Western&Graduate&PostdoctoralStudies

Electronic Thesis and Dissertation Repository

8-11-2021 9:45 AM

Deep Learning for High-Impedance Fault Detection and Classification

Khushwant Rai, *The University of Western Ontario*

Supervisor: Grolinger, Katarina, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Engineering
Science degree in Electrical and Computer Engineering
© Khushwant Rai 2021

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Other Computer Engineering Commons, Other Electrical and Computer Engineering Commons, and the Power and Energy Commons

Recommended Citation

Abstract

High-Impedance Faults (HIFs) are a hazard to public safety but are difficult to detect because of their low current amplitude and diverse characteristics. Supervised machine learning techniques have shown great success in HIF detection; however, these approaches rely on resource-intensive signal processing techniques and fail in presence of non-HIF disturbances and even for scenarios not included in training data. This thesis leverages unsupervised learning and proposes a Convolutional Autoencoder framework for HIF Detection (CAE-HIFD). In CAE-HIFD, Convolutional Autoencoder learns only from HIF signals by employing cross-correlation; consequently, eliminating the need for diverse non-HIF scenarios in training. Furthermore, this thesis proposes a novel HIF classification approach based on the transformer network stacked with the convolution neural network. To discriminate HIFs from non-fault disturbances, probability distribution-based kurtosis analysis is utilized. The proposed approaches reliably detect HIFs with 100% success rate in terms of all five metrics of protection system performance, namely accuracy, security, dependability, safety, and sensitivity. The evaluation studies show that proposed approaches outperform the state-of-the-art HIF detection techniques and are robust against noise.

Keywords: High-Impedance Fault; Power System Protection; Deep Learning; Convolutional Autoencoder; Convolutional Neural Network; Transformer Network

Summary for Lay Audience

A High-Impedance Fault (HIF) typically occurs in a distribution network when a live conductor in a distribution network touches surfaces, such as tree limbs, ground, and gravel. The HIF current is usually too low to cause any direct damage to the power system equipment; however, undetected HIFs are a potential hazard to public safety. The inception of HIFs often causes arcing ignition, which can result in wildfires, life-threatening injuries, and other irreparable damages. Moreover, the undetected HIFs cause grid instability and hinder power system integrity. The traditional protection relays fail to reliably detect the majority of HIFs because of diverse characteristics and the low current magnitude of HIFs. The existing approaches for HIF detection are associated with resource-intensive signal processing techniques and supervised Machine Learning algorithms that are not reliable under diverse non-HIF disturbances. Consequently, this thesis proposes Convolutional Autoencoder framework for HIF detection (CAE-HIFD), an unsupervised learning-based approach that solely learns from the fault data and omits the requirement of the diverse non-fault scenarios during the training process. Also, this thesis proposes a Transformer-CNN (T-CNN) framework for HIF detection and classification, a deep learning-based approach for reliable identification of fault type in the presence of diverse HIF and non-HIF conditions without requiring resource-intensive signal processing. The non-fault disturbances, such as capacitor and load switching, exhibit characteristics similar to the HIFs; therefore, a probability distribution-based kurtosis analysis is utilized to provide security against non-HIF disturbances. The performance of the proposed approaches is evaluated on the diverse data generated by the IEEE 13-node test feeder, and the response is measure on various challenging case studies. The results show that CAE-HIFD and T-CNN achieve 100% accuracy and outperformed the state-of-the-art approaches for HIF detection.

Acknowledgements

Above all, I would like to express my sincere gratitude to my supervisor, Dr. Katarina Grolinger, for giving me the opportunity to do research and providing invaluable guidance throughout this research.

I am profoundly grateful for her continued support, guidance, and encouragement. Her vision, sincerity, and motivation have deeply inspired me. I am thankful to her for placing faith in me, without which none of this research would have been possible. It has been a privilege to work and study under her guidance.

I would like to thank my family, my parents, Pardeep Kumar and Satinder, whose love, support, and blessings have driven me to work hard persistently and strengthen me to achieve such success, and my sister, Nimarta, for always being supportive and cheering me up .

Finally, I would like to extend my sincere thanks to all my friends, Anjali, Aaditya, Navid, and David for their support and help. Thanks for always being there for me and making this experience so fun and lively.

Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations, Symbols, and Nomenclature

Acc	Accuracy
ANN	Artificial Neural Network
CAE	Convolutional Autoencoder
CAE-HIFD	Convolutional Autoencoder framework for HIF Detection
CC	Cross-Correlation
CNN	Convolutional Neural Network
Dep	Dependability
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
FN	False Negatives
FP	False Positives
GB	Gabor Transform
GSCV	Grid Search Cross-Validation
GRU-AE	Gated Recurrent Units Autoencoder
HIF	High-Impedance Fault
K	Kurtosis
LSTM	Long Short Term Memory
ML	Machine Learning
MSE	Mean Squared Error
NLP	Natural Language Processing
PDF	Probability Density Functions
ReLU	Rectified Linear Unit
RMS	Root Mean Square
RNN	Recurrent Neural Network
Saf	Safety
Sen	sensibility
STFT	Short Time Fourier Transform
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
T-CNN	Transformer-CNN
TFA	Time-Frequency Analysis
TN	True Negatives
TNR	True Negative Rate
TP	True Positives
TPR	True Positive Rate
VMD	Variational Mode Decomposition
WT	Wavelet Transform

Chapter 1

Introduction

1.1 Motivation

High-Impedance Faults (HIFs) typically occur when a live electrical conductor unintentionally comes in contact with a highly resistive surface, such as trees branches, sand, and concrete [1–4]. The undetected HIFs pose a potential threat to personnel, environment, and grid stability [5]. As shown in Fig. 1.1, the HIF is often followed by the occurrence of arcing ignition. Thus, the downed conductor is a safety hazard, as the fire caused by arcing can cause great loss to life and property [4, 6]. In late 2017, the deadliest fires in Mendocino and Napa Counties killed 15 people, burned a total of approximately 88,000 acres, and destroyed nearly 1200 structures [7]. The California Department of Forestry and Fire Protection reported the cause of these incidents to be power lines that ignited fires when they came in contact with trees and branches [7]. Moreover, undetected HIFs have been reported to cause instability of renewable energy systems [8]. As a result, reliably detecting and clearing HIFs on time is crucial to ensure the safety of personnel and maintain the power system integrity [4, 6, 9]. Furthermore, classification of the type of HIFs assists in isolating the fault, which ensures timely removal of faults and reduces the number of power outages.

HIFs commonly have characteristics such as asymmetry of the current waveform, random-

ness, and nonlinearity of the voltage-current relationship. These characteristics are diverse and are affected by multiple factors, including surface types and humidity conditions [4]. As a result, the HIF current magnitude typically ranges from 0 to 75 A [1]. Such low current magnitudes compared to normal load current levels, together with a high diversity of characteristics and patterns, make HIFs difficult to detect. Diversity and randomness make it difficult to quantify and characterize the HIFs and, consequently, complex techniques, such as those based on machine learning, are required for HIF detection. Moreover, due to the low fault current magnitude, the conventional overcurrent relays fail to discriminate most HIFs from a load unbalance [1, 2, 9].

Figure 1.1: Arcing ignition caused by power lines touching tree branch.

As shown in Fig. 1.2, the HIFs are characterized as the distortion of the voltage and current signals. These distortions are so minute that they are hard to note on the current signal plots. As shown in Fig. 1.2 (b), they can be hardly seen even on the magnified plots of the voltage signals. As a result, it is hard to detect HIFs by directly using the voltage or current signals; thus, these signals are typically preprocessed by utilizing various signal processing or domain transformation techniques.

Numerous protection schemes have been proposed to detect HIFs that utilizes signal pro-
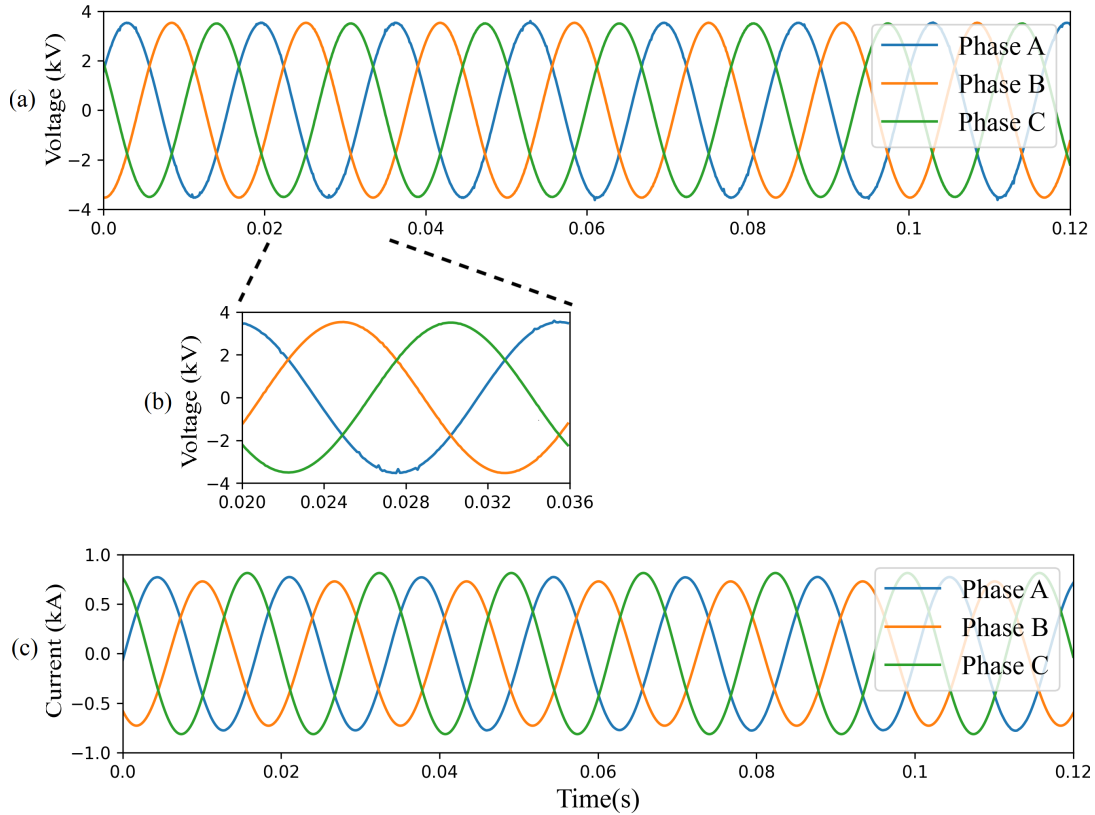
Figure 1.2: The phase-A type HIF: (**a**) voltage signals, (**b**) magnified distortions, and (**c**) current signals.

cessing techniques, such as, harmonic transformations [10–16], Discrete Wavelet Transformations (DWTs) [5, 17–21], Kalman Filtering [22], Empirical Mode Decomposition (EMD) [23], Variational Mode Decomposition (VMD) [6, 9], fractal theorem [24], and energy variance criterion [25]. Recently, Machine Learning (ML) has been gaining popularity in the field of HIF detection. The combination of signal processing and ML-based algorithm, such as, Artificial Neural Network (ANN) [26–29], Support Vector Machine (SVM) [30], decision tree [31], Recurrent Neural Network (RNN) [32], and Convolutional Neural Network (CNN) [33] have been proposed in the literature to detect the HIF scenarios.

The signal processing-based protection strategies have achieved high accuracy in detecting the HIFs. However, the signal processing used for preprocessing of the data is resource-intensive, and the HIF detection techniques based on the signal processing fail to reliably detect HIFs in presence of diverse non-fault disturbances and nonlinear loads. The widely utilized

harmonic-based approaches may falsely detect non-HIF disturbances (e.g., capacitor and load switching), transformer energization, and may falsely classify nonlinear loads as HIFs. Furthermore, the DWT-based techniques are vulnerable to noise and harmonic variations in the system. Most of the signal processing-based HIF detection approaches neglect the effects of noise [19, 21, 27, 30] and non-fault disturbances, such as capacitor switching and load switching in their evaluation studies [17, 22, 28].

Although the ML-based HIF detection techniques have high pattern recognition capability, these studies apply a supervised approach to learn the mapping from the input to the output based on a limited set of HIF and non-HIF scenarios present in the training set. Here, the non-HIF scenario refers to the normal steady-state operation and any transient response to non-HIF disturbance such as capacitor switching and load variations. A supervised learning system may fail to reliably identify any fault scenario or any non-fault disturbance, which is not present in the training set. However, there is a wide range of non-fault operating conditions and it is difficult to include them all in the training set. Therefore, a different way of training ML models is required that does not require a variety of non-fault scenarios. Furthermore, there are very few studies capable of detecting the fault type, as discriminating phases in a HIF is a challenging task [4]. Thus, there is a need for a protection technique that can reliably detect on which phase the fault occurred.

Consequently, this thesis proposes HIF detection and classification techniques that reliably detect HIFs and the fault type regardless of the type of surface involved and the fault distance. For the HIF detection and classification, two different deep neural network architectures are designed without making use of resource-intensive signal processing techniques. The Convolutional Autoencoder (CAE) framework has been developed for HIF Detection (CAE-HIFD), which utilizes an unsupervised approach to learn solely from the fault data, thus avoiding the need to take into account all possible non-fault scenarios in the learning stage. Furthermore, for the HIF classification, we have developed the transformer network-based framework that employs supervised learning to reliably detect the HIFs and the fault type [34].

1.2   Contributions

The primary objective of this thesis is to leverage deep learning-based algorithms to design and develop reliable and accurate HIF detection and classification frameworks. The main contributions of this thesis are as follows:

- Proposing the Convolutional Autoencoder (CAE) framework for HIF Detection (CAE-HIFD), which learns solely from the fault data by utilizing unsupervised learning. The proposed approach does not need to consider diverse non-fault scenarios in the learning stage. The ability of the CAE to model the relationship between the data points constituting a signal enables the CAE-HIFD to learn complex HIF patterns. Furthermore, the CAE discriminates non-HIF steady-state conditions from HIFs by identifying deviations from the learned HIF patterns using Cross-Correlation (CC).

- Proposing the probability distribution-based kurtosis analysis for achieving security against falsely treating non-HIF disturbances, such as capacitor and load switching, and non-linear loads as HIFs. Non-HIF disturbances can cause voltage and current distortions similar to the HIFs, making HIF detection even more difficult. Consequently, the kurtosis analysis is utilized in the proposed CAE-HID and T-CNN frameworks to detect these disturbances.

- Proposing the HIF detection and classification approach based on the state-of-the-art deep neural network architecture, the transformer network, that reliably detect fault type. The transformer network is stacked with a convolutional layer to provide robustness against noise.

- Evaluating the performance of the proposed CAE-HIFD through comprehensive studies conducted on the IEEE 13-node test feeder taking into account various HIF conditions involving seven different fault surfaces and diverse non-HIF scenarios.

- Evaluating the performance of the proposed Transformer network-based HIF detection

and classification approach. The training and testing are done on the data comprising of diverse non-HIF scenarios and HIFs occurring at the three different phases. The data was obtained from the IEEE 13-node test feeder that considered HIF conditions at seven different fault surfaces.

- Conducting case studies to measure the response of the proposed CAE-HIFD on HIF scenarios, intermittent HIFs, frequency deviations, and non-HIF disturbances such as capacitor switching, nonlinear load, and transformer energization.

- Conducting case studies to examine response of the proposed transformer-based HIF detection and classification approach on HIF scenario and non-HIF disturbances such as capacitor switching, nonlinear load, and transformer energization.

- Demonstrating the robustness of the proposed CAE-HIFD and transformer-based approach against various levels of noise. Additionally, the performance of the transformer-based approach has been examined with the input data having sampling frequencies ranging from 1.25 kHz to 10 kHz.

The proposed CAE-HIFD and T-CNN frameworks achieved 100% accuracy and outperformed the existing state-of-the-art techniques for HIF detection and classification. Furthermore, the proposed approaches exhibit robustness against noise and perform well under all the challenging case studies.

1.2.1   Thesis Outline

The remainder of this thesis is organized as follows: Chapter 2 describes the background, which includes Autoencoders in Section 2.1, Convolutional Neural Network in Section 2.2, and Transformer Network in Section 2.3.

Chapter 3 discusses the related works. First, Section 3.1 discusses signal processing techniques for HIF detection and classification, and then Section 3.2 introduces machine learning techniques for HIF detection and classification.

Chapter 4 presents the methodology applied in the Convolutional Autoencoder for HIF Detection, including data preprocessing, offline training, and HIF detection.

Chapter 5 describes the methodology applied in the Transformer-CNN Framework for HIF Classification, including data preprocessing, offline training, and HIF classification.

Chapter 6 discusses the results and analyzes the findings: Section 5.1 presents the study system. Section 5.2 presents evaluation conducted for the proposed CAE-HIFD, including CAE-HIFD Model training, effects of CAE-HIFD's components, CAE-HIFD response to different case studies, comparison with other approaches, the robustness of the proposed CAE-HIF against noise, and discussion. Section 5.2 describes evaluation conducted for the transformer-CNN, which consists of T-CNN Model training, the effect of kurtosis and differencing, system response for representative case studies, the effect of the sampling frequency, and robustness of the proposed T-CNN against noise.

Finally, Chapter 7 concludes the thesis and discusses future work.

Chapter 2

Background

This section provides an overview of the autoencoders, the convolutional neural networks, and the transformer network in terms of structure and functions.

2.1   Autoencoders

A supervised machine learning model learns the mapping function from the input to output (label) based on example input-output pairs provided in the training dataset. In contrast, an unsupervised learning model discovers patterns and learns from the input data on its own, without the need for labeled responses, which makes this approach a go-to solution when labels are not available. Autoencoders are trained through unsupervised learning where the model learns data encoding by reconstructing the input [35]. They are commonly used for dimensionality reduction [35, 36], but the non-linear feature learning capability has made them also successful in denoising and anomaly detection [37–39].

As shown in Fig. 2.1, a conventional autoencoder is a feed-forward neural network consisting of an input, an output, and one or more hidden layers. The encoder part of the autoencoder reduces dimensionality. The input x of dimension f is multiplied by the weights W and, together with the bias b, is passed through the activation function σ to produce representation z of dimension m, $m < f$ [35], as follows.

$$z = \sigma(Wx + b) \tag{2.1}$$

Next, the decoder attempts to reconstruct the input x from the encoded value z. The product of weights W' and z is added to biases c, and the resultant is utilized as an input to the activation function σ to generate the reconstructed signal y as follows:

$$y = \sigma(W'z + c) \tag{2.2}$$

Over a number of iterations (epochs), the autoencoder optimizes the weights and biases by minimizing an objective function such as Mean Squared Error (MSE) [35]:

$$MSE = \frac{1}{f} \sum_{i=1}^{f} (x_i - y_i)^2 \tag{2.3}$$
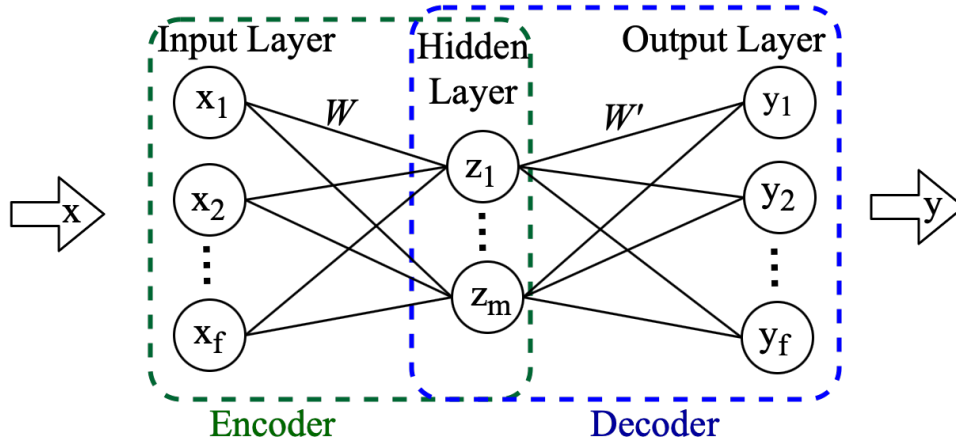
Figure 2.1: Conventional Autoencoder structure

where $x_1, x_2, x_3, ..., x_f$ is the input vector, $y_1, y_2, y_3, ..., y_f$ is the output vector produced by the autoencoder, and f is the number of features as per Fig 2.1. The described feed-forward autoencoder ignores the spatial structure present in data and, to address this issue, the CAE was introduced [40]. A CAE replaces the fully connected layers of the feed-forward autoencoder with convolutional and deconvolutional layers.

