. 22 and Fig. 23. Also, the elapsed time of each execution of the proposed framework is monitored and recorded to verify the real-time operating performance of the proposed framework.

Table 13: Feature label

| Label-1 | SSO Event | Specification |
|---------|-----------|---------------|
| C0 | None | No SSO is observed |
| C1 | Detected | SSO observed |

| Label-2 | SSO Event | Specification |
|---------|-----------|---------------|
| C1-0 | Damping | SSO quickly damped within 95ms |
| C1-1 | Non-damping | the active power oscillation failed to damp within 2% of the rated power within 95ms |
| C1-2 | Started | SSO localized within the first observation of the input feature sequence |

#### 4.3.3.1   Non-damping SSO Event

The detection and prediction result of an escalating SSO event is shown in Fig. 21, where the SSO event occurs at $t = 53$ms. The active power in Fig. 21(a) started oscillating at the occurrence of SSO event, and its oscillating amplitude increases at around t = 100ms. The C1 (SSO) event is detected at t = 70 ms in detection result plot (Fig. 21(b)) and the detection module keeps reporting C1 as the SSO event lasts. Meanwhile, for the prediction module, once the C1 event is reported, the C1-2 event (SSO started) is reported at t = 85ms, then a prediction of C1-1 event (non-damping SSO) occurrence is reported at t = 115ms and this prediction result lasts as this escalating SSO event lasts in Fig. 21(c). In this test, the detection time for SSO event is 17ms and the SSO growth prediction time is 62ms. One should notice that, in Fig. 21(d), the elapsed time for the detection and prediction is around or under 2ms, which is less than the execution rate of the entire proposed framework. Therefore, there would be sufficient time redundancy for the whole process.

(a) Active Power

(b) SSO Detection Result

(c) SSO Prediction Result

(d) Elapsed Time
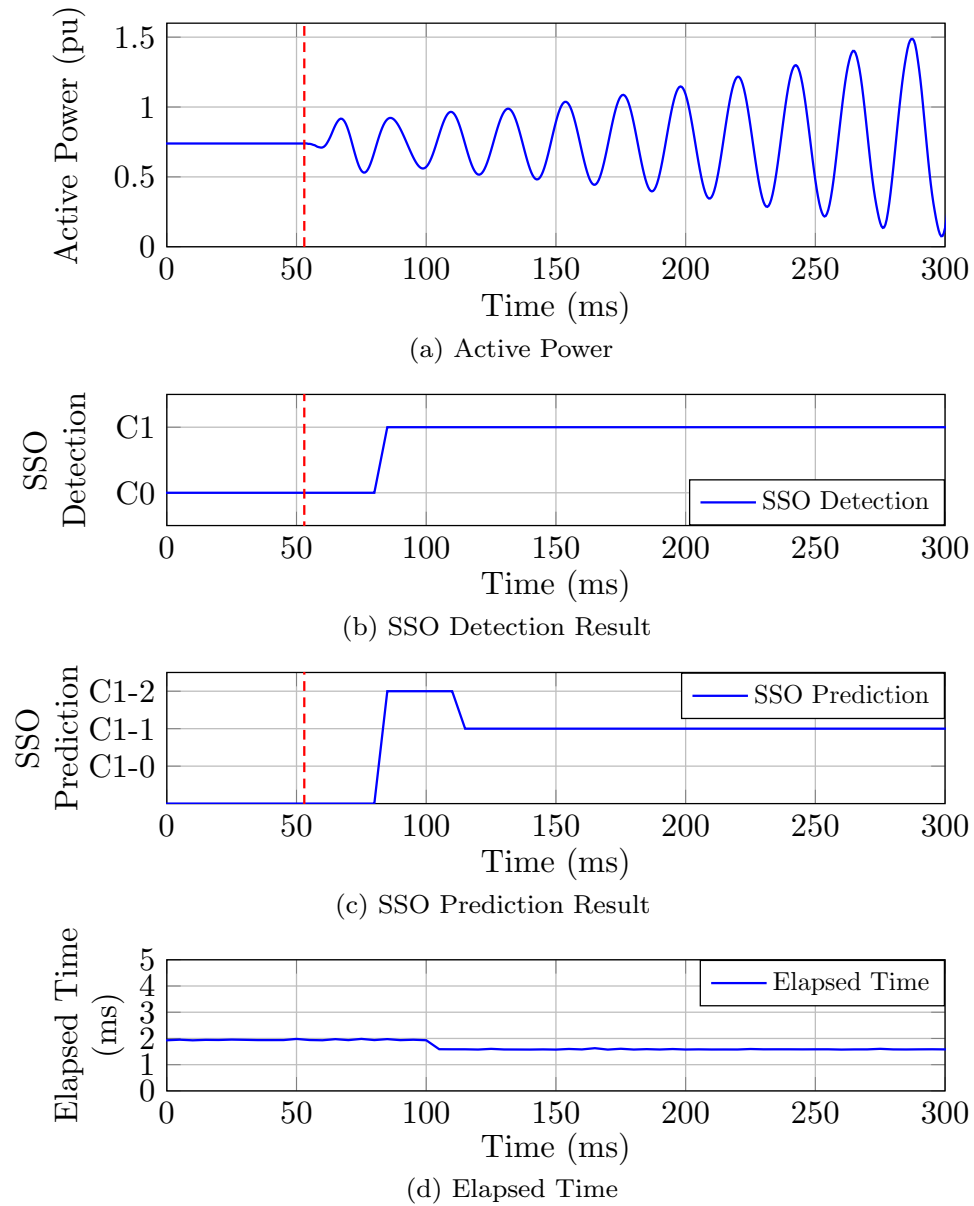
Figure 21: Online test results: Case 1 — escalating SSO event.