2.2   Convolutional Neural Network

Convolutional Neural Networks (CNNs) are the type of deep neural networks for learning patterns from the data that have grid-like topology. As depicted from the name, the CNNs apply convolution operation on local spatial regions of the input. The CNNs have been tremendously successful with image data, which can be perceived as 3D matrix with RGB grid of pixels [41]. Consequently, the CNNs have been employed to accomplish classification, segmentation, and detection tasks in radiology, detect tampering in the images, and conduct medical image analysis [42–44]. The one-dimensional (1D) convolutions have led to significant advancements in sensor data processing [45–47]. For example, the 1D-CNN has been successful in achieving high accuracy for applications, such as motor fault detection [45], photovoltaic power forecasting [46], and human activity recognition [47].
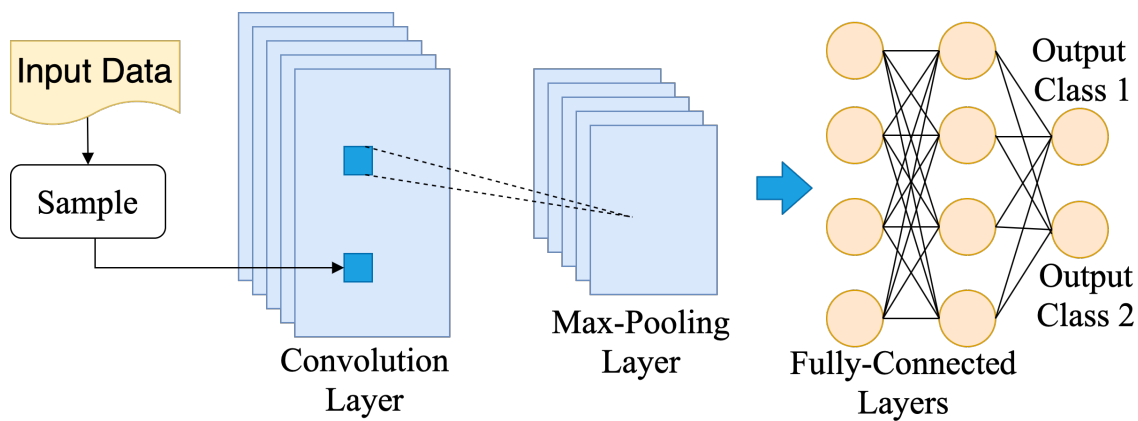
Figure 2.2: Convolutional Neural Network

In this thesis, the 1D-CNN is utilized for HIF detection and classification. A shown in Fig. 2.2, a typical CNN has layers, such as input, convolution, max-pooling, fully connected, and output layer. The convolution layer consists of kernels, also known as filters, which move across the timesteps and perform element-wise multiplication with the input matrix to generate an activation map. Next, the pooling layer (e.g., max-pooling) performs down-sampling to reduce dimensionality and associated computational complexity. The output dimension of the max-pooling layer is flattened to connect with all the nodes of the fully connected layers. For the classification task, the output classes are generated by the output layer. Next, loss

function, such as categorical crossentropy calculates the error between predicted and actual labels. Finally, backpropagation and gradient descent are applied to the CNN to update the kernels and weights in the convolution and fully connected layers.

2.3 Transformer Network

Recurrent Neural Networks (RNNs) are well suited for sequential data as they are capable of capturing temporal dependencies. However, their sequential nature restricts them from processing the whole input sequence at once resulting in a long training time. To remedy this issue, in natural language processing, the transformer network has been recently introduced: this network enables high parallelism by processing the whole input sequence at once [34].

As shown in Fig. 2.3, the transformer network comprises of the encoder and decoder [34]. The embedding and positional encoding block in Fig. 2.3 signify that the text data input in the Natural Language Processing (NLP) is transformed into a numerical vector before feeding it to the encoder. Furthermore, the positional embedding is added to the input embedding in order to introduce the order of the sequence information in the temporal data.

The encoder takes as the input vector sentence embeddings together with positional encoding. The multi-head self-attention block comprised of parallel placed fully connected neural network layers simultaneously processes different portions of the input vector to learn associations within vector time steps [34]. Next, the normalization and the feed-forward layers generate an abstract representation of the complex input patterns, which is passed to the decoder.

Similar to the encoder, the decoder has the multi-head self-attention, the feed-forward, and normalization layers. Additionally, the decoder inserts a masked multi-head attention layer that processes right-shifted output embeddings. The masking of future values ensures that the prediction for the current position only depends on the outputs of the previous positions. The masking of the future values preserves the auto-regressive properties of the transformer network and allows it to employ a teacher-forcing learning procedure. As a result, the trans-

Output Probabilities
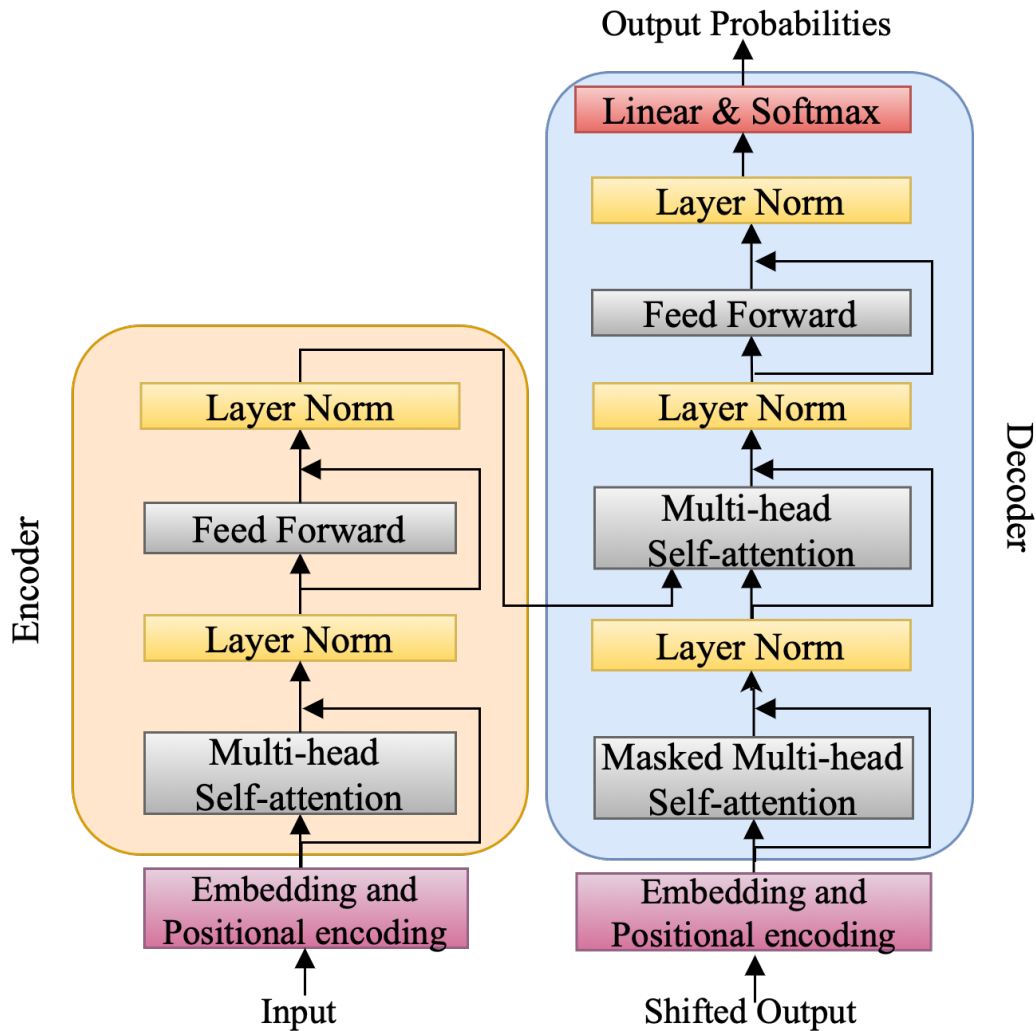
Figure 2.3: Transformer Network

formers outperform existing approaches for the applications, such as language translation [34].

Finally, linear transformation and the softmax function generate the output in the form of prob-

abilities.

Several recent works in the field of NLP apply transformer networks to achieve better re-

sults than widely used RNNs for applications such as machine translation, text summarization,

and text generation [34]. Moreover, the temporal data handling attribute of the transformer

network encoder has shown good initial results in time series classification, time-series fore-

casting, and music generation [48–50]. Consequently, the complex pattern learning capability

of the transformer network is utilized in this thesis to detect the HIFs and identify the fault

type.

Chapter 3

Related Work

This chapter first reviews various signal processing-based techniques for HIF detection. Next, different machine learning algorithms for HIF detection and classification are discussed.

3.1   Signal Processing Techniques for HIF detection and classification

Reliably detecting HIFs in a distribution network remains a challenge due to their random, as well as dynamic nature. Various HIF detection techniques have been proposed in the literature that exploits different characteristics of the HIFs. A significant number of HIF detection methods apply signal processing techniques on the current and voltage signals observed at the relay substation, and this subsection reviews those techniques.

Arcing caused by the HIFs results in the formation of low and high-frequency components in the current spectrum [4]. Various approaches take advantage of the frequency components to detect the HIFs [10–16]. The early works used the third and fifth harmonics magnitude [10, 11], even-order harmonics [12], and inter-harmonics [13] to detect HIF in the distribution network. Emanuel *et al.* [14] compared the low-frequency spectrum with current harmonics to understand the importance of 120 Hz and 180 Hz components in HIF detection. Their experimental studies confirmed that the second harmonic can be applied to detect the HIFs caused

by the low amplitude arcs on the sandy surface. Jeerings *et al.* [15, 16] relied on a change in the magnitude and phase angle of the third harmonic current to detect high impedance ground faults. In general, Harmonic-based HIF detection techniques are not sufficiently sensitive to detect remote HIFs [1]. Moreover, the harmonic-based approaches may falsely detect transient events, such as capacitor switching, transformer energization, and the presence of nonlinear loads as HIFs [1].

An HIF detection scheme based on a power line communication system is proposed by Milioudis *et al.* [51]. This method discriminates the HIFs from normal operating conditions by observing deviations in the input impedance of the network at particular frequencies. Although the power line communication system-based method is successful in detecting and locating HIFs, it requires costly communication systems and is not suitable for large and complex networks [52].

The Discrete Wavelet Transformation (DWT)-based HIF detection techniques [20] operate by examining the measured signals in both the time and frequency domains. The DWT has been widely utilized in literature to detect HIFs [5, 17–21]. In these methods, the mother wavelets, such as Daubechies [5, 17–19, 21] and Rbior [20] are used to decompose the current or voltage signals into approximation and detail coefficients. Next, the HIFs are discriminated from the non-HIF scenarios by analyzing these extracted coefficients. Vulnerability to noise and harmonics are the main shortcomings of the DWT-based approaches [53] with their accuracy being highly dependent on the choice and order of the mother wavelet [4].

The Time-Frequency Analysis (TFA) has also been proposed for detecting the HIFs, and techniques based on this typically consider the energy of the input signal at all time and frequency coordinates [4]. Lima *et al.* [54] used Short Time Fourier Transform (STFT) to extract lower order harmonics, which assisted in analyzing the HIF behavior in time and frequency domains. In this technique, a time delay of five cycles is required to discriminate HIFs from distribution network disturbances, such as capacitor or load switching. Cheng *et al.* [55] applied TFA by combining the Gabor Transform (GT) and Wigner distribution to detect HIFs and

distinguish them from non-stationary signals. In general, TFA algorithms are highly sensitive to discontinuities and require extensive computational resources [53].

The presence of randomness and flicker in currents signals can be the result of HIFs in distribution systems. Mamishev *et al.* [24] applied Fractal theorem on the Root Mean Square (RMS) values of the current signals to measure the degree of chaos caused by the HIFs. Lien *et al.* [25] utilized a three-phase unbalance current as an input signal to an HIF detector that estimated randomness by calculating the energy variance for the second, fourth, and sixth harmonics of the input signal. Typically, randomness-based HIF detection approaches achieve high detection rate [24, 25]. However, the HIF current is too low to cause any significant changes to the measured current signals at the substation [56], especially under the unbalanced load conditions. Hence, these techniques fail to provide dependable protection against HIFs in distribution systems.

HIF detection by Kalman filtering [22] is based on tracking randomness in the current waveform. In this approach, time variations in the fundamental and harmonic components have been extracted by applying the Kalman filter to detect the HIF scenarios. Because the small HIF current does not cause significant variations in the current waveforms seen by the relay at the substation [56], the Kalman filtering-based technique exhibits low sensitivity.

The presence of randomness and disturbance in the voltages or current signals may be an indication of HIFs; thus, a mathematical morphology has been employed [2]. However, when HIF occurs, the fundamental frequency components dominate the HIF signatures and thus, this technique causes a significant time delay in HIF detection. Moreover, under remote HIFs or the presence of nonlinear loads, this strategy fails to provide a reliable HIF detection [53].

The HIF can also be detected by analyzing temporal inconsistencies in the time domain. Carr *et al.* [57] developed proportional relay based on a time-domain analysis to detect the HIFs. In the study by Sarlak *et al.* [58], time-domain features are extracted by measuring the magnetic-field strength signal near the conductor of the feeder, and these features are utilized as an input by a pattern recognition algorithm. Although methods based on time-domain analysis

exhibits low computational complexity, they invalidate HIF detection under unbalanced load conditions and switching events scenarios [16].

Empirical Mode Decomposition (EMD) [23] and Variational Mode Decomposition (VMD) [6, 9] have been recently proposed for HIF detection. These techniques operate by extracting features from the signals. However, the EMD-based methods suffer from modal mixing and are sensitive to noise [59]. On the other hand, the VMD-based technique is more robust against noise; nevertheless, it requires a large number of decomposition modes as well as a high sampling frequency to perform well, which results in a large computational burden. Moreover, the VMD technique [9] for HIFs detection cannot reliably detect HIFs when a large nonlinear load is in operation.

In contrast to the reviewed works [2, 5, 6, 9–25, 51, 54, 55, 57, 58], preprocessing in the proposed CAE-HIFD and T-CNN frameworks does not require resource-intensive signal processing. In addition, approaches that operate on the fundamental frequency components risk failures in the presence of frequency deviations [10–16]. However, CAE-HIFD does not operate based on the fundamental frequency components of the input signals and, consequently, is not sensitive to frequency deviations as examined in the evaluation section. Furthermore, the proposed techniques are robust against non-fault disturbances, such as capacitor or load switching, nonlinear loads, and inrush currents.

3.2 Machine Learning Techniques for HIF detection and classification

Machine Learning (ML) algorithms can learn patterns in the data and thus, in recent years, ML-based approaches have been gaining popularity in many domains, including HIF detection.

Artificial Neural Networks (ANNs) are the most prevalent algorithms utilized in detecting and classifying the HIFs [26–28, 60, 61]. They are inspired by the biological neural networks in the human brain and can achieve high accuracy in classifying the faults [27, 35]. Baqui

et al. [27] combined an ANN with Discrete Wavelet Transform (DWT) for HIF detection in medium-voltage networks. They applied DWTs to extract the relevant features from the current waveforms and then used the extracted features as an input to the ANN classifier. In a study by Michalik *et al.* [26], the high-frequency components of zero-sequence voltage and current signals are suppressed by using Butterworth filter, and ANN is trained over this preprocessed signal. Routray *et al.* [61] applied Stockwell Transform to extract time and frequency-based features from the current and voltage signals. The ANN classifier is utilized to identify the fault and non-fault scenarios from the extracted features. They reported average accuracy of 96.5% under 50 dB noise.

Tonelli-Neto *et al.* [28] applied Wavelet Transform (WT) on the current signals to conduct multi-resolution analysis and then, utilized the energy concept to extract features. They used multiple fuzzy-ARTMAP neural networks and multiple fuzzy inference systems along with the extracted features to detect and classify HIFs. Moravej *et al.* [29] took advantage of dual-tree complex WTs to extract the features that were used as input by probabilistic neural networks to discriminate HIF cases from non-HIFs. Veerasamy *et al.* [62] used classifiers such as adaptive neuro-fuzzy inference system and multi-layer perceptron along with DWT-based feature extraction to detect HIFs and to discriminate them from other transients in medium-voltage distribution systems. Although the neural networks achieve acceptable accuracy in detecting the HIFs, they are sensitive to frequency changes and noise [4].

The Support Vector Machine (SVM) is another commonly used classifier that can categorize data by mapping it to higher dimensions [30]. Ghaderi *et al.* utilized TFA to determine the energy of low-frequency component features that are used as an input for the SVM classifier to separate the HIF from the non-HIF scenarios [30]. Although this method could only achieve 93.6% detection accuracy, it eliminated the need for modification of the technique in case of new observed HIF samples by preserving the information during the feature extraction. In another study, an SVM-based model is used with features extracted by VMD for HIF detection in distribution lines incorporating distributed generators [6]. This technique prevents false

detection of switching events as HIFs; however, it is sensitive to noise for SNRs below 30 dB. Moreover, the transformations associated with the VMD require high computational capability [4].

The decision tree algorithm with the fast Fourier transform is also an alternative HIF detection approach, which utilizes the HIF features, such as the phase of the third harmonic, magnitude of the second, third, and fifth harmonics, and RMS value of current signals [31]. Although this technique is successful in discriminating HIFs from non-fault disturbances, it is sensitive to the presence of noise.

Recurrent Neural Network (RNN) architecture is well suited for sequential data [63]; therefore, its variant, Long Short Term Memory (LSTM) was used with features obtained by DWT analysis to detect the HIFs in the solar photovoltaic integrated power system [32]. However, this method detects HIFs with a success rate of only 92.42% and, moreover, this study uses DWT-based features which are sensitive to noise [64].

Another study relies on modified Gabor WT for extracting the Two Dimensional (2-D) scalograms from the input signal [33]. The extracted 2-D features are used as an input to the Convolutional Neural Network (CNN) classifier for detecting the HIFs [33]. This method was accurate in the presented experiments; however, as it is a supervised learning method, it may fail to reliably detect HIF and non-HIF scenarios that are not present in the training data.

The huge success of deep learning algorithms has motivated this thesis to explore and leverage deep learning to detect and classify HIFs with promising accuracy. The reviewed works [6, 26–28, 28–33, 60–62] rely on supervised learning, which utilizes a limited set of HIF and non-HIF scenarios to learn the patterns in the data. As there could be a wide range of HIF and non-HIF scenarios, the supervised learning-based method may fail to reliably detect any new HIF or non-HIF scenarios. In contrast, the proposed CAE-HIFD utilizes an unsupervised approach that learns solely from the fault data, which avoids taking into account diverse non-HIF scenarios. Furthermore, as the HIFs do not cause significant changes in the voltage or current signals, very few studies are able to detect the fault type [4]. However, the proposed T-

CNN can reliably perform the HIF detection and classification. Many studies neglect the effects of noise [19, 21, 27, 30] and omit to examine non-fault disturbances [17, 22, 28]. However, the robustness of the proposed CAE-HIFD and T-CNN against noise has been examined in this thesis.

Chapter 4

High-Impedance Fault Detection

This chapter aims to provide a detailed explanation of the proposed deep learning-based framework for the High-Impedance Fault (HIF) detection. Subsections within the chapter discuss different components of the proposed framework, such as data preprocessing, offline training, and HIF detection.

4.1 Convolutional Autoencoder for HIF Detection

Traditionally, for fault detection, autoencoders are trained with normal data and then used to detect abnormal operating conditions by identifying deviations from the learned normal data [38]. On the contrary, in the proposed CAE-HIFD, the CAE is trained using fault data only and recognizes non-fault operating condition by detecting deviations from the learned fault scenarios. The spatial feature learning and generalization capability of the CAE assist the CAE-HIFD to detect new HIF scenarios that are not present in the training set. Furthermore, since the CAE is only trained on the fault data, any non-fault cases will not be identified as HIF, which increases the security of the proposed protection strategy.