(a) Active Power

(b) SSO Detection Result

(c) SSO Prediction Result
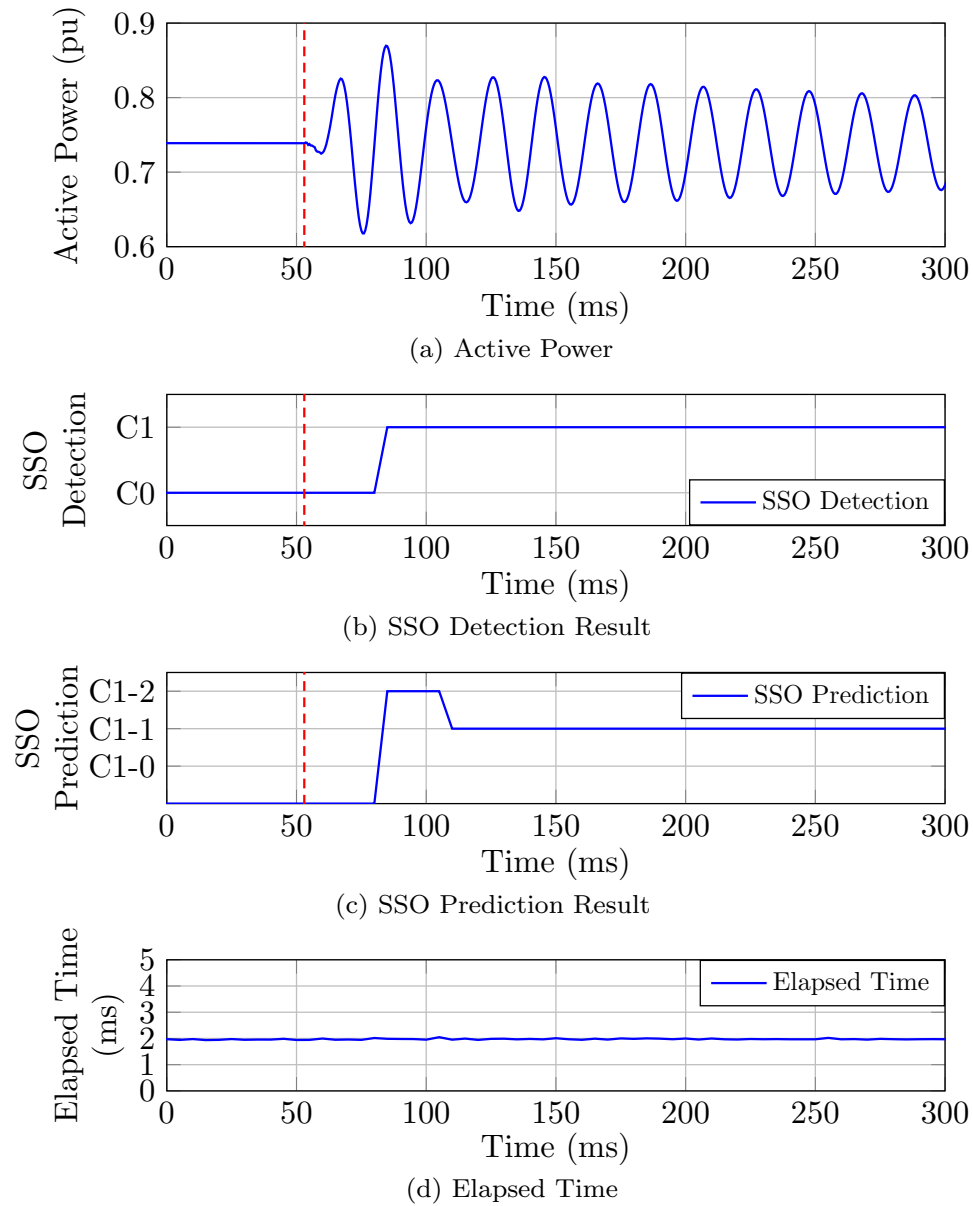
(d) Elapsed Time

Figure 22: Online test results: Case 1 — slow damping SSO event

### 4.3.3.2 Slow-damping SSO Event

During an SSO event with a slow damping amplitude (Fig. 22(a)) occurred at t = 53ms, the amplitude is still above the threshold, therefore, the detection result in Fig. 22(b) constantly reports a C1 event starting at t = 85ms. As the proposed framework is able to predict the growth of the SSO event based on the amplitude trend, even though a slow-damping SSO has occurred but the tendency of the SSO amplitude shows no sign of decaying by the preset time limit (95ms), then a C1-1 prediction is made at t = 110ms in Fig. 22(c). This result proves that the proposed framework can be used for online protection applications since the slow damping SSO can still cause undesired energy exchange and lead fatigue and critical failure on the generator turbine shaft. In this slow-damping SSO event, the overall detection time is 32ms and prediction time is 57ms, which can be potentially used for issuing the trip command for the wind turbines during SSO events. And the elapsed time for each execution of the proposed framework is still around 2ms in Fig. 22(d).

### 4.3.3.3 Quick damped SSO Event

Due to load change or transmission line switching, variation of the output power can be observed in forms of a quick damped SSO event. In this scenario, SSO events might be observed and the corresponded features would reveal, however, the trip actions are unnecessary for this case. So in this test scenario, an rapid damped SSO event is simulated in Fig. 23, where the active power is 0.74 pu before the SSO at t = 53ms (Fig. 23(a)) but its amplitude begins to decrease at t = 80ms and finally damped below the 2% threshold at t = 140ms. In Fig. 23(b) the detection module reports C1 starting at $t$ = 85ms and C0 at t = 175ms. Between t = 85ms and 175ms in Fig. 23(c), the prediction module is active and an SSO growth prediction—C1-0 (damped SS)—is given at t = 115ms and this prediction result ends as the detection reports C0 again. In this test, the detection time is 32ms and 35ms, respectively; the prediction time is 62ms, which again verified the performance of the proposed SSO detection and prediction framework. Also, the elapsed time is around 2ms.
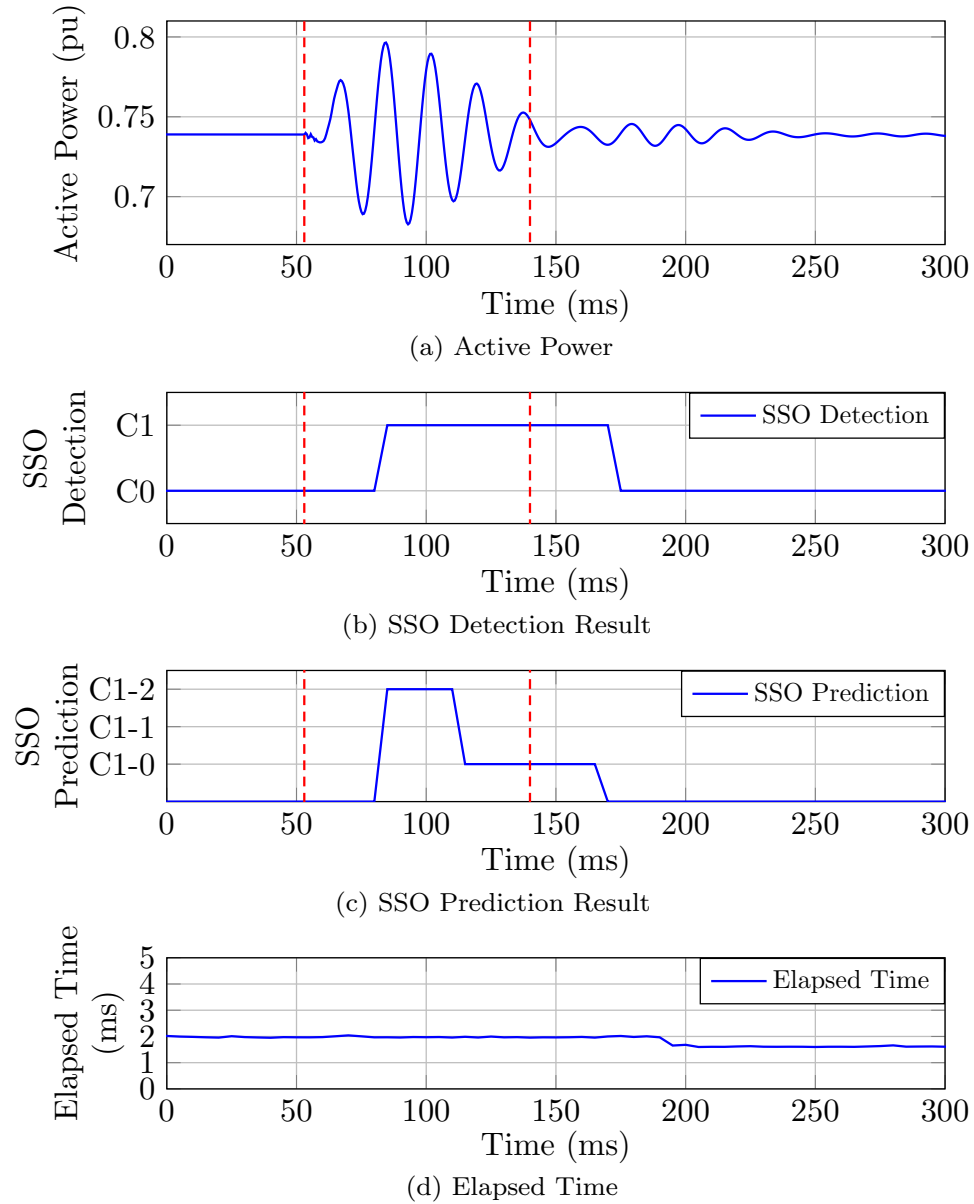
(a) Active Power



(b) SSO Detection Result



(c) SSO Prediction Result



(d) Elapsed Time

Figure 23: Online test results: Case 1 — damping SSO event.

### 4.3.4    Experimental Results and Analysis: Case 2

#### 4.3.4.1    Transmission line switched shunts out of service scenario

The test result of Case 2 under switched shunts out of service scenario is shown in Fig. 24. The occurrence of the radial condition between the wind farm and the series-compensated line is placed at $t = 53$ms (Fig. 24(a)). The frequency of oscillation is detected at 12Hz. The SSO event is reported at t = 70ms for the proposed approach (see Fig. 24(b)), which is only 17ms after the occurrence of the SSO event; the prediction module provides the non-damping alert at $t = 105$ms (Fig. 24(c)), i.e., 52ms after non-damping SSO occurred. One should notice that, the result from the proposed approach is reported every 5ms, so slops can be observed in the detection and prediction results. For "Hybrid" method, the SSO detection is reported at $t = 85$ms, which is 32ms after SSO starts. In "Hybrid" method, the SSO non-damping alert (protection signal) is always designated to be sent after there cycles (50ms) of the SSO detection is reported, which is equivalent to the non-damping prediction used in this paper at Fig. 24(c). So the earliest protection action can be issued at $t = 135$ms from "Hybrid" method, which 82ms after SSO starts equivalently in this paper.