As depicted in Fig. 4.1, the CAE-HIFD is comprised of offline training and online HIF detection. The analog three-phase voltage and current signals are sampled and converted to digital signals using A/D converters. Training happens in the offline mode using a dataset

prepared from multivariate time series consisting of three-phase voltage and current signals. The online HIF detection uses the weights and thresholds obtained from the offline training. As the data preprocessing step is the same for both training and detection, the preprocessing is described first, followed by the explanation of training and detection.
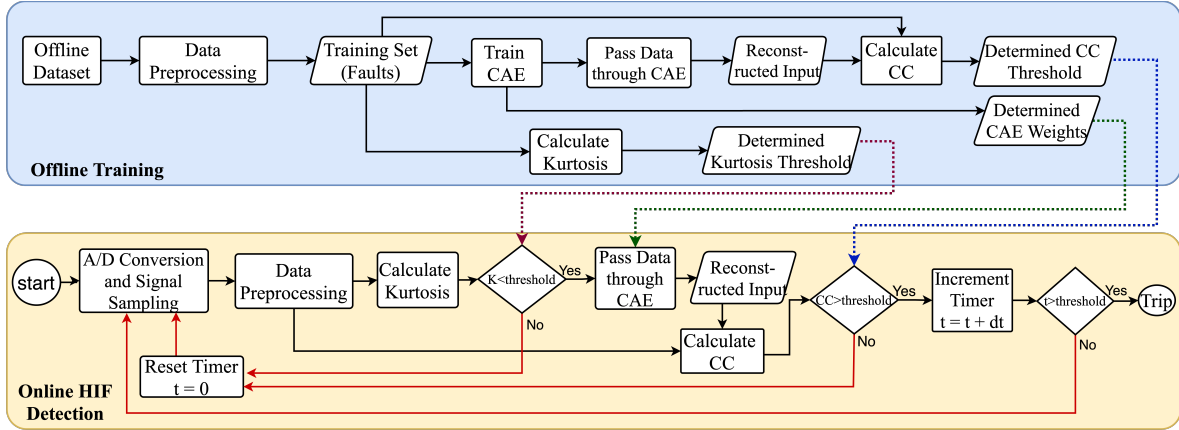
Figure 4.1: CAE Framework for HIF Detection (CAE-HIFD

4.1.1   Data Preprocessing

The first step of data preprocessing entails the sliding window approach applied to time-series data to transform it into a representation suitable for the CAE. As illustrated in Fig. 4.2, the first n samples (readings) make the first data window; thus, the data window dimension is n time steps $\times$ f number of features. In each iteration of the algorithm, the data window slides for s time steps, where s is referred to as a *stride*, to create the second data window and so on. Note that Fig. 4.2 illustrates a case where $s = n$.

Each voltage and current phase signal (each feature) in the sliding window is processed individually by differencing. The first-order d^1 and second-order d^2 differencing of signal $y(t)$, e.g., phase A voltage, are as follows:

$$d^1(t) = y(t) - y(t - 1) \tag{4.1}$$

$$d^2(t) = d^1(t) - d^1(t - 1) \tag{4.2}$$

The HIF causes small distortions in the voltage and current waveforms. The second-order differencing helps the CAE to learn and detect the HIF pattern by amplifying these distortions and suppressing the fundamental frequency component of each input signal. Differencing also amplifies noise; nevertheless, the generalization and spatial feature extraction capabilities of the CAE make the CAE-HIFD robust against noise as demonstrated in Section 6.2.4.
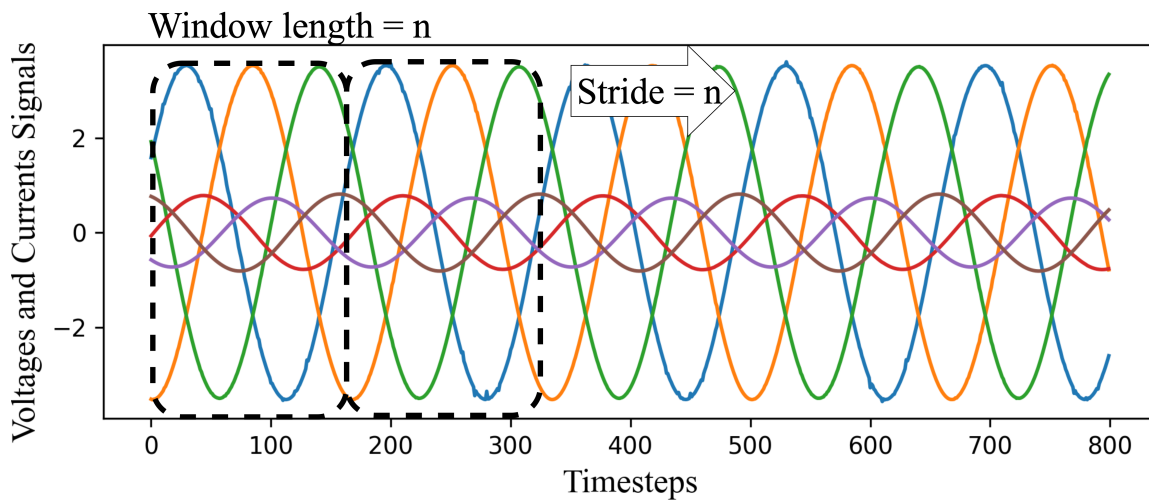
Figure 4.2: Sliding window approach

4.1.2 Offline Training

The CAE is trained solely with the fault data, and the non-fault data are only utilized for the system validation. As illustrated in Fig. 4.1, the preprocessed data are passed to the CAE as one data window, $n \times f$ matrix, at a time. As shown in Fig. 4.3, the CAE is composed of two main components: the encoder and decoder [35].

The first layer in the encoder performs the 1D convolution operation on the $n \times f$ input matrix with the kernel of size $k \times f$. This kernel moves across the time steps of the input and interacts with k time steps (here $k < n$) of the input window at a time; thus, during the CAE training, the kernel learns the local spatial correlations in the input samples. There
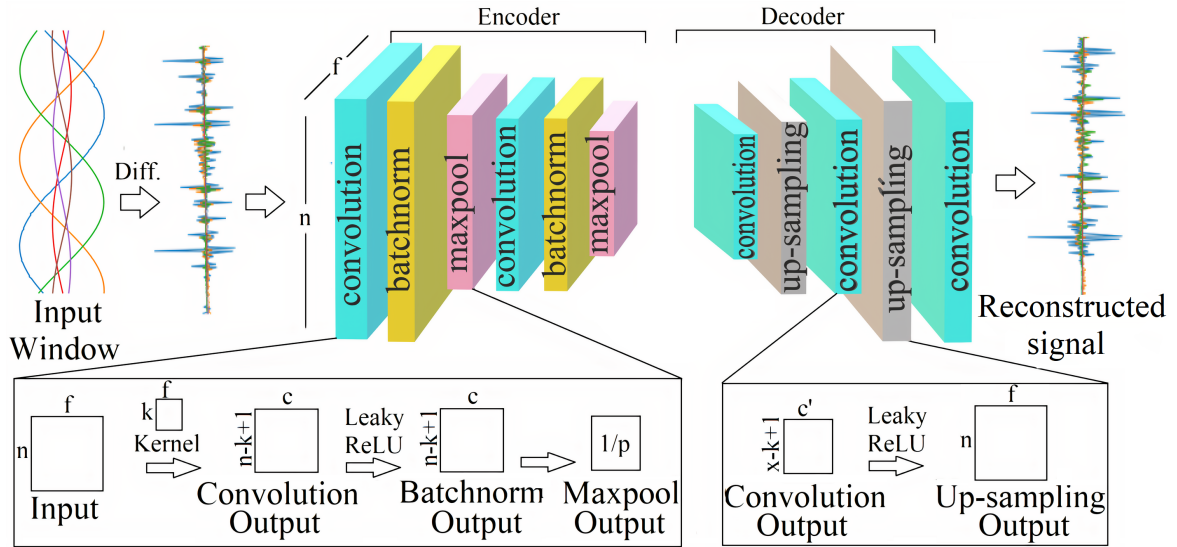
Figure 4.3: Convolutional Autoencoder (CAE) structure

are m kernels in the first layer and each kernel convolves with input to generate an activation map. Consequently, the output of the first layer has a dimension of $(n - k + 1) \times m$, and every column of this output matrix corresponds to the weights of one kernel. These kernel weights are learned during the CAE training process. Rectified Linear Unit (ReLU) activation function is often used to introduce non-linearity after the convolution. However, here LeakyReLU, a leaky version of ReLU, is used instead because ReLU discards the negative values in the sinusoidal wave [35]. Next, the batch normalization layer re-scales and re-centers data before passing them to the next layer in order to improve the training convergence. The batch normalized data are passed to the max-pooling layer to reduce data dimensionality and the associated computational complexity. The size of the max-pooling operation is p; therefore, the output of the pooling layer is $\frac{1}{p}$ of the convolved input. As illustrated in Fig. 4.3, the convolution, batch normalization, and max-pooling layers are repeated two times to extract features on different levels of abstraction. These encoder layers create an encoded representation of the input signal which is passed to the decoder.

While the encoder decreases the dimensionality of the input, the decoder reconstructs the original signal from these encoded values. In the decoder, as illustrated in Fig. 4.3, the convolu-

tional layer first generates the activation map, and then the up-sampling operations increase the dimensionality of the down-sampled feature map to the input vector size. During up-sampling, the dimensionality of the input is scaled by repeating every value along the time steps in the signal with the scaling factor set according to the max-pooling layer size in the encoder. Similar to the encoder, in the decoder, the convolutional and up-sampling layers are repeated twice (Fig. 4.3).

The CAE optimizes the weights and the biases using the backpropagation process in which the gradient descent is applied based on the loss function, typically MSE. In the proposed CAE-HIFD, the MSE is utilized as the loss function for training the CAE using fault data. In autoencoder, the MSE is also referred to as a reconstruction error as it evaluates the similarity between the input signal and the reconstructed signal given by the autoencoder output. As the objective of the gradient descent algorithm is to minimize the MSE for training data, the MSE is expected to be low for the training data and high for any deviations from the training patterns.

In the CAE-HIFD, the CAE sees only the fault data during training, and consequently, the trained CAE is expected to fail in reconstructing the non-fault data input. Therefore, the MSE for the non-fault data is expected to be higher than the MSE for the learned fault data. Traditionally, in autoencoders, the separation between fault and non-fault data is done based on a threshold which is determined using the MSE of the training dataset. However, in HIF detection, when CAE is trained with fault data, MSE is not a reliable metric for calculating the threshold. As illustrated in Fig. 4.4, the differentiated fault data forms a complex pattern with a high number of fluctuations causing the dissimilarities between the CAE output and input. The magnitude of these fluctuations varies from -2.0 to 1.0 and, as a result, even a small mismatch between input and CAE output leads to high MSE: for example, in Fig. 4.4 (a), MSE for fault data window is 0.0244. On the other hand, in Fig. 4.4 (b), MSE for steady state data window is 0.0002 because of a relatively simpler pattern compared to HIFs and small amplitudes of differentiated signal oscillations varying from -0.04 to 0.04. Consequently, the MSE is not a reliable indicator to discriminate between HIF and non-fault cases.

In signal processing, a metric commonly used to evaluate the similarity between signals is the cross-correlation (CC) [65] which is defined as:

$$CC = (f * g)(\tau) = \int_{-\infty}^{\infty} \overline{f(t)}g(t + \tau)dt \tag{4.3}$$

where f and g are two signals and τ is a time shift in the signal.
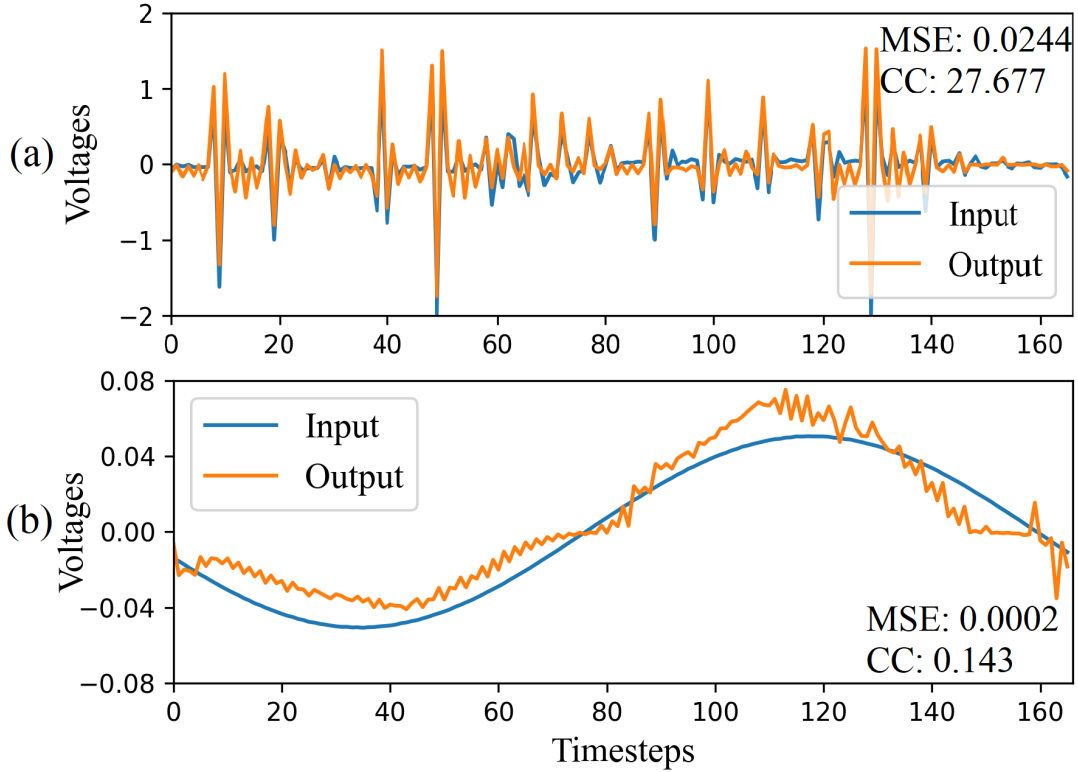
Figure 4.4: Input and output of the CAE for (a) HIF and (b) Non-HIF scenario.

In CAE-HIFD, CC is used to measure the similarity between the CAE input and output signals. As illustrated in Fig. 4.1, after the CAE training is completed, the trained CAE reconstructs all data windows from the training set and obtains reconstructed signals. Next, for each window in the training set, the CC value is calculated for the input signal and the corresponding CAE output. As seen in Fig. 4.4 (a), the HIF data window has a CC value of 27.677 because the input and output signals of the CAE are similar. On the contrary, a normal steady-state operating condition data window, Fig. 4.4 (b), has a low CC value of 0.143 as the

output deviates from the input. As the minimum CC value from the training set represents the least similar input-output pair from the training set, this minimum CC value serves as the CC threshold for separating HIF and non-HIF cases.

The CAE perceives the responses to disturbances such as capacitor and load switching to be HIFs because these disturbances cause waveform distortions. However, these disturbances usually occur for a shorter duration of time than HIFs making their statistical distribution different from those of HIFs and steady-state operation. Fig. 4.5 shows that the disturbances and HIFs both exhibit Gaussian behavior, but the disturbances have a thinner peak and flatter tails on the Probability Density Function (PDF) plot. In contrast, steady-state operation data (sinusoidal waveforms) have an arcsine distribution.

To distinguish disturbances from HIFs, the statistical metrics kurtosis is used. The kurtosis provides information about the tailedness of the distribution relative to the Gaussian distribution [65]. For univariate data $y_1, y_2, y_3, ..., y_n$ with standard deviation s and mean $\bar{y}$, the kurtosis is:

$$K = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^4/n}{s^4} \tag{4.4}$$

As Fig. 4.5 shows, flatter tails and thinner peaks results in higher kurtosis values. For example, the distribution of the differentiated capacitor switching disturbance in Fig. 4.5 (b) has a kurtosis value of $K = 76.6$ which is higher than the $K = 1.9$ for the HIF distribution in Fig. 4.5 (f).

The kurtosis is calculated from the training set individually for each data window after applying differencing. To prevent misinterpretation of the K values and avoid treating HIFs as non-fault disturbances, the kurtosis threshold must be higher than every K value present in the training set. Accordingly, the kurtosis threshold is the value below which all the K values of the training data lie.

The artifacts of the offline training are the CC threshold, the learned CAE weights, and the kurtosis threshold. These artifacts are used for online HIF detection.
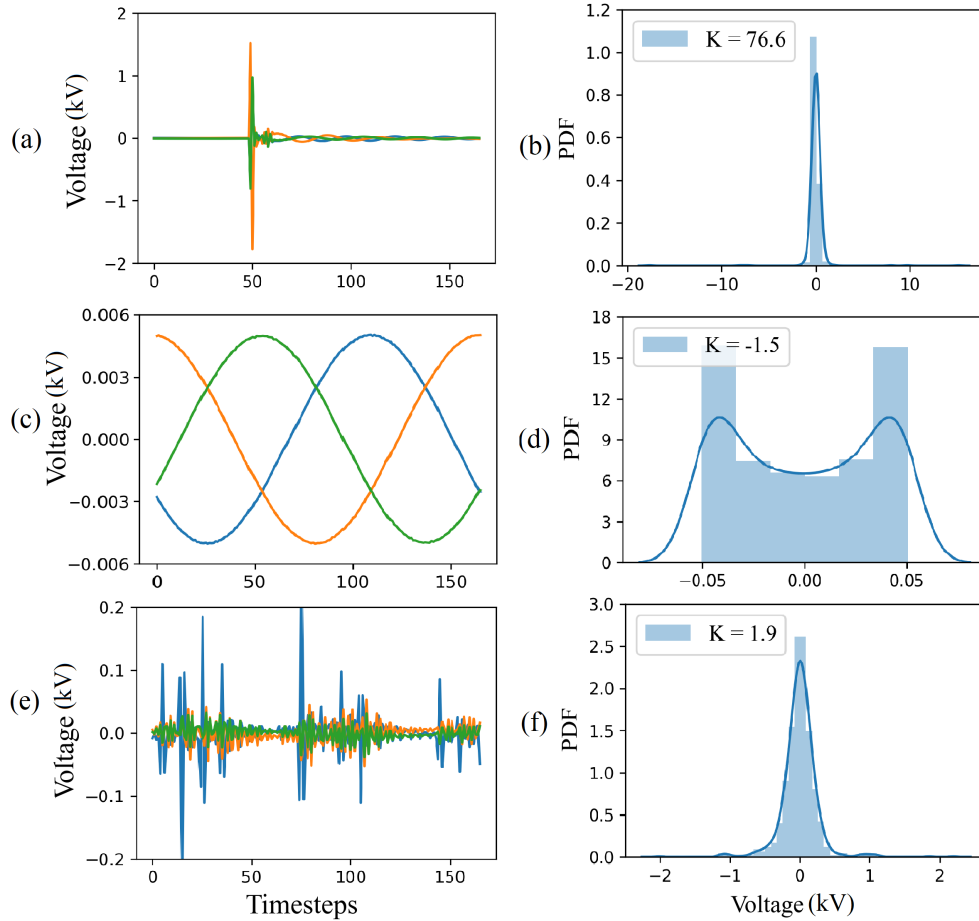
Figure 4.5: Kurtosis analysis: (**a**) differentiated voltage signal corresponding to a capacitor switching disturbance, (**b**) PDF of the voltage signal corresponding to a capacitor switching disturbance (**c**) differentiated normal steady-state voltage, (**d**) PDF of the normal steady-state voltage,(**e**) differentiated HIF voltage, and (**f**) PDF of the HIF voltage.

4.1.3 HIF Detection

The online HIF detection algorithm uses the artifacts generated by offline training as illustrated in Fig. 4.1. First, the analog input signal is converted to digital by the A/D converter and the data preprocessing module generates data windows which proceed through the remaining HIF detection components, one window at the time.