#### 4.3.4.2    Transmission line switched shunts in service scenario

The frequency of oscillation is detected at 15Hz in test results of Case 2 with switched shunts in service scenario. Similarly, the occurrence of SSO is shifted and placed at $t = 53$ms in Fig. 25(a). In this test scenario, the SSO detection of the proposed method is at $t = 75$ms (22ms after SSO starts, see Fig. 25(b)) and the non-damping prediction is first reported at $t = 105$ms (52ms after SSO starts, see Fig. 25(c)). And for "Hybrid" method. the detection and non-damping prediction are reported at $t = 90$ and $t = 140$ms respectively, which is 37ms and 87ms after the SSO event starts.

In all, the test results in Case 1 and Case 2 verified the efficiency and performance of the proposed approach. And the comparison in Case 2 with and without switched shunts in service confirms the performance improvement of the proposed work in this paper. Also, during a transition period when SSO starts, as the waveform which contains the SSO feature

(a) Active Power

(b) SSO Detection Result

(c) SSO Prediction Result

(d) Elapsed Time
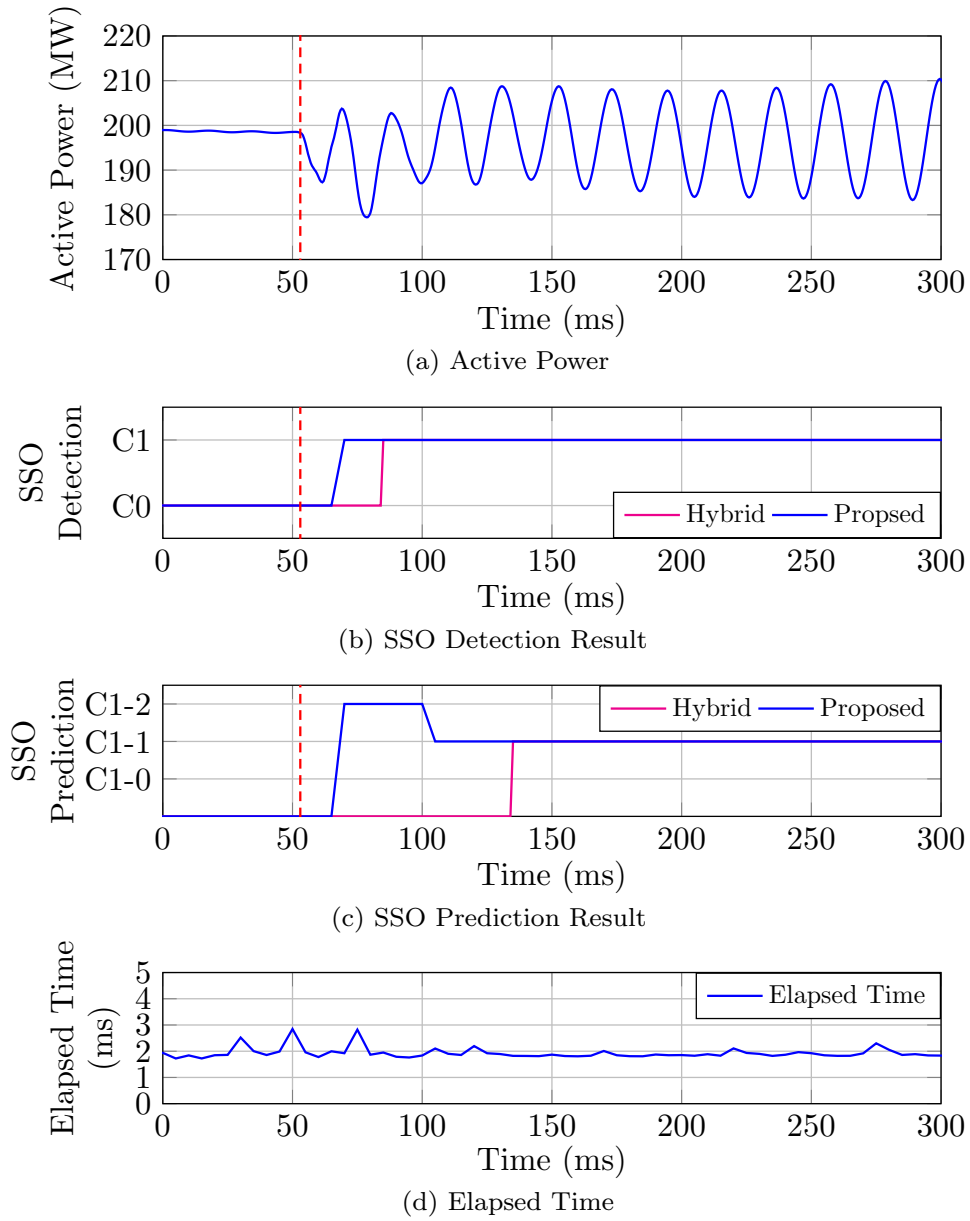
Figure 24: Case 2: test results under switched shunts out of service.

69

(a) Active Power

(b) SSO Detection Result

(c) SSO Prediction Result
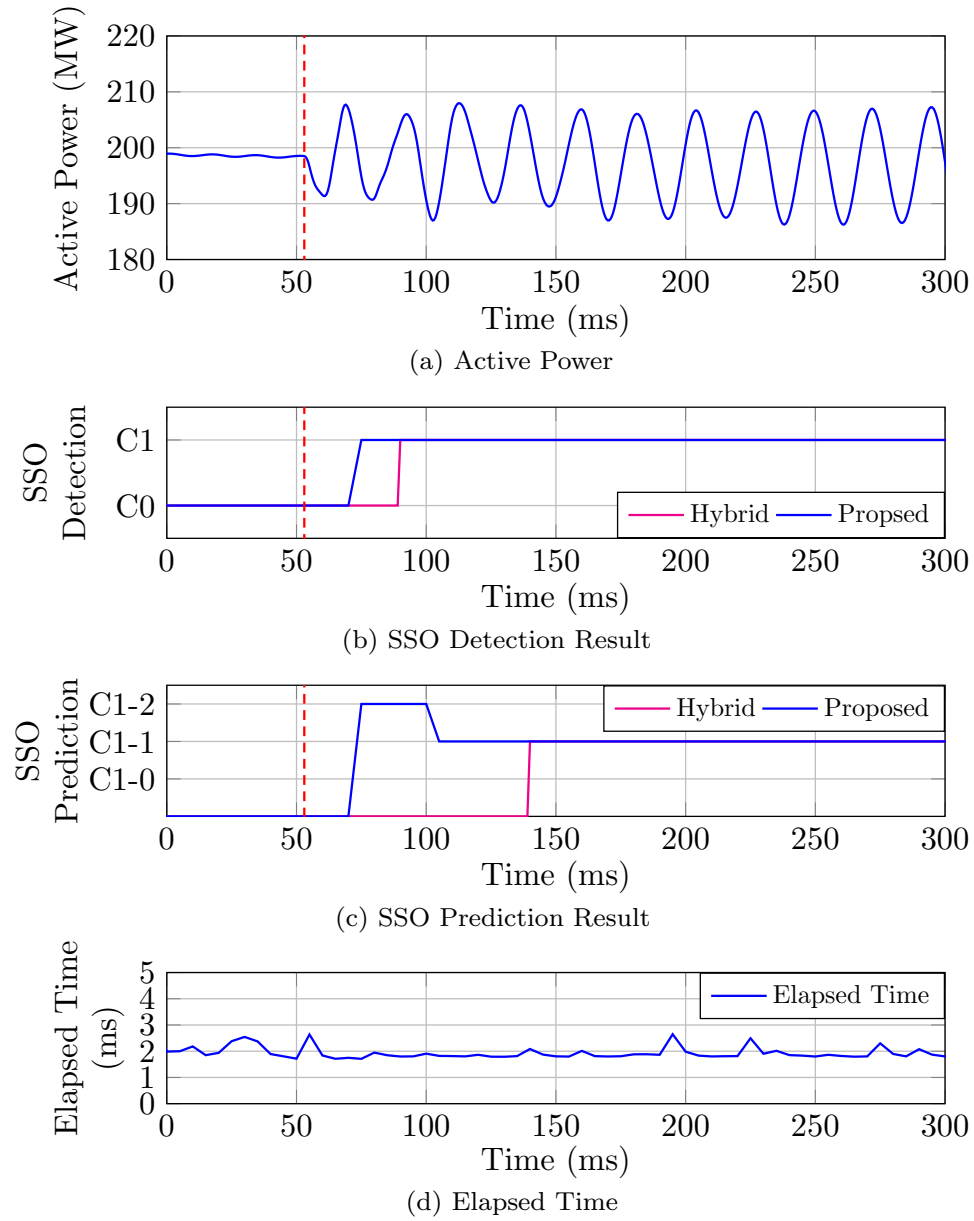
(d) Elapsed Time

Figure 25: Case 2: test results under switched shunts in of service.

takes time to fully move into the waveform data buffer which is equivalent to 45ms in length, thus a detection delay is observed. However, a noticeable finding is that, the proposed approach itself shows improved performance in Case 2 compared to Case 1, this is because at the beginning of SSO in Case 2, the active power reveals immediate variations, which indicates that, there are sudden changes in current waveform caused by SSO. The proposed work utilizes and processes feature extraction from the current waveform, so the sudden changes in current waveform would provide more distinctive features that are in support of SSO detection for the proposed approach.

The prediction module in this paper requires a sequence of feature images to gather sufficient information, thus the prediction time is slightly longer than the detection time, but the prediction result can be provided in 62ms at maximum. Nevertheless, there would be a potential to improve the detection time and prediction time as the execution rate of the proposed framework can be increased. But even using such a conservative execution interval of 5ms, this proposed framework still provides promising performances in all the experimental tests, and therefore, it can be applied for the SSO online monitoring and protection applications when connected with protective relays.