The value of kurtosis is calculated for each data window and compared with the corresponding threshold obtained from the offline training. Any data window with the kurtosis value above the threshold is identified as a non-fault disturbance case for which the CAE is

disabled because there is no need for additional processing as the signal is already deemed to be a disturbance. Next, the timer is reset for processing the next input signal segment.

If the kurtosis value is less than the threshold, the data window is sent to the trained CAE which encodes and reconstructs the signal. As the CAE is trained with fault data, for HIFs, the reconstructed signal is similar to the original signal. This similarity is evaluated by calculating the CC between the reconstructed signal and the original signal. If the CC value of the data window is greater than the CC threshold determined in the training process, the signal is identified to be corresponding to a HIF.

Under transient disturbances such as capacitor switching, the value of CC may exceed the corresponding threshold for a short time period immediately after the inception of disturbance. False identification of disturbances as HIFs is prevented using a pick-up timer. The timer is incremented when the CC exceeds its threshold and is reset to zero whenever the CC or K indicates a non-HIF condition, as shown in Fig. 4.1. A tripping (HIF detection) signal is issued when the timer indicates that the time duration of the HIF exceeds a predetermined threshold.

Chapter 5

High-Impedance Fault Classification

One of the objectives of this thesis is to detect the fault type; therefore, a deep learning-based framework for HIF detection and classification is proposed. Consequently, this chapter introduces the proposed framework and discusses its components, such as data preprocessing, model training, and HIF classification.

5.1   Transformer-CNN Framework for HIF classification

The transformer network outperforms sequence-based models in learning complex patterns from time-series data; however, it is not good at capturing the spatial structures present in the data. On the other hand, CNNs are capable of learning the multiple discriminative features in time-series data for classification tasks [45, 47]. Consequently, the Transformer-CNN (T-CNN) framework proposed here combines the temporal data handling capabilities of the transformer network and the spatial feature learning ability of the CNN for the HIF classification. As shown in Fig. 5.1, the T-CNN framework has two main stages: 1) model training and 2) online HIF classification. The training is conducted in an offline setting using a dataset prepared from multivariate time-series data. The online HIF classification utilizes the weights and threshold obtained from offline training. The data preprocessing is conducted for both training and classification; therefore, it is described first, followed by training and classification.
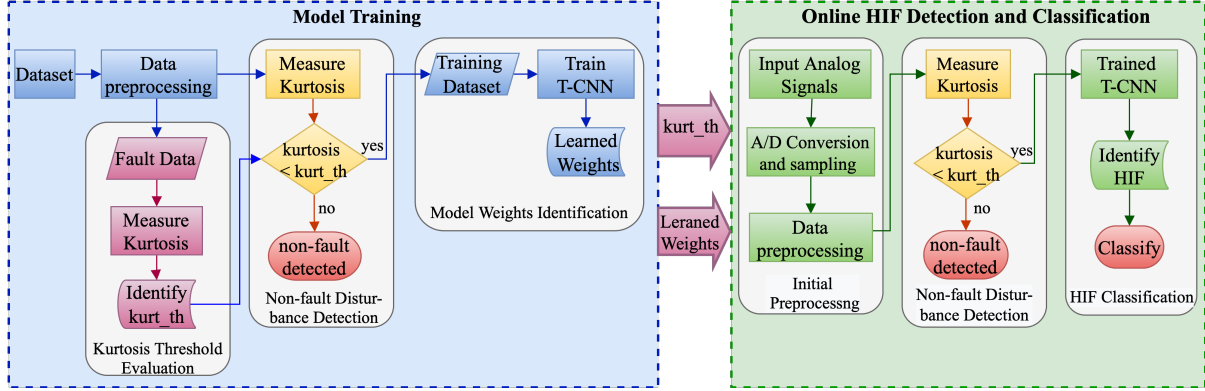
Figure 5.1: Transformer-CNN Framework for HIF Classification (T-CNN)

5.1.1   Data preprocessing

The digital signals obtained by sampling the analog voltage and current signals using Analog to Digital (A/D) converters comprise the multivariate time-series data utilized for the model training. The sliding window technique transforms this time series data into data windows of $w \times f$ size, where w is the number of time steps, and f is the number of features. The window slides for s time steps to form the next data window; s is referred to as a stride. The first order df_1 and second-order df_2 differencing is applied on every data window, where df_1 and df_2 are described as follow:

$$d^1(t) = y(t) - y(t-1) \tag{5.1}$$

$$d^2(t) = d^1(t) - d^1(t-1) \tag{5.2}$$

where $y(t)$ is signal at time t; for example phase A current. HIFs introduce distortions to the waveforms, and the second-order differencing amplifies these distortions and suppresses the fundamental frequency. Consequently, the resulting complex pattern is a mixture of temporal and amplitude variations. The differencing leads to amplification of noise; however, as demonstrated in experiments, the CNN layer in the proposed T-CNN enables generalization and makes T-CNN robust against noise.

5.1.2 Model Training

Non-HIF disturbances, such as the capacitor and load switching can cause signal distortions similar to the HIF data, and thus can result in false tripping. The T-CNN framework (Fig. 5.1) deals with these disturbances through kurtosis analysis. Non-fault disturbances occur for a short duration of time which makes their statistical distribution different from that of faults. The Probability Density Functions (PDF) from Fig. 4.5 in Chapter 4 illustrate that both non-fault disturbances and HIF data windows have leptokurtic distribution (i.e., normal distribution with more concentration around the mean) while the steady-state operation has arcsine distribution. As the kurtosis indicates the tailedness of the distribution relative to the Gaussian distribution, it is used to differentiate non-fault disturbances from the HIF and the steady-state operation. As mentioned in Chapter 4, for univariate data $y_1, y_2, y_3, ..., y_n$ with standard deviation s and mean $\bar{y}$, the kurtosis is:

$$kurtosis = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^4/n}{s^4} \tag{5.3}$$

The kurtosis value for each HIF data window in the training set is calculated individually, and the maximum value is selected as the kurtosis threshold. The data windows with the kurtosis above this threshold are deemed non HIFs and the remaining data windows comprise the training dataset (Fig. 5.1). As shown in Fig. 5.2, the proposed transformer-CNN model utilized in the T-CNN framework has multiple layers: data windows $w \times f$ are first passed through the transformer network encoder and then through CNN layers. Here, only the encoder part of the original transformer network [66] is used, as it is capable of generating meaningful representations for successfully HIF classification.

As shown in Fig. 5.2, similar to the transformer network, the encoder of the transformer-CNN model consists of multi-head attention, feed-forward, and normalization layers. The multi-head attention applies a self-attention mechanism to the input vector, which associates each time step in the data window (input vector) to other time steps in the input. An attention function is defined as mapping a query and set of key-value pairs to an output, where attention
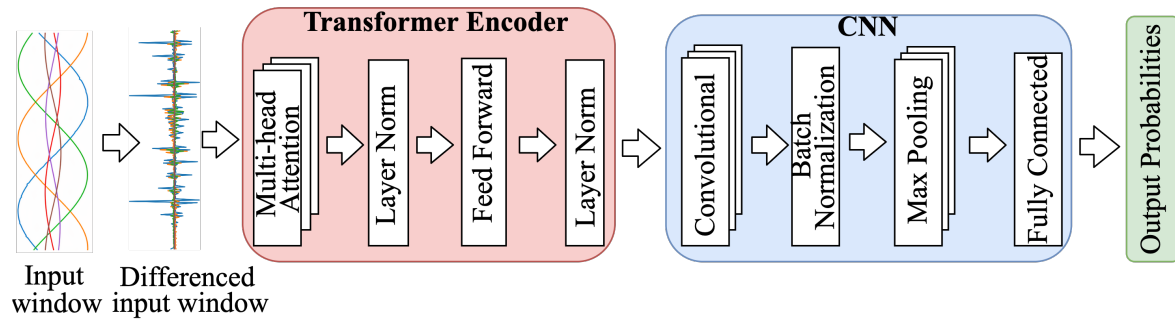
Figure 5.2: Transformer-CNN Model

weights are queried in the timesteps (keys) and output is the weighted sum of values [66]. To achieve self-attention in the encoder, the input vector is passed through fully connected layers to generate the query q, key k, and value v vectors [66]. The attention function utilized in the transformer network is described as follows:

$$Attention(q, k, v) = v \times softmax(\frac{qk^T}{\sqrt{d_k}}) \tag{5.4}$$

where *softmax* function produces class probabilities, and d_k is the dimension of k and q vectors. The dot product of q and k produces a score matrix that defines relative dependencies and associations within a data window. The scores are scaled down by a factor of $\sqrt{d_k}$ to prevent exploding gradient problem; where d_k is the dimension of keys and queries. Finally, the attention weights are computed by applying *softmax* operation on scaled-down scores, and an output vector is obtained by multiplying attention weights and v vector. In the multi-head attention mechanism, each head generates a separate part of the output vector $Attention(q, k, v)$, and each part learns something different, ultimately boosting performance. All the parts of output vectors from every head are concatenated and passed through a linear layer to generate an abstract encoder representation. The multi-head mechanism enables the parallelization, for example, the input with 6 features and *heads* = 3, each head can process $6/3 = 2$ features parallelly.

The output of the encoder is passed through the CNN, wherein, the 1D convolutional layer

extracts spatial features using the kernel ($k \times f$), which moves across the time steps of the encoder output. Furthermore, l kernels produce an activation map of dimension $(n - k + 1) \times l$ and columns of this map are weights of the l kernels. The Rectified Linear Unit (ReLU) activation function is typically applied to the output of the convolution layer to add non-linearity. Here, Leaky ReLU is used as the activation function instead of common ReLU because of its advantage in processing negative values [35]. Next, the output is re-centered and re-scaled in the batch normalization layer to reduce the time taken for the convergence during the training. The dimensionality of the batch normalized output is reduced in the pooling layer by a factor of $\frac{1}{pool-size}$, where $pool - size$ is max-pooling operation size. The max-pooling also assists in reducing the computational time complexity of the training process. Next, the fully connected layer calculates the final output probabilities by applying the *softmax* operation. Finally, the output of the proposed T-CNN framework is the class that has the highest probability: non-HIF, Phase *A*, Phase *B*, or Phase *C* fault.

The model is trained for *N* epochs by updating the weights to minimize negative *log-likelihood* loss between actual and predicted labels. The learned transformer-CNN weights and the kurtosis threshold are the outputs of the model training: they are used for the online HIF classification

5.1.3 HIF classification

The online HIF classification first transforms the analog signals into digital signals by using A/D converters as seen in Fig. 5.1. Next, these digital signals are preprocessed in the same way as for the model training: the sliding window technique is applied to generate data windows, and then the second-order differencing is applied on each individual data window.

The kurtosis value of each differentiated data window is calculated and compared with the kurtosis threshold obtained from the training process. If the kurtosis is greater than the threshold, the data window corresponds to non-HIF disturbance and is classified as non-HIF. Data windows with the kurtosis less than the threshold, the window proceeds to the trained

T-CNN model which discriminates the non-HIF cases from the HIF cases and identifies the fault type. Hence, the proposed T-CNN framework detects four classes: non-fault, phase A, B, and C fault.

Chapter 6

Evaluation

This chapter first describes the study system and the process of obtaining data for the performance verification studies. Next, evaluations of the proposed CAE-HIFD and the T-CNN frameworks are discussed.

6.1   Study System

The data set utilized for model training and evaluation is obtained through time-domain simulation studies performed in the PSCAD software. The study system is the IEEE 13 node test feeder of Fig. 6.1, a realistic 4.16 kV distribution system with a significant load unbalance. This test feeder was selected in order to examine the system behavior under challenging load unbalance conditions and because of its common use in HIF studies [67]. Detailed information regarding the line and the load data are provided in the appendix, and further information about this benchmark system can be found in [67]. For an accurate representation of the HIF behavior, the antiparallel diode model of Fig. 6.2, [2, 3, 6, 9, 68] is utilized. The HIF model parameters representing seven different faulted surface types are given in Table 6.1 [2, 69]. These parameters lead to effective fault impedances as high as 208 ohms in 4.16 kV distribution system.

In total, 210 faulty cases were simulated: 7 different surfaces, 10 fault locations, and 3
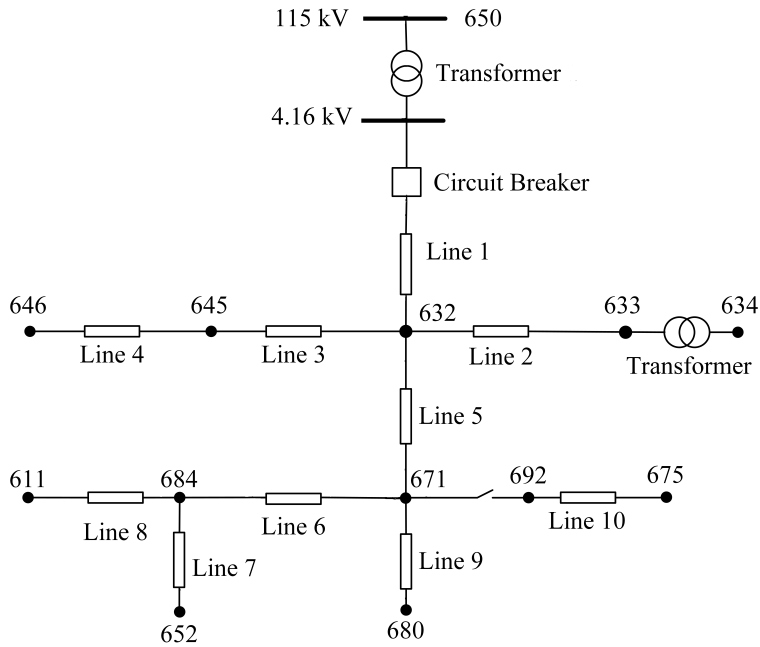
Figure 6.1: IEEE 13 node test feeder

phases. After the windowing technique, this resulted in 1372 HIF data windows. Additionally, the dataset obtained from simulations contained 272 non-fault data windows. Using data obtained from simulation studies enables considering diverse fault types, locations, and surfaces while obtaining such diverse data from real-world experiments would be difficult or even impossible. The 80% of the fault data is assigned for the model training and the rest for testing. As CAE-HIFD requires only fault data for training, all non-fault data is assigned for testing. The 10% of the training set is used as a validation set for the hyperparameter optimization.

Table 6.1: HIF Model Parameters For Different Surfaces.

Surfaces	R1 (Ω)	R2 (Ω)	V1 (V)	V2 (V)
Wet Sand	138±10%	138±10%	900±150	750±150
Tree Branch	125±20%	125±20%	1000±100	500±50
Dry Sod	98±10%	98±10%	1175±175	1000±175
Dry Grass	70±10%	70±10%	1400±200	1200±200
Wet Sod	43±10%	43±10%	1550±250	1300±250
Wet Grass	33±10%	33±10%	1750±350	1400±350
Rein. Concrete	23±10%	23±10%	2000±500	1500±500
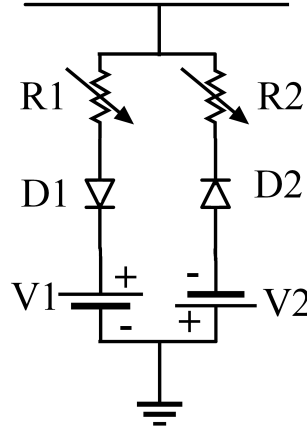
Figure 6.2: Nonlinear HIF model utilized for time-domain simulation studies.

6.2 CAE-HIFD Evaluation

This section first describes the details of CAE-HIFD model training and the effects of different CAE-HIFD components are presented. Next, the response of the CAE-HIFD to different case studies is demonstrated. Finally, the CAE-HIFD performance is compared with other HIF detection approaches, and its sensitivity to noise is examined.

6.2.1 CAE-HIFD Model Training

The performance of the CAE is evaluated using accuracy (Acc), security (Sec), dependability (Dep), safety (Saf), and sensibility (Sen) [30].

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{6.1}$$

$$Sec = \frac{TN}{TN + FP} \times 100\% \tag{6.2}$$

$$Dep = \frac{TP}{TP + FN} \times 100\% \tag{6.3}$$

$$Saf = \frac{TN}{TN + FN} \times 100\% \tag{6.4}$$

$$Sen = \frac{TP}{TP + FP} \times 100\% \tag{6.5}$$

Here True Positives (TP) and True Negatives (TN) are the numbers of correctly identified fault

and non-fault cases, and False Negatives (FN) and False Positives (FP) are the numbers of miss-classified fault and non-fault cases. The accuracy is the percentage of overall correctly identified states, the security is the healthy state detection precision, the dependability is the fault state detection precision, the safety is resistance to faulty tripping, and the sensibility is resistance to unidentified faults [30]. Note that dependability is also referred to as True Positive Rate (TPR) or sensitivity, while security is referred to True Negative Rate (TNR) or specificity [38].

To achieve high accuracy, the CAE hyperparameters must be tuned; this includes the number and size of kernels, learning rate, optimizer, and batch size. Hyperparameter tuning is performed with Grid Search Cross-Validation (GSCV), wherein an exhaustive search is conducted over pre-specified parameter ranges. The GSCV determines the best performing parameters based on the scoring criteria provided by the model, in our case accuracy. The tuned CAE has 256-128-128-256 filters of size 3×3 in the four convolution layers, the optimizer is Adam, the learning rate is 0.001, and the batch size is 16.

The window size after differencing is 166 voltage/current samples, which corresponds to one cycle of the 60 Hz power frequency signal sampled at a rate of 10 kHz. The proposed method does not operate based on the fundamental frequency components of the input signal and thus, is not sensitive to frequency deviations as shown in Subsection 6.2.3, case study VII.

The sliding window stride during training impacts the CAE-HIFD performance: its value is determined on the training data set. Once the system is trained, it is used with the stride of one. The upper bound for the stride value is 166 as a higher value would lead to skipped samples. The performance metrics for varying stride values are shown in Fig. 6.3. It can be observed that the safety and dependability are not affected by the change in the stride value as none of the non-HIF cases are misclassified as a HIF case. The accuracy, security, and sensibility are 100% for the stride size of 166, and for shorter strides, these metrics are slightly lower. As the stride decreases, a few data windows are mistakenly identified as faults resulting in decrease of security, sensibility, and accuracy. Hence, the stride value of 166 is selected for the sliding

window in the preprocessing of the training data set.
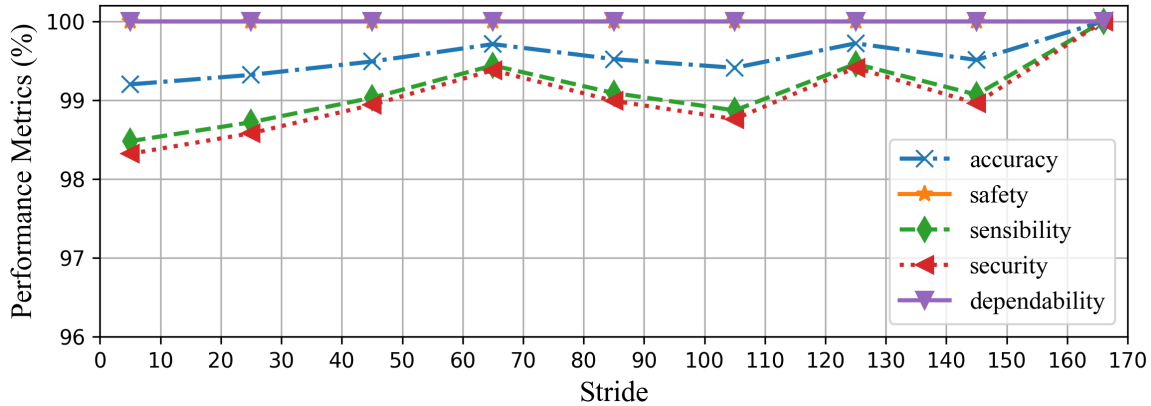
Figure 6.3: Impact of the sliding window stride on the performance metrics of the CAE-HIFD.