# 5.0    Conclusions and Future Work

## 5.1    Conclusions

In this thesis, three topics are conducted to improve the optimal operation, security and resiliency of the power system: 1) The proposed SLP and SLPBB algorithms are able to find the global near-optimal solutions for the non-convex ACOPF problems. Preliminary experiments illustrated the effectiveness of these two algorithms, and the results also satisfied the KKT conditions; 2) Two intelligent models (Model-C & Model-L) based on ConvNet technique are developed to classify the fault and determine its location in power distribution systems. The main achievements of this method are: quick fault type classification in 100% accuracy; meaningful and significant feature vectors' extraction for the fault localization in both $(a, b, c)$ and $(Z, N, P)$ domains as a function of fault distance $d$; robust and more effective ConvNet with single observability characteristic for different fault types (average accuracy of 85%). As a result, these two intelligent models are more practical and economical when there are limited micro-PMUs installed, and limited access to the data in the system; 3) A novel SSO detection & prediction framework that utilizes waveform feature extraction and deep learning techniques is , which can detect the SSO within the 45ms input waveform buffer length and provide trip signal within 70ms. This framework detects the SSO by using waveform features and bypasses the frequency measurement process, which can be applied to monitor the system with different configurations if the system model is provided. And this approach is built under the developers' specification and settings, thus it can be adjusted to different protective requirements. Test results prove that the framework is capable of detecting the presence of SSO and providing high accuracy in both SSO detection & prediction for different dispatch capacities. Therefore, this framework is very promising in real-time SSO monitoring applications. Also, this proposed design can be used to assist relays to guarantee a reliable trip action or used independently for providing trip commands.

## 5.2 Future Work

In the future, some improvements of these three research areas will be considered: 1) For the ACOPF problems, the complexity of SLPBB algorithm grows exponentially in terms of the increasing number of negative eigenvalues of the Hessian matrix in the objective function. This indicates SLPBB may not be effective for larger scale power systems because of the large number of negative eigenvalues. For the future researches, one interesting direction is to study whether we can develop effective global algorithms for generic QP problems, as well as further investigation on the minimum number of penalized terms added in the objective function; 2) About the fault recognition problems, the single observability characteristic causes the accuracy rate of some fault types and situations is not much high (less than 90%). This is one problem we will improve in the future. Next, we will work more on the feature vectors and ConvNet optimization for fault localization to improve its accuracy rate, especially with single observability and limited access to the data. Also, we will take researches on computing the distance between the fault and the terminal bus to provide more exact information, as well as validate our methodology on real daily-life data, not merely the simulated ones; 3) For the SSO detection & recognition problem, current intelligent model is based on ConvNet and LSTM algorithms. However, there are some limitations about LSTM such as it cannot provide proper attention to lengthy input sequences. As an attention-based technique, transformer can avoid this problem and perform better than LSTM. Therefore, transformer technique will be considered to replace the LSTM in the future research, and develop a more efficient intelligent model for SSO detections.

# Bibliography

[1]     I. Penn. Power lines and electrical equipment are a leading cause of california wildfires. *Los Angeles Times*, 2017.

[2]     Christopher Olah. Understanding lstm networks, 2015.

[3]     B. C. Lesieutre and I. A. Hiskens. Convexity of the set of feasible injections and revenue adequacy in ftr markets. *IEEE Transactions on Power Systems*, 20(4):1790–1798, 2005.

[4]     J. Lavaei and S. H. Low. Zero duality gap in optimal power flow problem. *IEEE Transactions on Power Systems*, 27(1):92–107, 2012.

[5]     H. Hijazi, Carleton Coffrin, and Pascal Van Hentenryck. Polynomial SDP cuts for Optimal Power Flow. *2016 Power Systems Computation Conference (PSCC)*, pages 1–7, 2016.

[6]     B. Kocuk, S. Dey, and X. Sun. Strong SOCP relaxations for the Optimal Power Flow problem. *Oper. Res.*, 64:1177–1196, 2016.

[7]     L. Bobo, A. Venzke, and S. Chatzivasileiadis. Second-order cone relaxations of the Optimal Power Flow for active distribution grids. *ArXiv*, 2020.

[8]     Minyue Ma, Lingling Fan, Zhixin Miao, Bo Zeng, and H. Ghassempour. A sparse convex AC OPF solver and convex iteration implementation based on 3-node cycles. *Electric Power Systems Research*, 180, 2020.

[9]     S. Sojoudi, R. Madani, and J. Lavaei. Low-rank solution of convex relaxation for optimal power flow problem. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 636–641, 2013.

[10]    R. Madani, M. Ashraphijuo, and J. Lavaei. Promises of conic relaxation for contingency-constrained optimal power flow problem. *IEEE Transactions on Power Systems*, 31(2):1297–1307, 2016.

[11]  F. Zohrizadeh, M. Kheirandishfard, E. Q. Jnr, and R. Madani. Penalized parabolic relaxation for optimal power flow problem. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1616–1623, 2018.

[12]  W. A. Bukhsh, A. Grothey, K. I. M. McKinnon, and P. A. Trodden. Local solutions of the optimal power flow problem. *IEEE Transactions on Power Systems*, 28(4):4780–4788, 2013.

[13]  M. Kezunovic and I. Rikalo. Detect and classify faults using neural nets. *IEEE Computer Applications in Power*, 9(4):42–47, 1996.

[14]  Q. H. Alsafasfeh, I. Abdel-Qader, and A. M. harb. Fault classification and localization in power systems using fault signatures and principal components analysis. *Energy and Power Engineering*, 4(6):506–522, 2012.

[15]  Ieee guide for performing arc-flash hazard calculations. *IEEE Std 1584-2018 (Revision of IEEE Std 1584-2002)*, pages 1–134, 2018.

[16]  W. Atkinson. The link between power lines and wildfires, 2018.

[17]  E. R. Sanseverino, V. Li Vigni, A. Di Stefano, and R. Candela. A Two-End Traveling Wave Fault Location System for MV Cables. *IEEE Transactions on Industry Applications*, 55(2):1180–1188, 2019.

[18]  Mohamed Gabr, D. Ibrahim, E. S. Ahmed, and M. Gilany. A New Impedance-based Fault Location Scheme for Overhead Unbalanced Radial Distribution Networks. *Electric Power Systems Research*, 142:153–162, 2017.