The CAE-HIFD prevents false detection of disturbances as HIFs using kurtosis value. As the distributions of the HIFs and disturbances both exhibit Gaussian behavior, an HIF window can have a K value close to the K value of a disturbance. The kurtosis threshold is determined starting from the kurtosis values for which all HIFs scenarios in the training data lie below this threshold: in this case 10. Next, accuracy on the training data is examined with thresholds close to this initial threshold. As illustrated in Fig. 6.4, the accuracy is 100% when kurtosis thresholds are between 9.5 and 10.5. The accuracy decreases when threshold is below 9.5 because some of the HIF scenarios are mistakenly detected as non-HIF scenarios ($FN > 0$). Furthermore, the threshold above 10.5 leads to low accuracy as some non-HIF scenarios are falsely declared as HIFs ($FP > 0$). Consequently, the kurtosis threshold of 10 is selected to discriminate the non-fault disturbances from the HIFs.

6.2.2   Effects of CAE-HIFD's Components

The proposed CAE-HIFD uses differencing and cross-correlation in addition to the main component, the CAE, to increase various performance metrics. Furthermore, kurtosis is utilized to improve the security of the proposed method. Consequently, as depicted in Table 6.2, the CAE-HIFD achieves 100% performance in all five considered metrics regardless of the surface
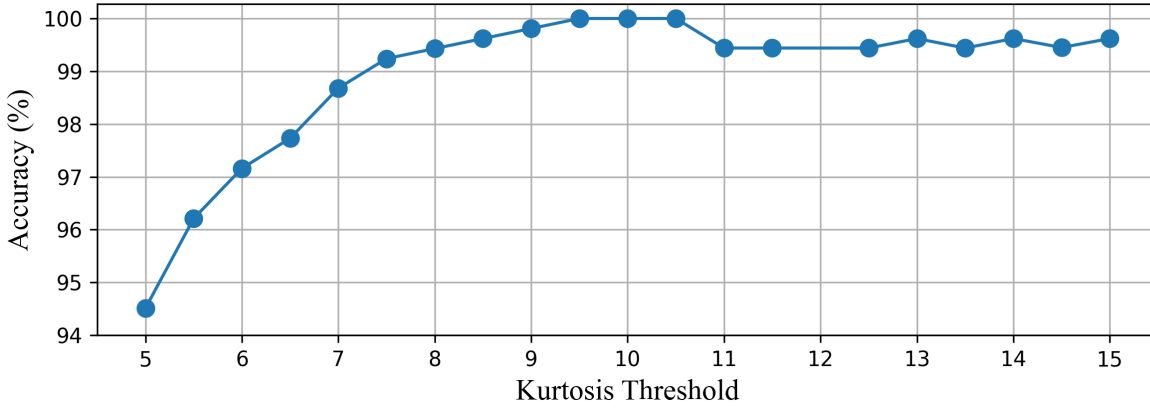
Figure 6.4: Impact of the Kurtosis threshold on the accuracy of the CAE-HIFD.

type, inception angle, and fault locations.

Table 6.2: Impact of differencing, cross-correlation, and kurtosis on CAE-HIFD performance.

Model	Acc	Saf	Sen	Sec	Dep
Proposed CAE-HIFD	100	100	100	100	100
CAE with CC, K	51.19	100	51.10	0.37	100
CAE with K, Diff	48.29	4.76	50.10	0.40	92.67
CAE with CC, Diff	96.38	100	93.31	92.70	100
CAE with CC	51.10	100	50.64	1.84	100
CAE with Diff	48.9	37.93	49.51	4.04	93.43
CAE with K	89.64	96.67	84.16	82.77	96.35

Additionally, Table 6.2 includes variants of the CAE-HIFD with only some of the three components included. With only CC and kurtosis, the accuracy and sensibility drop to nearly 51%, and the security decreases by 99.6%. In the absence of differencing, the CAE cannot learn patterns to distinguish between the HIF and the non-HIF data windows and, as a result, a large number of the non-HIF data windows are falsely classified as the HIF which means high FP, thus low security.

To examine the impact of CC, the traditional MSE is used in place of CC to measure the similarity of the input and reconstructed signal. As shown in Table 6.2, in the absence of CC, the values of accuracy and sensibility drop to nearly 50%. This happens because the MSEs calculated for the HIF and non-HIF data windows are similar and, thus, non-HIFs are falsely detected as HIFs. Furthermore, the security value is low (0.40%), whereas the dependability

value is high (92.67%) as there are only a few TN and FN compared to TP and FP.

Omitting the kurtosis evaluation results in only a small increase in the number of FP cases which are the non-fault disturbances falsely declared as HIFs. Therefore, as shown in Table 6.2, all the performance metrics values decrease by less than 8%.

Finally, only one out of the three components is included in the CAE-HIFD framework. Whereas the simultaneous use of differencing and cross-correlation achieves relatively high performance metrics, with only one of the two components, there is a major decrease in security (more than 95%). With kurtosis only, all metrics are between 82% and 97% in comparison to 100% obtained in the presence of all three components. This is caused by the absence of differencing which assists in amplifying sign wave distortions and omission of cross-correlation which facilitates signal comparisons. The results shown in Table 6.2 highlight the necessity of each CAE-HIFD component and the contribution of each component to the HIF detection performance.

6.2.3   CAE-HIFD Response to Different Case Studies

In this section, seven case studies have been conducted to depict the response of the proposed CAE-HIFD.

Case study I − Close-in HIF

Fig. 6.5 illustrates the performance of the CAE-HIFD under both normal and HIF conditions. In this case study, the HIF is applied at Node 632 starting at 0.05 s, as seen in Fig. 6.5 (a). The input voltage and current signals observed at the substation relay are shown in figures 6.5 (b) and (c), and the kurtosis calculated from those voltages and currents is displayed in Fig. 6.5 (d). During normal operation, the kurtosis is below the threshold; upon the HIF inception, it raises over the threshold for approximately 8-10 ms returning quickly back to below threshold values. The HIF causes the CC value to rise above the threshold, Fig. 6.5 (e), and, therefore, a trip signal is issued approximately 60 ms after the HIF inception as seen in Fig. 6.5 (f).

Figure 6.5: CAE-HIFD performance under normal and close-in HIF conditions: (**a**) HIF current, (**b**) three-phase input voltages, (**c**) three-phase input currents, (**d**) kurtosis, (**e**) cross-correlation, and (**f**) trip signal.

Case study II − Remote HIF

Fig. 6.6 depicts the result of the CAE-HIFD in presence of a remote HIF: the HIF is applied at Node 652 starting at 0.05 s, as seen is Fig. 6.6 (a). The input voltage and current signals

are observed in figures 6.6 (b) and (c), and the calculated kurtosis is shown in Fig. 6.6 (d). Due to the remote location of HIF, the HIF influence on the voltages and current signals is highly attenuated. As a result, the kurtosis surpasses the threshold for shorter duration of time (approximately 1-2 ms) as compared to Case study I. As shown in Fig. 6.6 (e), the CC value raises above the threshold after the inception of HIF. Consequently, a trip signal is issued approximately 50 ms after the HIF inception as seen in Fig. 6.6 (f).

Case study III − Capacitor switching

The proposed HIF detection method successfully discriminates HIFs from switching events as shown in Fig. 6.7 with a three-phase capacitor bank located at node 675. Fig. 6.7 (a) depicts phase *A* current caused by the capacitor energization at t = 0.05 s. The current and voltage signals seen by the relay at substation exhibit significant oscillations as shown in figures 6.7 (b) and (c). This switching event causes sudden increase in the kurtosis for a short duration of time, approximately 15 ms (Fig. 6.7 (d)). Although the CC for the switching event is higher than its threshold, Fig. 6.7 (e), this disturbance is not falsely identified as an HIF, due to the high kurtosis value. Moreover, the CC for remaining non-HIF signal is below the threshold. Consequently, a trip signal is not issued throughout the switching event, Fig. 6.7 (f).

Case study IV − Nonlinear load

Fig. 6.8 shows the performance of the proposed CAE-HIFD in presence of a nonlinear load which causes significant harmonics. The load at node 634 is replaced by a DC motor fed by a six-pulse thyristor rectifier. The motor is started at t = 0.05 s. Fig. 6.8 (a) illustrates the phase A current of the nonlinear load, while figures 6.8 (b) and (c) show voltages and currents measured by the relay at the substation. Although the CC is higher than its threshold (Fig. 6.8 (e)), the trip signal (Fig. 6.8 (f)) is not issued because the kurtosis surpasses its threshold (Fig. 6.8 (d)).
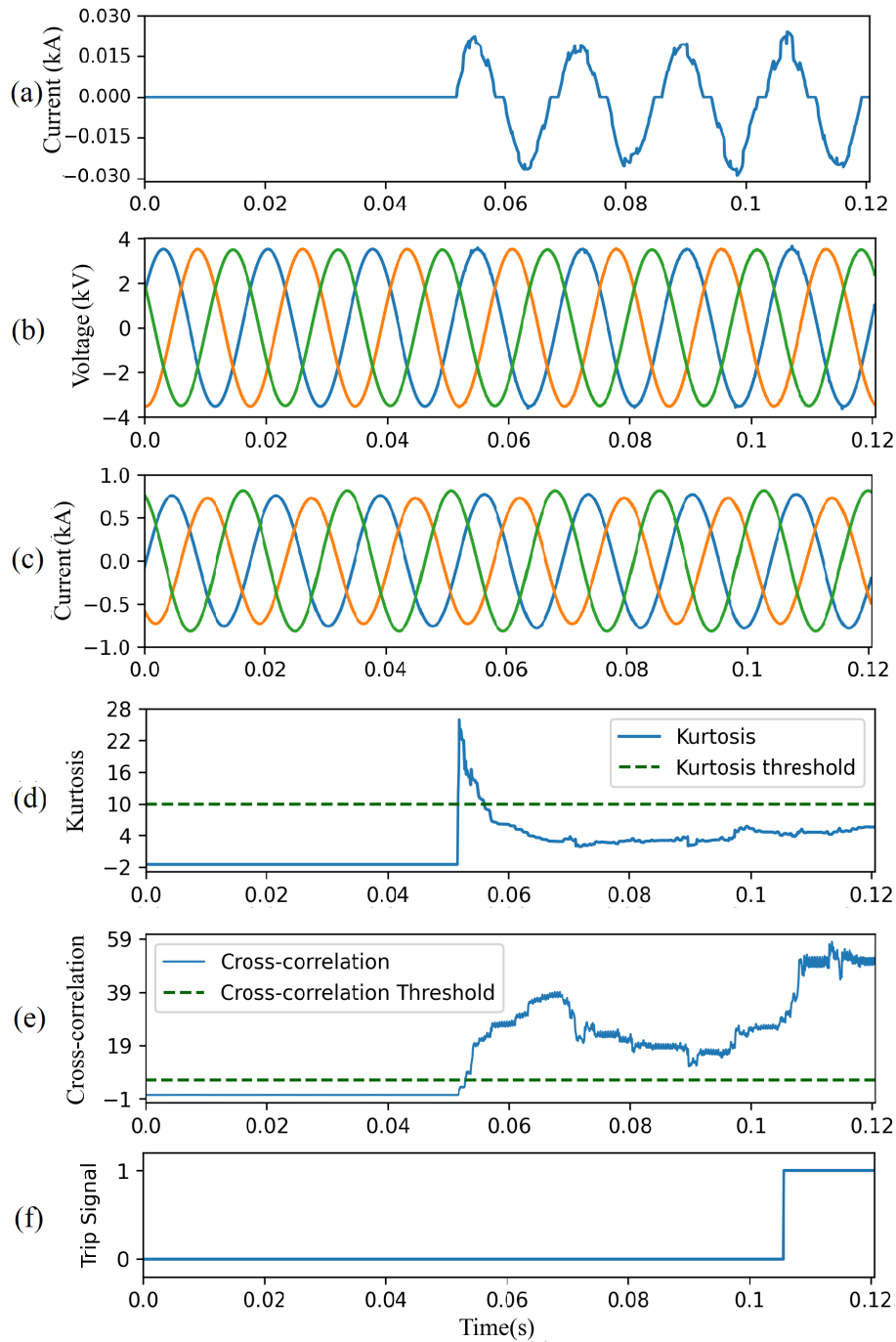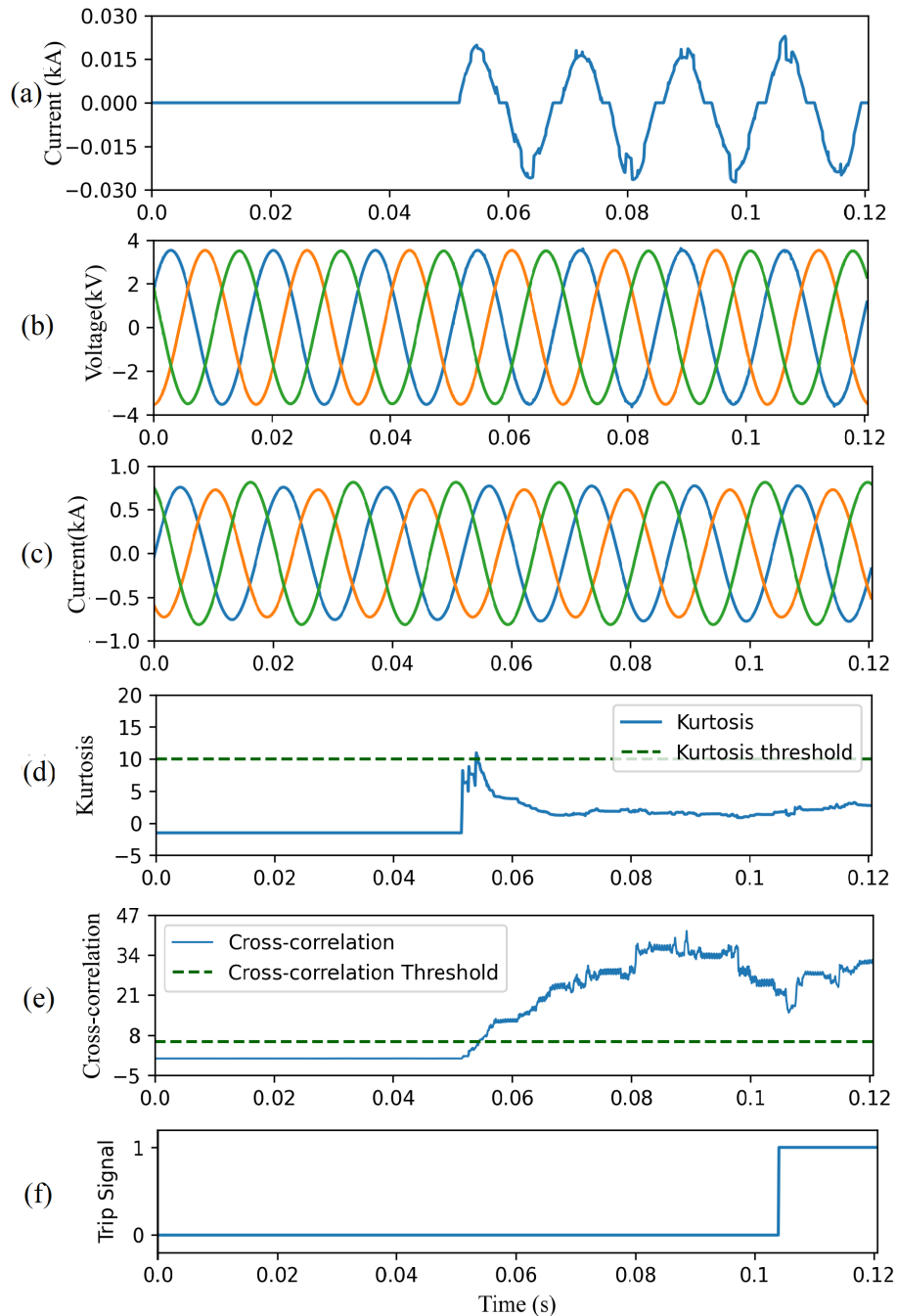
Figure 6.6: CAE-HIFD performance under normal and remote HIF conditions: (**a**) HIF current, (**b**) three-phase input voltages, (**c**) three-phase input currents, (**d**) kurtosis, (**e**) cross-correlation, and (**f**) trip signal.

Case study V − Transformer energization

This case study investigates the performance of the CAE-HIFD under a transformer energiza-

tion scenario: the transformer at node 633 is energized at t=0.05 s. The inrush current for
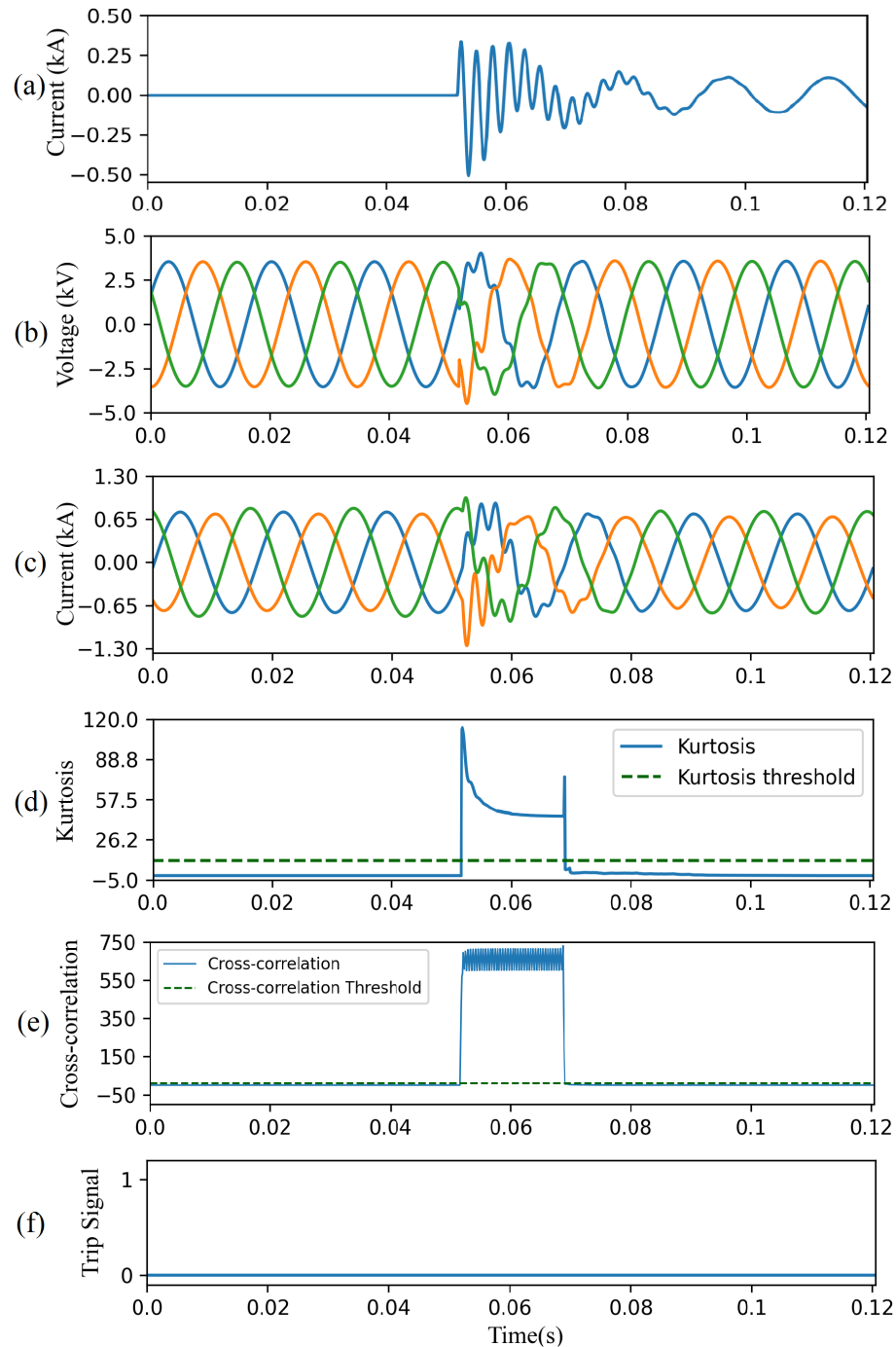
Figure 6.7: CAE-HIFD performance under a capacitor switching scenario: (**a**) capacitor current, (**b**) three-phase input voltages, (**c**) three-phase input currents, (**d**) kurtosis, (**e**) cross-correlation, and (**f**) trip signal.

phase-A is shown in Fig. 6.9 (a) while figures 6.9 (b) and (c) display voltages and currents measured at the substation. Both the resulting kurtosis shown in Fig. 6.9 (d), and the CC
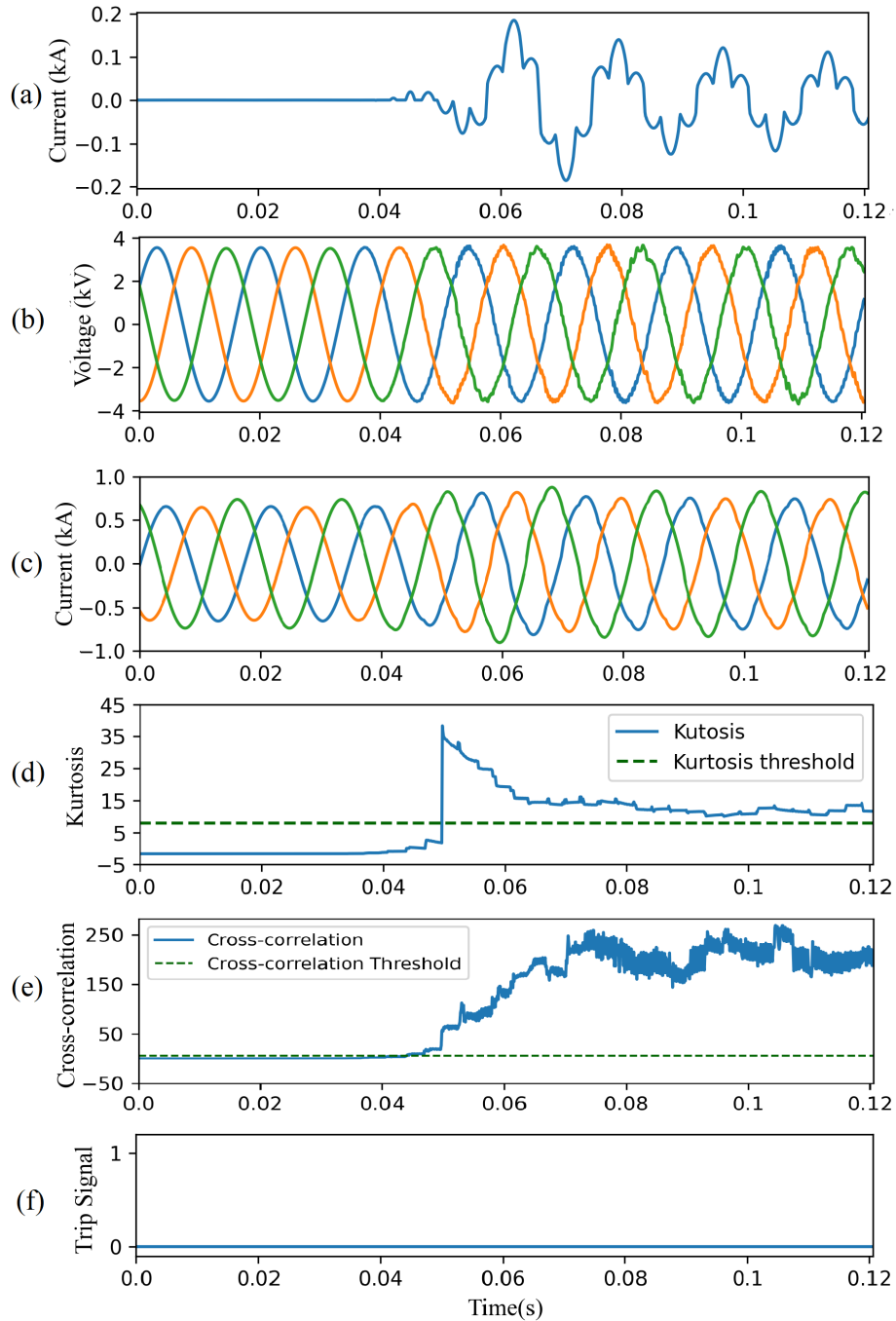
Figure 6.8: CAE-HIFD performance under non-linear load switching: (**a**) non-linear load current, (**b**) three-phase input voltages, (**c**) three-phase input currents, (**d**) kurtosis, (**e**) cross-correlation, and (**f**) trip signal.

shown in Fig. 6.9 (e) are below their corresponding thresholds. As a result, the proposed protection strategy does not cause any unnecessary tripping (Fig. 6.9 (f)) under the transformer energization scenario.
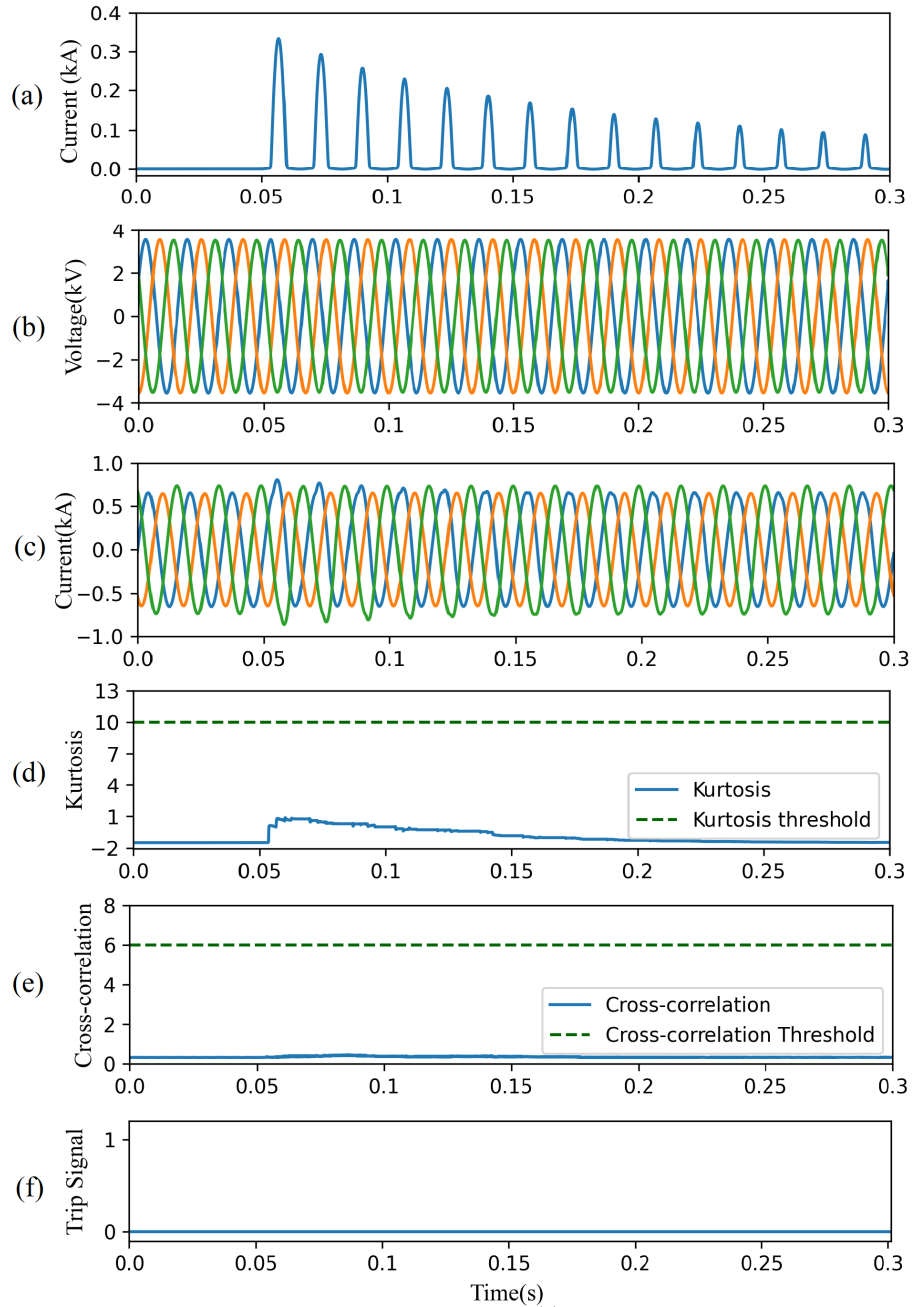
Figure 6.9: CAE-HIFD performance under inrush currents: (**a**) Inrush current, (**b**) three-phase input voltages, (**c**) three-phase input currents, (**d**) kurtosis, (**e**) cross-correlation, and (**f**) trip signal.

Case study VI − Intermittent HIFs

This case study demonstrates the effectiveness of the proposed CAE-HIFD in detecting intermittent HIFs. A tree branch momentarily connects the phase-A to the ground for approximately

3.5 cycles (55 ms) as illustrated in Fig. 6.10 (a). The voltage and current signals shown in figures 6.10 (b) and (c) are measured by the relay at the substation. As depicted in Fig. 6.10 (d), the kurtosis does not exceed the threshold. The CC in Fig. 6.10 (e) crosses the threshold during the intermittent faults. As shown in Fig. 6.10 (f), the trip signal is issued after 50 ms. The trip signal is reset after the intermittent fault is cleared.

Case study VII – Frequency Deviations

To demonstrate the effectiveness of the proposed method in presence of frequency deviations, the system frequency is increased to 61 Hz in this case study. The HIF is initiated at t = 0.05 s. figures 6.11 (a) and (b) represent currents and voltages measured by the relay at the substation. As shown in Fig. 6.11 (c), before the HIF takes place, the kurtosis is below the threshold. As the HIF samples enters the sliding window, the kurtosis exceeds the threshold because the distribution suddenly changes during the transition. Next, the kurtosis returns to values below the threshold. The CC in Fig. 6.11 (d) is above the CC threshold; therefore, the system trips within three cycles of the HIF inception.

6.2.4   Comparison with Other Approaches

This section first compares the proposed CAE-HIFD with other supervised and unsupervised learning algorithms. The two supervised models selected for the comparison are: Support Vector Machine (SVM) [6] and Artificial Neural Network (ANN) [27]. As supervised models require the presence of both, HIF and non-HIF data in the training set, these models are trained with a data set containing an equal number of the HIF and non-HIF instances. Moreover, as those models have originally been used with the DWT applied on the current waveform [6, 27], DWT is used here too. DWT extracts features by decomposing each phase current into seven detail level coefficients and one approximate level coefficient using the db4 mother wavelet. The features are formed by computing the standard deviation of coefficients at each level; therefore, eight standard deviations from each phase form a new input sample with 24 elements
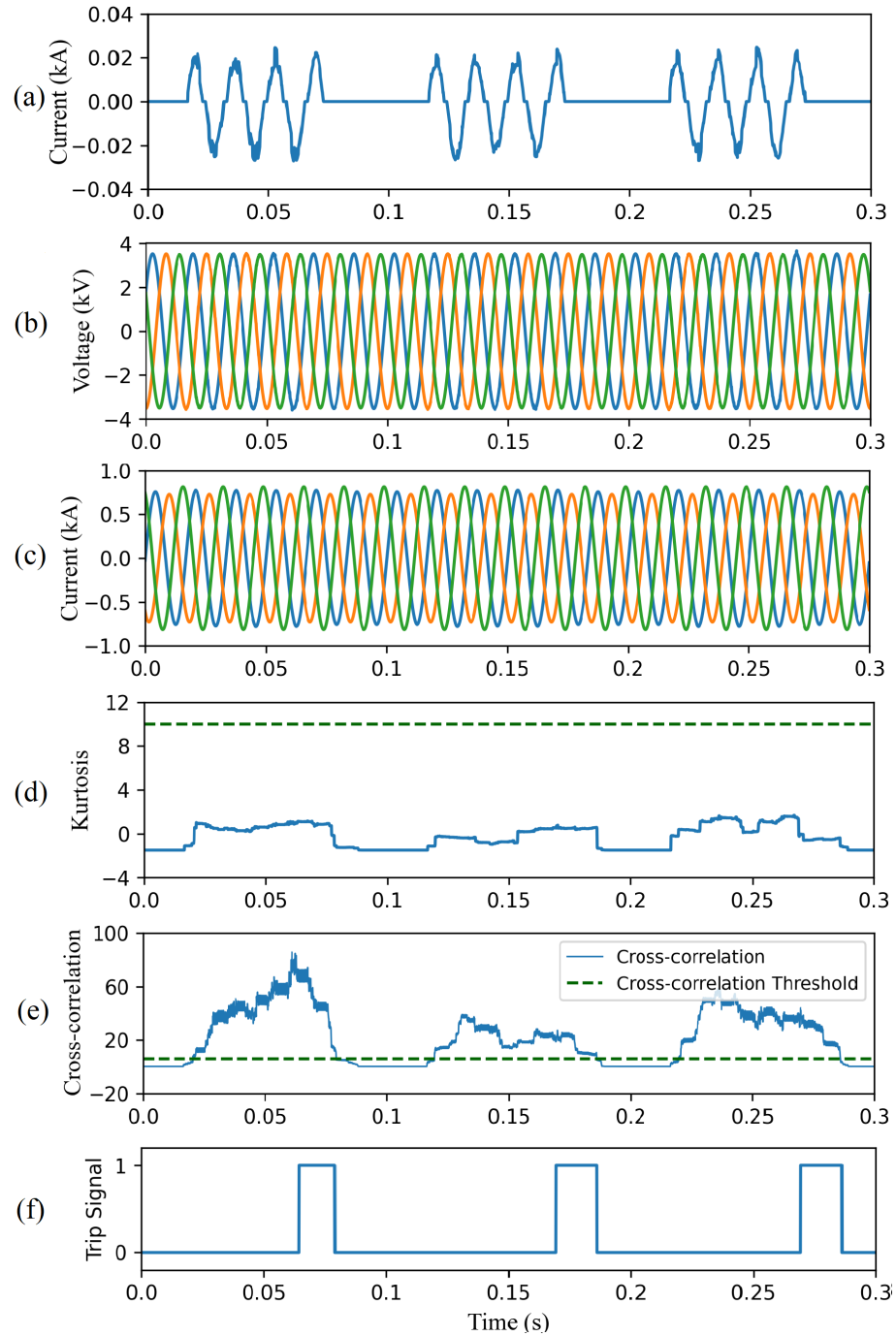
Figure 6.10: CAE-HIFD performance under intermittent HIF condition: (**a**) fault current, (**b**) three-phase input voltages, (**c**) three-phase input currents, (**d**) kurtosis of the input, (**e**) cross-correlation, and (**f**) trip signal.

[27]. As with CAE-HIFD, the SVM and ANN hyperparameters are tuned using GDCV. The SVM kernel is RBF with γ of 0.05. The ANN has three layers with 24-18-1 neurons, the
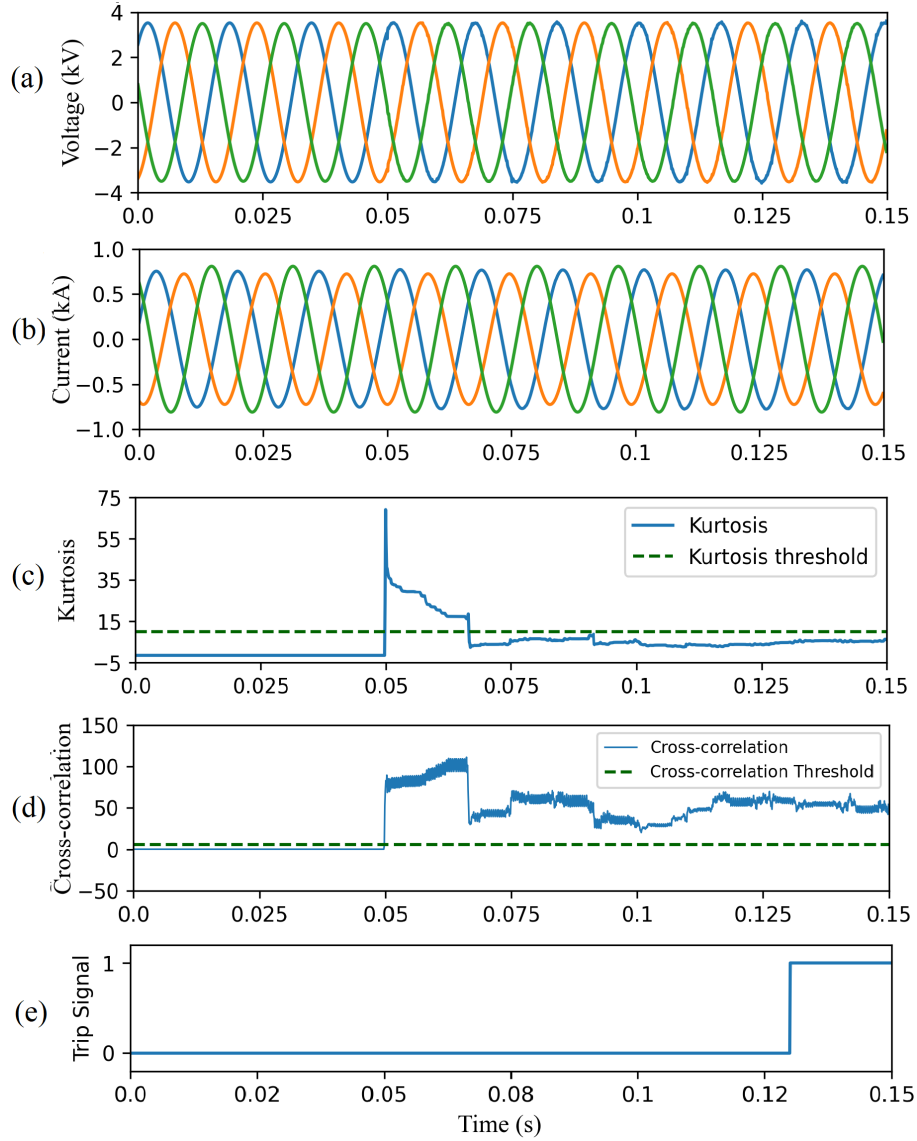
Figure 6.11: CAE-HIFD performance under 61 Hz fundamental frequency: (**a**) three-phase input voltages, (**b**) three-phase input currents, (**c**) kurtosis, (**d**) cross-correlation, and (**e**) trip signal.

activation function for input and hidden layers is ReLU, and binary cross-entropy is the loss function.

As other studies have only used supervised learning, to examine unsupervised learning techniques, variations of the proposed approach are considered in this evaluation. Fig. 6.12 shows the flowchart for the unsupervised ML models. Preprocessing, Kurtosis, and CC calculation components are exactly the same as in the proposed CAE-HIFD, while two options are

considered for the autoencoder algorithm and the training dataset. As the autoencoder, the proposed CAE-HIFD uses CAE while here we consider a variant of Recurrent Neural Network, Gated Recurrent Units Autoencoder (GRU-AE). GRU-AE is selected because it issuccessful in extracting patterns from time-series data such as those present in the current and voltage signals. The GRU-AE tuned with GDCV has two hidden layers, each one with 32 GRU cells and the ReLU activation function. For both, CAE-HIFD and GRU-AE, two types of studies are conducted: training on HIFs data only and training on non-HIFs data only.
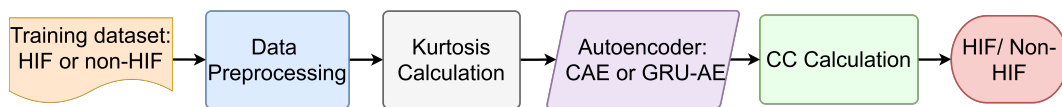
Figure 6.12: Flowchart for the alternative unsupervised HIF detection models

For the comparison with an unsupervised learning model, a variant of Recurrent Neural Network, Gated Recurrent Units Autoencoder (GRU-AE) is selected because the GRU-AE is successful in extracting patterns from time-series data such as those present in the current and voltage signals. The GRU-AE tuned with GDCV has two hidden layers, each one with 32 GRU cells and the ReLU activation function. As show in Fig. 6.12, the GRU-AE is used in place of CAE in the proposed CAE-HIFD approach keeping data preprocessing, kurtosis, and cross-correlation the same. For both, CAE-HIFD and GRU-AE, two types of studies are conducted: training on HIFs data only and training on non-HIFs data only.