[19]  S. S. Gururajapathy, H. Mokhlis, and H. A. Illias. Fault location and detection techniques in power distribution systems with distributed generation: A review. *Renewable and Sustainable Energy Reviews*, 74:949–958, 2017.

[20]  S. Wang, P. Dehghanian, and L. Li. Power grid online surveillance through PMU-embedded convolutional neural networks. *IEEE Transactions on Industry Applications*, 56(2):1146–1155, March 2020.

[21]  M. H. Rezaeian Koochi, P. Dehghanian, S. Esmaeili, P. Dehghanian, and S. Wang. A synchrophasor-based decision tree approach for identification of most coherent generating units. In *Annual Conference of the IEEE Industrial Electronics Society*, pages 71–76, Oct 2018.

[22]    S. Wang, P. Dehghanian, L. Li, and B. Wang.  A machine learning approach to detection of geomagnetically induced currents in power grids. *IEEE Transactions on Industry Applications*, 56(2):1098–1106, March 2020.

[23]    B. Shinde, S. Wang, P. Dehghanian, and M. Babakmehr.  Real-time detection of critical generators in power systems: A deep learning HCP approach. In *2020 IEEE Texas Power and Energy Conference (TPEC)*, pages 1–6, Feb 2020.

[24]    T. S. Rao, S. T. Ram, and J. Subrahmanyam.  Fault Recognition and Diagnosis Based on a Decision Tree for Power Distribution Systems. *Journal of Computational Mechanics, Power System and Control*, 2(2), 2019.

[25]    A. Pessoa and M. Oleskovicz. Fault Location in Radial Distribution Systems based on Decision Trees and Optimized Allocation of Power Quality Meters. *2017 IEEE Manchester PowerTech*, pages 1–6, 2017.

[26]    M. Dashtdar, R. Dashti, and H. R. Shaker. Distribution Network Fault Section Identification and Fault Location using Artificial Neural Network. *2018 5th International Conference on Electrical and Electronic Engineering (ICEEE)*, pages 273–278, 2018.

[27]    S. Gururajapathy, H. Mokhlis, H. Illias, and L. J. Awalin.  Support Vector Classification and Regression for Fault Location in Distribution System using Voltage Sag Profile. *IEEJ Transactions on Electrical and Electronic Engineering*, 12:519–526, 2017.

[28]    R. Perez, E. Inga, A. Aguila, C. Vásquez, L. Lima, Amelec Viloria, and M. Henry. Fault Diagnosis on Electrical Distribution Systems based on Fuzzy Logic. In *ICSI*, 2018.

[29]    J. Liang, T. Jing, H. Niu, and J. Wang.  Two-Terminal Fault Location Method of Distribution Network based on Adaptive Convolution Neural Network. *IEEE Access*, 8:54035–54043, 2020.

[30]    J. C. Gu, Z. J. Huang, J. M. Wang, L. C. Hsu, and M. T. Yang. High Impedance Fault Detection in Overhead Distribution Feeders Using a DSP-Based Feeder Terminal Unit. *IEEE Transactions on Industry Applications*, 2020.

[31]    W. Li, D. Deka, M. Chertkov, and M. Wang.  Real-time faulted line localization and pmu placement in power systems through convolutional neural networks. *IEEE Transactions on Power Systems*, 34(6):4640–4651, 2019.

[32]  G. D. Irwin, A. K. Jindal, and A. L. Isaacs. Sub-synchronous control interactions between type 3 wind turbines and series compensated ac transmission systems. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–6, 2011.

[33]  F. Salehi, I. B. M. Matsuo, A. Brahman, M. A. Tabrizi, and W. Lee. Sub-synchronous control interaction detection: A real-time application. *IEEE Transactions on Power Delivery*, 35(1):106–116, 2020.

[34]  Z. Zhang, S. Liu, G. Zhu, and Z. Lu. Ssci detection and protection in doubly fed generator based on dtft. *The Journal of Engineering*, 2017(13):2104–2107, 2017.

[35]  H. Li, M. Abdeen, Z. Chai, S. Kamel, X. Xie, Y. Hu, and K. Wang. An improved fast detection method on subsynchronous resonance in a wind power system with a series compensated transmission line. *IEEE Access*, 7:61512–61522, 2019.

[36]  T. Rajaram, J. M. Reddy, and Y. Xu. Kalman filter based detection and mitigation of subsynchronous resonance with sssc. *IEEE Transactions on Power Systems*, 32(2):1400–1409, 2017.

[37]  B. Gao, R. Torquato, W. Xu, and W. Freitas. Waveform-based method for fast and accurate identification of subsynchronous resonance events. *IEEE Transactions on Power Systems*, 34(5):3626–3636, 2019.

[38]  B. Badrzadeh, M. Sahni, Y. Zhou, D. Muthumuni, and A. Gole. General methodology for analysis of sub-synchronous interaction in wind power plants. *IEEE Transactions on Power Systems*, 28(2):1858–1869, 2013.

[39]  X. Xie, W. Liu, H. Liu, Y. Du, and Y. Li. A system-wide protection against unstable ssci in series-compensated wind power systems. *IEEE Transactions on Power Delivery*, 33(6):3095–3104, 2018.

[40]  H. Cho, S. Oh, S. Nam, and B. Lee. Non-linear dynamics based sub-synchronous resonance index by using power system measurement data. *IET Generation, Transmission Distribution*, 12(17):4026–4033, 2018.