The results of the comparison between CAE-HIFD and the other approaches are shown in Table 6.3. It can be observed that CAE-HIFD outperforms other approaches and is the only one not susceptible to false tripping as indicated by the security metrics. In addition, the CAE-HIFD trained only with HIF data is highly efficient in discriminating non-HIF instances by detecting deviations from the learned HIF patterns. This prevents the algorithm from false tripping in the case of a new non-HIF pattern not present in the training set. The results of the studies presented in Table 6.3 indicate that the CAE-HIFD achieves equally good results regardless of whether it is trained on the HIF data or non-HIF data; however, when trained

with non-HIF data, there is a risk of identifying new non-HIF patterns as HIFs. The supervised learning-based approaches category can only recognize the pattern present in the training set and, thus, may recognize new non-HIF events as HIFs. Overall, the proposed CAE-HIFD achieves better performance than the other approaches.

Table 6.3: Comparison of CAE-HIFD with other HIF detection approaches.

Model	Acc	Saf	Sen	Sec	Dep
Other Approaches - Supervised					
DWT+SVM [6]	97.97	100	97.78	78.99	100
DWT+ANN [27]	97.72	100	97.55	76.47	100
Variants of our approach - Unsupervised					
Train on non-faults					
GRU-AE	99.92	100	96.92	99.92	100
Proposed CAE-HIFD	100	100	100	100	100
Train on faults					
GRU-AE	34.61	32.24	83.22	97.52	5.66
Proposed CAE-HIFD	100	100	100	100	100

6.2.5   Robustness of the Proposed CAE-HIF Against Noise

To examine CAE-HIFD robustness against noise, studies are conducted by introducing different levels of noise. The white Gaussian noise is considered because it covers a large frequency spectrum. The noise is added to the current signals because current waveforms are more susceptible to noise [1]. The ratio of the signal power to the noise power (SNR) is considered to measure the level of noise added to the system. The SNR is measured in decibel and given as:

$$SNR = 20 \log \frac{S_p}{N_p} \tag{6.6}$$

Where S_p and N_p refers to the signal and noise power, respectively. As shown in Fig. 6.13, the proposed CAE-HIFD approach is immune to noise when SNR value is higher than 40 dB. In a case of high nose, SNR below 40 dB, the accuracy reduces to 97%. Thus, more than one consecutive window is needed to be processed before making a tripping decision in order to

avoid undesired tripping and to ensure accurate HIF detection. Therefore, three consecutive windows are utilized in all performance evaluation studies in order to accurately detect all HIFs. Increasing the timer threshold improves the resiliency against unnecessary tripping, but prolongs HIF detection time. Even with the extremely noisy condition of 1 dB SNR, the accuracy, security, and sensibility do not fall below 97.65%, 94.03%, 96.28% respectively. The reason behind this robustness is the CAE de-noising ability and strong pattern learning capability. The inherent de-noising nature of the autoencoders assists the CAE to generalize the corrupted input. Also, the CAE-HIFD learns the complex HIF patterns because of spatial feature learning proficiency of the CAE. The model accurately detects non-HIF scenarios under considered noise levels; hence, the safety and dependability remains at 100% even with high levels of noise. Moreover, the values of other performance metrics are also greater than 90% throughout the SNR range of 5dB to 50dB demonstrating noise robustness of the CAE-HIFD.

Fig. 6.14 shows CAE-HIFD performance in presence of noise of 20 dB SNR. Before the HIF inception at 0.05 s, despite the significant noise, the kurtosis (Fig. 6.14 (b)) and the CC (Fig. 6.14 (c)) remain below their thresholds. The CC surpasses the threshold upon HIF inception and, as a result, the designed protection system issues trip signal (Fig. 6.14 (d)).
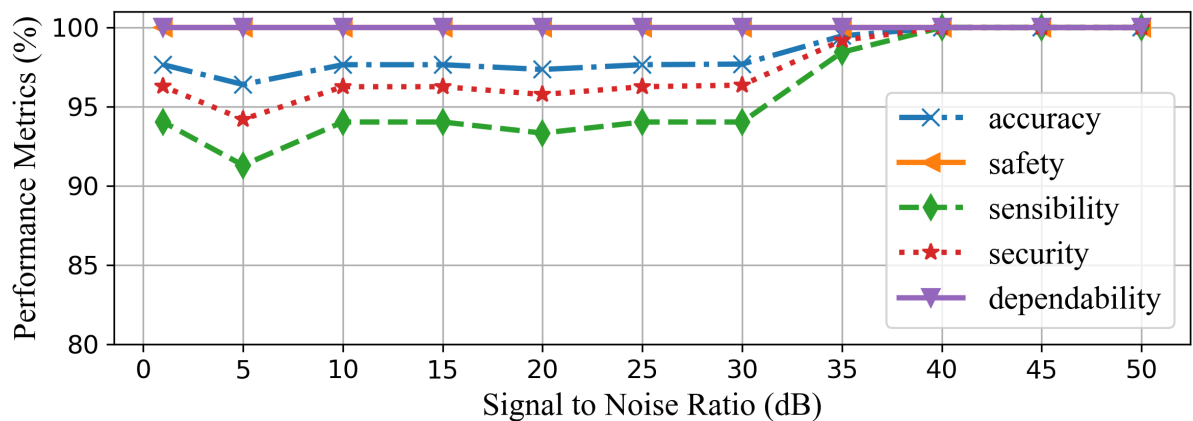
Figure 6.13: Effect of noise on the CAE-HIFD performance.
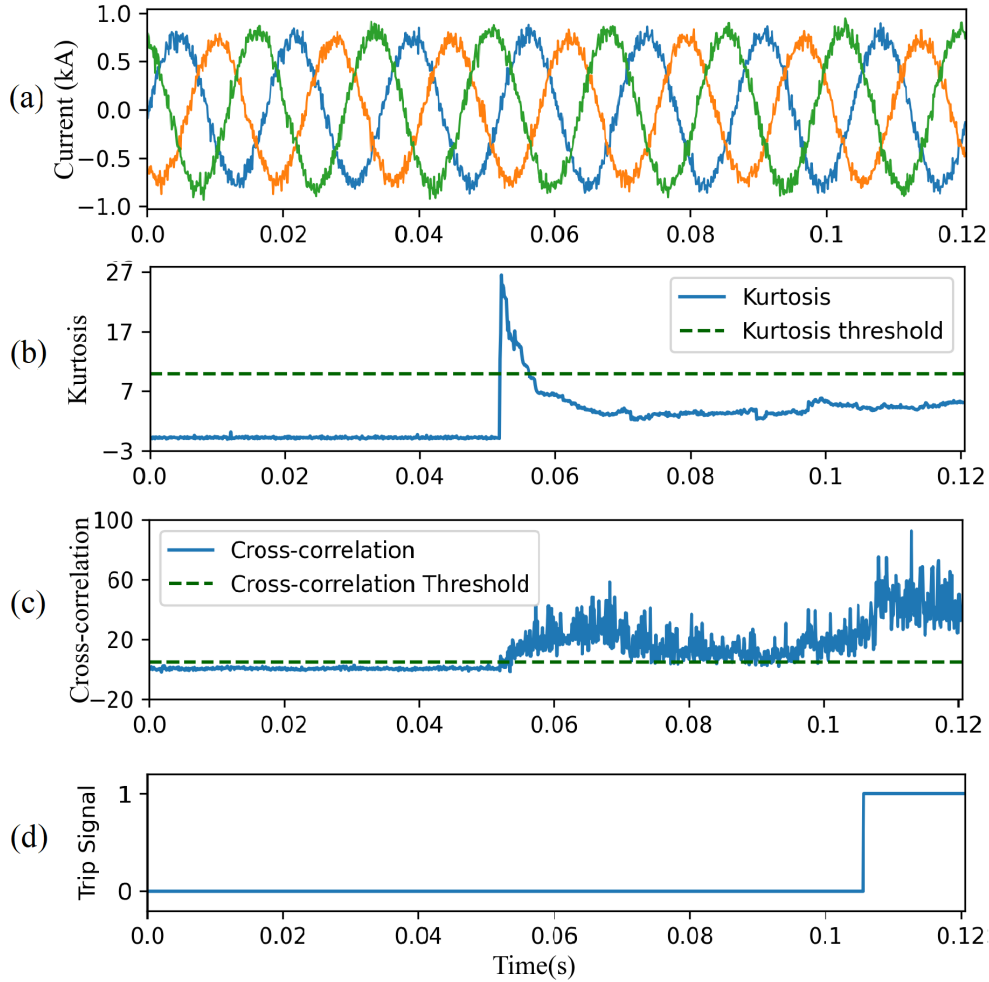
Figure 6.14: CAE-HIFD performance under 20 dB SNR: (**a**) three-phase input currents, (**b**) kurtosis, (**c**) cross-correlation, and (**d**) trip signal.

6.2.6 Discussion

The evaluation results demonstrate that the proposed CAE-HIFD achieves 100% HIF detection accuracy irrespective of the surface type, fault phase, and fault location. All metrics, including accuracy, safety, sensibility, security, and dependability are at 100% as shown in Table 6.2. Moreover, for all considered scenarios, the system trips within three cycles after the HIF inception.

The challenging part of machine learning for HIF detection is in the diversity of non-fault and fault signals together with similarities between non-HIF disturbances and HIFs. By training on faults only, the proposed approach does not require simulation of non-fault scenarios

for training. Distinguishing HIFs from the non-HIF steady-state operation can take advantage of the smoothness of non-HIF steady-state signal; however, non-HIF disturbances, such as capacitor and load switching, share many characteristics (e.g., randomness and non-linearity) with the HIF signals making it difficult for a neural network (in our case CAE) to distinguish between them. To address this challenge, the proposed approach takes advantage of differences in data distributions between non-fault disturbances and HIFs and employs kurtosis to differentiate between the two.

In experiments, 210 fault cases were considered corresponding to 1372 fault data windows as described in Subsection 6.1. Signals corresponding to these faults are different from each other as simulations included different surfaces, fault locations, and fault phases. From these fault cases, 80% is selected randomly for training, therefore, some cases are present only in testing. Moreover, all the case studies presented in Subsection 6.2.3 are conducted with data that is not seen by the proposed CAE-HIFD in training. The proposed system successfully distinguished between fault and non-fault signals for all scenarios, which demonstrates its abilities to detect previously unseen HIF and non-HIF scenarios.

Frequency deviations, as well as noise, impose major challenges for the HIF detection. Approaches that operate on the fundamental frequency components risk failures in presence of frequency deviations. However, CAE-HIFD does not operate based on the fundamental frequency components of the input signals and, consequently, is not sensitive to frequency deviations as shown in Subsection 6.2.3. As noise is common in distribution systems, it is important to consider it in HIF detection evaluation. HIF detection in presence of noise is difficult as noisy signals are accompanied by randomness and have characteristics that resemble HIFs. Nonetheless, experiments from Subsection 6.2.5 show that CAE-HIFD remains highly accurate even in presence of significant noise.

6.3   T-CNN Evaluation

This section describes the details of model training, the effect of kurtosis and differencing on the performance, and response of the T-CNN to various challenging scenarios. Furthermore, the proposed approach is compared with the state-of-the-art approaches and the effects of sampling frequency and noise are examined.

6.3.1   T-CNN for HIF Classification Model Training

The performance of the T-CNN framework is evaluated with accuracy and F1-score:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6.7}$$

$$F1 - score = \frac{2 * Sensibility * Dependability}{Sensibility + Dependability} \tag{6.8}$$

$$Sensibility = \frac{TP}{TP + FP} \tag{6.9}$$

$$Dependability = \frac{TP}{TP + FN} \tag{6.10}$$

where True Positives (TP) and True Negatives (TN) are the numbers of correctly detected fault types and non-fault cases, and False Positives (FP) and False Negatives (FN) are the number of falsely detected fault types and non-fault cases. Although accuracy depicts the overall performance of the machine learning algorithm, it fails to determine the success rate of detecting a class in a multi-class output scenario [70]. Consequently, the F1-score is utilized, which is the harmonic mean of sensibility and dependability.

Initially, hyperparameter optimization of a Machine Learning (ML) model training is necessary to attain high accuracy. The hyperparameters of the proposed T-CNN include the number of heads, number and size of kernels, optimizer, learning rate, and batch size. The Grid Search Cross-Validation (GDCV) is utilized to conduct hyperparameter tuning. In GDCV, exhaustive search is executed over pre-specified parameter ranges based on scoring criteria, e.g., accuracy. The tuned T-CNN has one encoder layer with four heads, a convolution layer containing 64
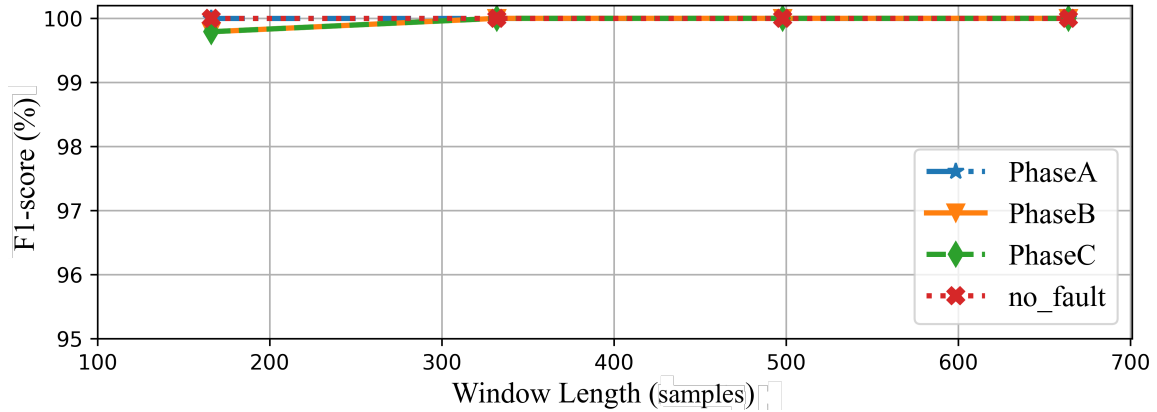
Figure 6.15: Impact of the window size on accuracy.

kernels of width 3, a fully connected layer, a linear layer, the optimizer is Adam, the batch size is 16, and the learning rate is 0.001.

For optimal performance, the size of the window is tuned using the training dataset. As one cycle of the 60 Hz power frequency signal contains 166 voltage/current samples, multiples of 166 are considered as window sizes. As seen in Fig. 6.15, for window sizes greater than 166, the F1-score in the case of all classes is 100%. As the data window size increases, T-CNN has more samples to learn from; however, the time complexity also increases. Hence, the window size of 332 is selected because it is the smallest window achieving 100% F1-score. Next, the stride of the sliding window is set 1 during the training process in order to provide the maximum possible data to train the T-CNN. The maximum value of stride cannot exceed 332 as it would lead to the omission of samples. The stride of 332 avoids the overlapping of the data, as a result, 332 is chosen as the stride value for the sliding window in the pre-processing of the testing dataset.

6.3.2   Effect of Kurtosis and Differencing

The proposed T-CNN for HIF classification utilizes a kurtosis threshold for detecting non-HIF disturbances. The kurtosis of all the HIF data windows in the training set is calculated, and the maximum value, i.e. 8 is selected as the kurtosis threshold. As shown in Fig. 6.16, F1-score of detecting Phase *B* and Phase *C* type fault decreases if threshold is selected below 7. This

happens because the lower threshold is mistakenly detecting HIF cases as non-HIF cases. On the other hand, raising the threshold above 9 leads to a decrease in the F-1score of detecting all the classes as the higher threshold is falsely detecting non-fault disturbances cases as HIF cases. As a result, the threshold value is selected as 8 to successfully discriminate between non-HIF disturbances and the HIF cases.

The proposed T-CNN along with components such as, differencing and kurtosis achieves 100% F1-score in detecting all phases and non-fault cases regardless of surface types, fault inception, and fault location, Fig. 6.17. The differencing is utilized for conducting feature extraction and kurtosis provides security against non-fault disturbances.
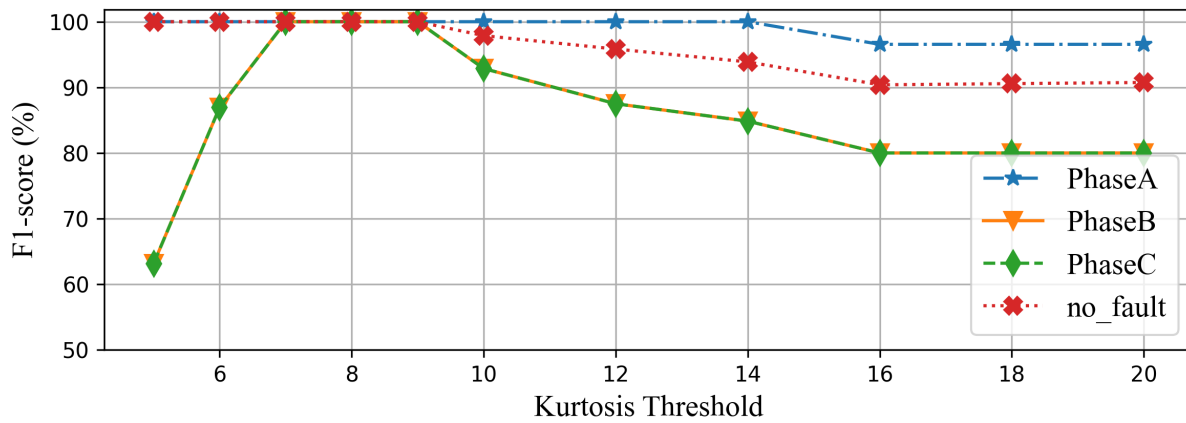
Figure 6.16: Impact of the kurtosis threshold on accuracy.

To examine the impact of kurtosis and differencing on the overall T-CNN framework performance, this section considers variants of the framework with only one of these two components. As seen from Fig. 6.17, while complete T-CNN achieves 100% F1-score for all classes, the omission of kurtosis leads to an increase in false positives reducing the F1-score for all classes. Without kurtosis, the F1-score for non-faults is approximately 89%, whereas phase classification F1-score is in the range of 75% to 93%.

The absence of differencing affects the F1-score even more, dropping it to below 50% for the three phases. The differencing amplifies distortions caused by the HIFs allowing the T-CNN model to comprehend the complex patterns present in the data and, thus, enabling distinction among phases. This comparison from Fig. 6.17 establishes the necessity of differencing and
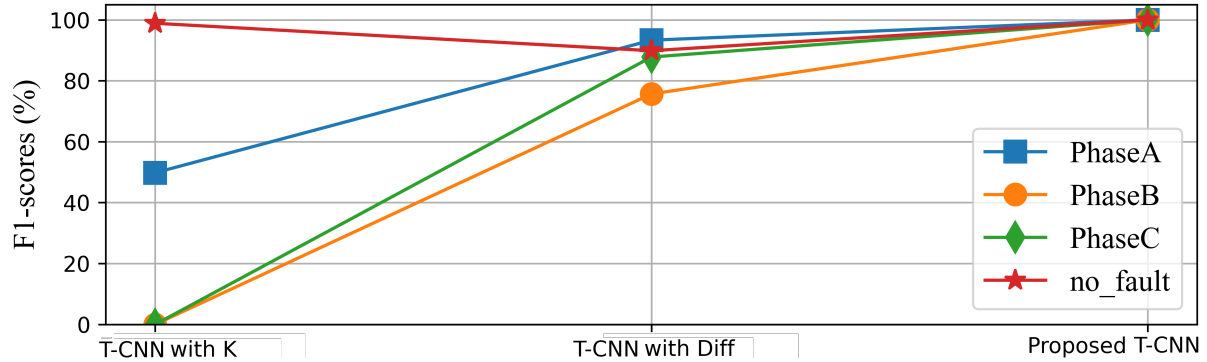
Figure 6.17: Impact of differencing and kurtosis on T-CNN performance.

kurtosis in the T-CNN framework.

6.3.3 System Response for Representative Case Studies

This section illustrates the response of the proposed framework for representative case studies: close-in HIF, capacitor switching, and nonlinear load.