[41]  S. Wang and P. Dehghanian. On the use of artificial intelligence for high impedance fault detection and electrical safety. *IEEE Transactions on Industry Applications*, pages 1–1, 2020.

[42] Shiyuan Wang, Payman Dehghanian, Li Li, and Bo Wang. A machine learning approach to detection of geomagnetically induced currents in power grids. *IEEE Transactions on Industry Applications*, 56(2):1098–1106, 2019.

[43] Shiyuan Wang, Payman Dehghanian, and Bei Zhang. A data-driven algorithm for online power grid topology change identification with PMUs. In *2019 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2019.

[44] Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. Soft + hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection. *Neural Networks*, 108:466 – 478, 2018.

[45] Yi-Ching Su, Kuo-Lung Lian, and Hsueh-Hsien Chang. Feature selection of nonintrusive load monitoring system using stft and wavelet transform. In *e-Business Engineering (ICEBE), 2011 IEEE 8th International Conference on*, pages 293–298. IEEE, 2011.

[46] Sheng-Huoo Ni, Kuo-Feng Lo, Lutz Lehmann, and Yan-Hong Huang. Time–frequency analyses of pile-integrity testing using wavelet transform. *Computers and Geotechnics*, 35(4):600–607, 2008.

[47] Shiyuan Wang, Payman Dehghanian, Li Li, and Bo Wang. A machine learning approach to detection of geomagnetically induced currents in power grids. In *IEEE Industry Applications Society Annual Meeting (IAS)*, pages 1–7, 2019.

[48] D. P. Mishra, S. R. Samantaray, and G. Joos. A combined wavelet and data-mining based intelligent protection scheme for microgrid. *IEEE Transactions on Smart Grid*, 7(5):2295–2304, Sept 2016.

[49] K. Thirumala, M. S. Prasad, T. Jain, and A. C. Umarikar. Tunable-q wavelet transform and dual multiclass svm for online automatic detection of power quality disturbances. *IEEE Transactions on Smart Grid*, 9(4):3018–3028, July 2018.

[50] Samuel Burer and Renato D. C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95:329–357, 2003.

[51] Joel A. Tropp, Alp Yurtsever, Madeleine Udell, and Volkan Cevher. Practical sketching algorithms for low-rank matrix approximation. *SIAM J. Matrix Anal. Appl.*, 38:1454–1485, 2016.

[52] Samuel Burer and Dieter Vandenbussche. A finite branch-and-bound algorithm for nonconvex quadratic programming via semidefinite relaxations. *Mathematical Programming*, 113:259–282, 2008.

[53] Samuel Burer and Dieter Vandenbussche. Globally solving box-constrained nonconvex quadratic programs with semidefinite-based finite branch-and-bound. *Computational Optimization and Applications*, 43:181–195, 2009.

[54] Jieqiu Chen and Samuel Burer. Globally solving nonconvex quadratic programming problems via completely positive programming. *Mathematical Programming Computation*, 4:33–52, 2012.

[55] Fernando Palacios-Gómez, Leon S. Lasdon, and Michael Engquist. Nonlinear optimization by successive linear programming. *Management Science*, 28:1106–1120, 1982.

[56] R. Horst and N. V. Thoai. Dc programming: Overview. *Journal of Optimization Theory and Applications*, 103(1):1–43, 1999.

[57] S. Boyd and L. Vandenberghe. *Convex Optimization*. 2004.

[58] C. A. Floudas and V. Visweswaran. Quadratic optimization. *Handbook of Global Optimization*, pages 217–270, 1994.

[59] J. Gorski, F. Pfeuffer, and K. Klamroth. Biconvex sets and optimization with biconvex functions: A survey and extensions. *Mathematical Methods of Operations research*, 66:373–407, 2007.

[60] V. Visweswaran and C. A. Floudas. New properties and computational improvement of the gop algorithm for problems with quadratic objective function and constraints. *Journal of Global Optimization*, 3:439–462, 1993.

[61] P. Loridan. Necessary conditions for $\epsilon$-optimality. *Mathematical Programming Studies*, 19:140–152, 1982.

[62] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.

[63]   M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5-6):555–559, 2003.

[64]   W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Neural Networks*, 29(32):4790–4797, 1990.

[65]   R. Venkatesan and B. Li. *Convolutional Neural Networks in Visual Computing: A Concise Guide*. CRC Press, 2017.

[66]   John Hubbard. Calculation of partition functions. *Physical Review Letters*, 3(2):77, 1959.

[67]   Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.

[68]   P. Li, M. Abdel-Aty, and Jinghui Yuan. Real-time crash risk prediction on arterials based on lstm-cnn. *Accident; analysis and prevention*, 135:105371, 2019.

[69]   A. Graves. Supervised sequence labelling with recurrent neural networks. In *Studies in Computational Intelligence*, 2008.

[70]   W. A. Bukhsh, A. Grothey, K. I. M. McKinnon, and P. A. Trodden. Local solutions of the optimal power flow problem. *IEEE Transactions on Power Systems*, 28(4):4780–4788, 2013.

[71]   R. Madani, S. Sojoudi, and J. Lavaei. Convex relaxation for optimal power flow problem: Mesh networks. *IEEE Transactions on Power Systems*, 30(1):199–211, 2015.

[72]   H. K. Nia, F. Salehi, M. Sahni, N. Karnik, and H. Yin. A filter-less robust controller for damping ssci oscillations in wind power plants. In *2017 IEEE Power Energy Society General Meeting*, pages 1–1, 2017.