Close-in HIF

Fig. 6.18 depicts the performance of the T- CNN framework under both normal and close-in HIF conditions. The input voltage and current signals observed by the substation relay and used for HIF detection are shown in figures 6.18 (a) and (b). The HIF is applied to Node 632, at 0.05 s, as shown in Fig. 6.18 (c). As seen from Fig. 6.18 (c), during the steady-state normal operation, the kurtosis is below the threshold and upon the HIF inception, it raises over the threshold for approximately 8-10 ms. As observed from Fig. 6.18 (e), after 0.06 s, phase A type fault has the highest output probability; thus, a trip signal, Fig. 6.18 (f), is issued approximately 45 ms after the HIF inception.

Capacitor Switching

The response of the proposed framework under the capacitor switching is shown in Fig. 6.19. As seen from figures 6.19 (a) and (b), capacitor switching causes significant oscillations in the current and voltage signals measured by the relay at substation when a three-phase capacitor
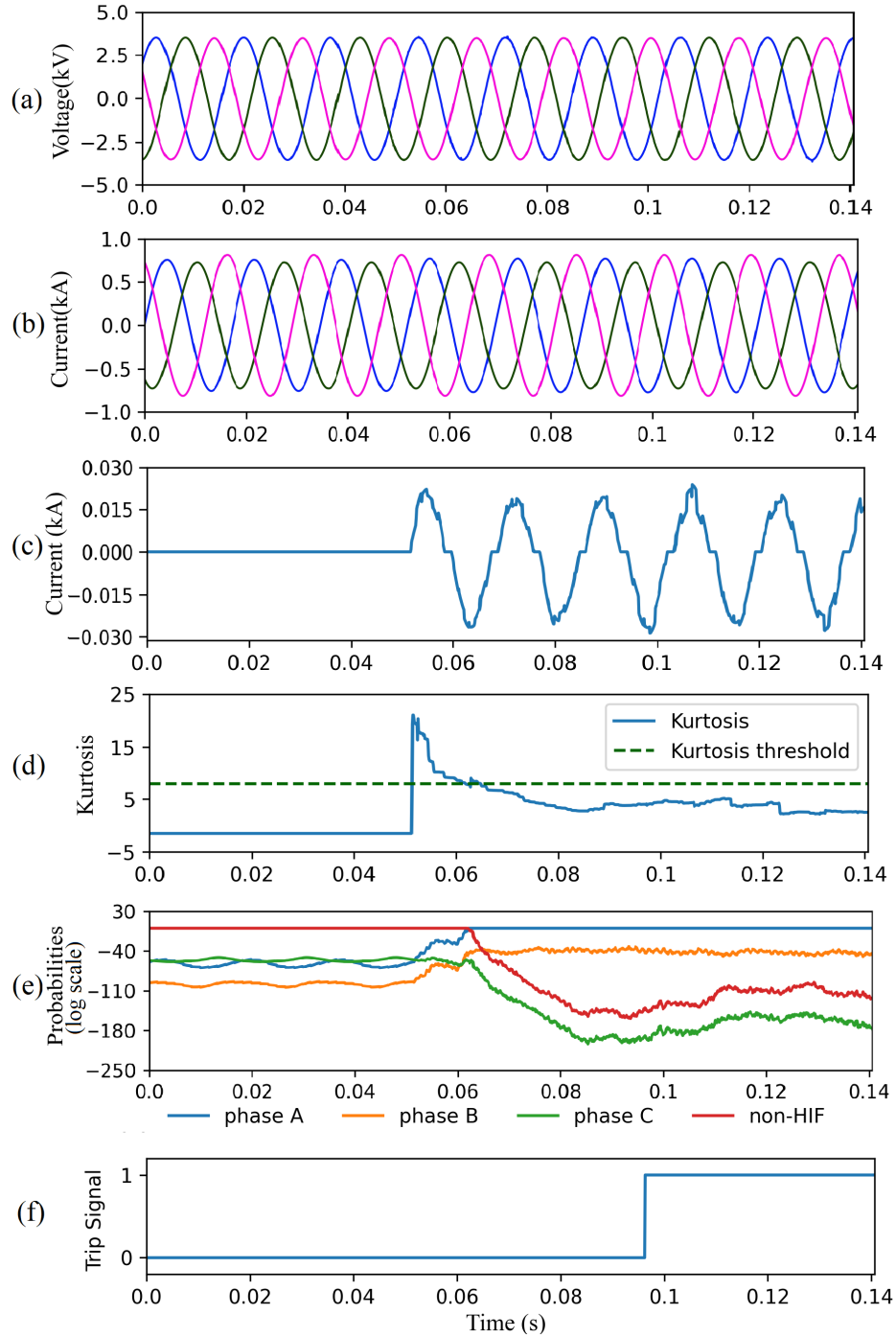
Figure 6.18: T-CNN performance under close-in HIF: (a) three-phase input voltages, (b) three-phase input currents, (c) HIF current, (d) kurtosis, (e) output probabilities, and (f) trip signal.

bank is energized at 0.05 s, as shown in Fig. 6.19 (c). These oscillations are captured by the sudden rise in the kurtosis value for 30 ms in Fig. 6.19 (d). Despite the output probability for phase *B* being the highest during the switching event (Fig. 6.19 (e)), the trip signal is not issued

as the kurtosis is above the threshold (Fig. 6.19 (d)). Thus, the proposed framework correctly detects switching events as non- HIF scenarios.
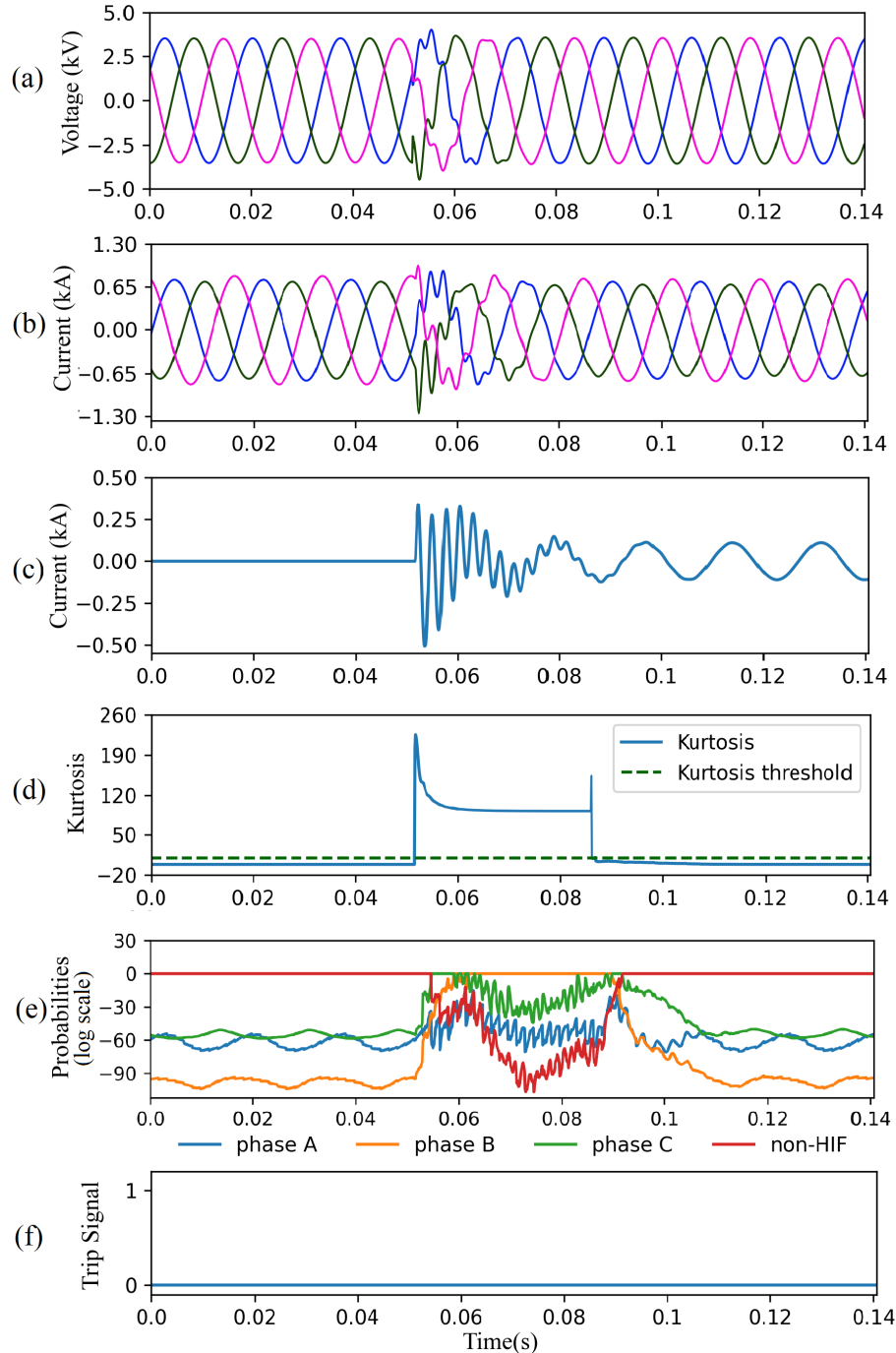
Figure 6.19: T-CNN performance under capacitor switching condition: (a) three-phase input voltages, (b) three-phase input currents, (c) capacitor current, (d) kurtosis, (e) output probabilities, and (f) trip signal.

Nonlinear Load

The DC motor fed by a six-pulse thyristor rectifier is utilized as a non-linear load. Figures 6.20 (a) and (b) illustrate the voltage and current signal observed by relay at substation while Fig 6.20 (c) depicts phase *A* current of the non-linear load. The high kurtosis value after 0.05 s, Fig. 6.20 (d), prevents the trip signal from occurring even though the probability of the non-HIF is lower than the probability of HIFs on the three phases, Fig. 6.20 (e). Hence, the proposed T-CNN successfully discriminates disturbance caused by non-linear load from HIFs.

Inrush Currents

This case study demonstrates the robustness of the proposed T-CNN in the presence of inrush currents. Figures 6.21 (a) and (b) represent the voltage and current signals measured by the relay at the substation. As shown in Fig. 6.21 (c), the inrush current at phase *A* is produced by the transformer being energized at $t = 0.05s$. The kurtosis value is lower than the threshold (Fig. 6.21), and the output probability for the non-HIF class is maximum throughout the inrush currents. Consequently, the proposed HIF detection method prevents any false tripping under inrush current (Fig. 6.21).

6.3.4   Comparison with Other Approaches

The proposed T-CNN framework is compared to the four other state-of-the-art ML techniques for HIF detection on the same study system: FFNN with DWT-Standard Deviation (STD) [62], FFNN with DWT-Energy [71], CNN, and transformer network. In FFNN with DWT-STD [62], the db9 mother wavelet is utilized to decompose each phase current to five detail level coefficients and one approximate level coefficient. Three input features for FFNN are extracted by taking the standard deviation of the fifth level detail coefficients [62].

FFNN with DWT-energy [71] decomposes a three-phase current signal using db6 mother wavelet and three levels of decomposition. Then, the input features for FFNN are calculated as
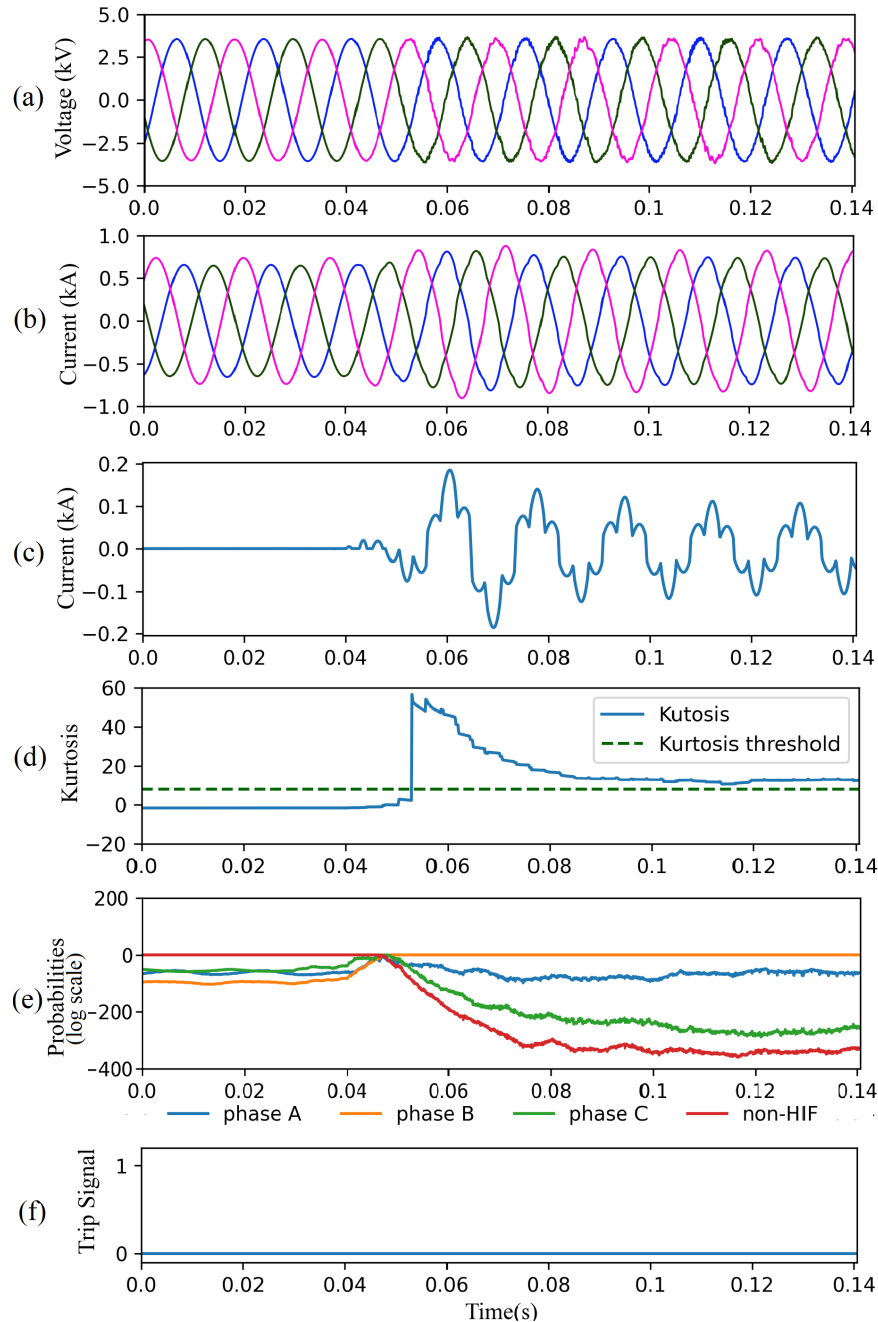
Figure 6.20: T-CNN performance under non-linear load switching: (a) non- linear load current, (b) three-phase input voltages, (c) three-phase input currents, (d) kurtosis, (e) output probabilities, and (f) trip signal.

the sum of the squares of the third level detail coefficient. The remaining two approaches, CNN and transformer network, are the variants of our solution, but instead of the T-CNN model from Fig. 5.1, they use individual CNN and transformer network, respectively.

Similar to the proposed T-CNN, the GDCV is used to select optimal hyperparameters in
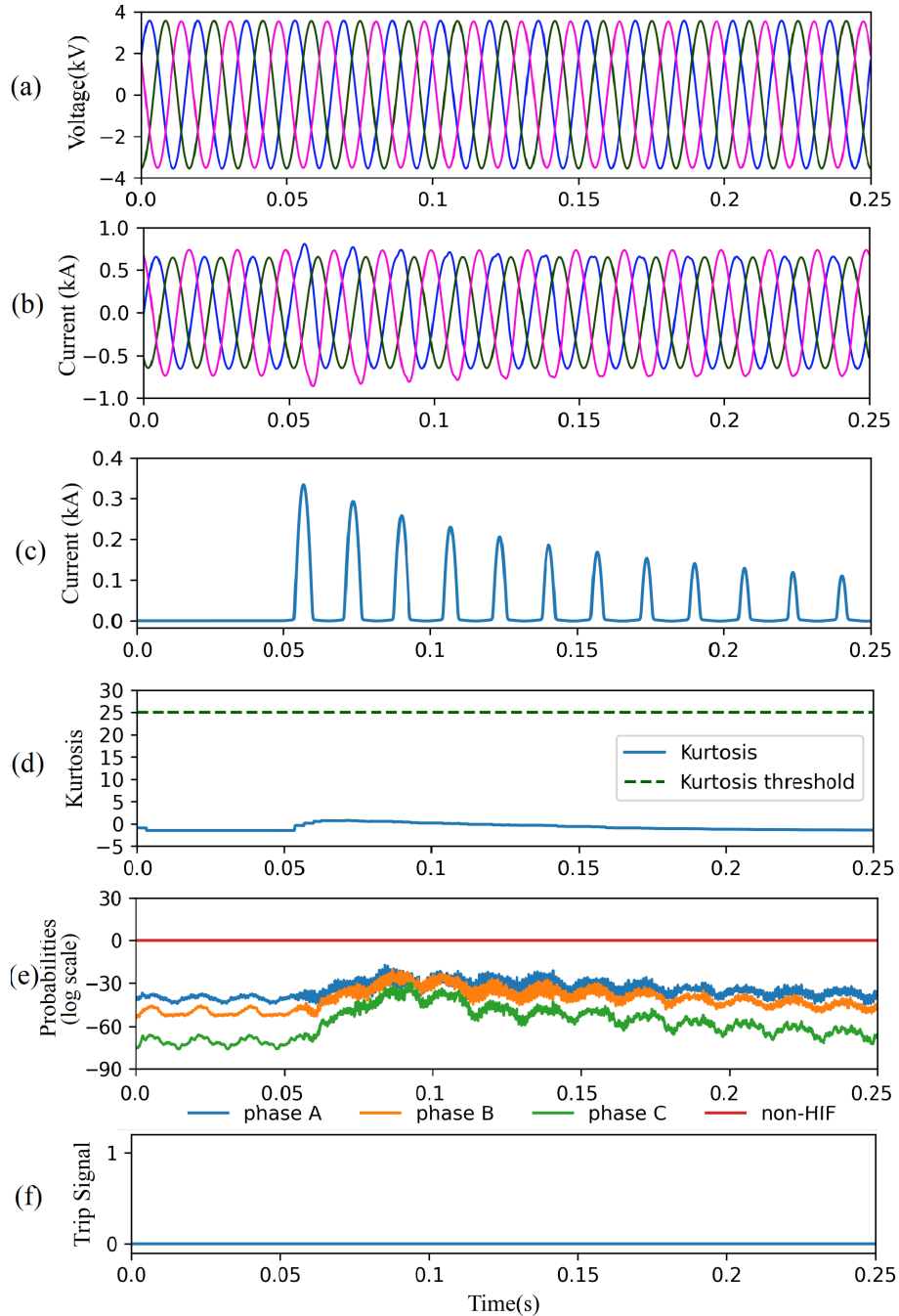
Figure 6.21: T-CNN performance under inrush currents: (a) three-phase input currents, (b) three-phase input voltages, (c) inrush current, (d) kurtosis, (e) output probabilities, and (f) trip signal.

the FFNN, the CNN, and the transformer network. The tuned SVM has kernel Radial basis function (RBF), regularization parameter (C) 1.0, and gamma scale. After hyperparameters optimization, the FFNN has three layers with 48-36-4 neurons, wherein the input and hidden

layer have a ReLU activation function. The output layer of FFNN has a softmax activation function to generate probabilities resulting in four types of classes, and the network is trained with a categorical-crossentropy loss function. The tuned hyperparameters of the CNN used for HIF classification are: one CNN layer containing 64 kernels of width size 3, one fully connected layer containing with LeakyReLU activation function, and a linear layer consisting of 4 neurons with LogSoftmax activation function. The transformer network used for comparison has one encoder layer with 2 heads, a fully connected layer with ReLU activation function, and a linear layer with 4 neurons generating output probabilities by the LogSoftmax activation function.

The results of the comparison are shown in Fig. 6.22. Both T-CNN and transformer network achieve 100% F1-score indicating that these models successfully learn the patterns and are capable of detecting HIFs and identifying the fault type. Because of the generalization and spacial-learning capability, the CNN achieved better performance than the FFNN approaches. Overall, the proposed T-CNN achieved the same F1-scores as the transformer network, but the T-CNN is more robust against noise as will be shown in subsection 6.3.6.
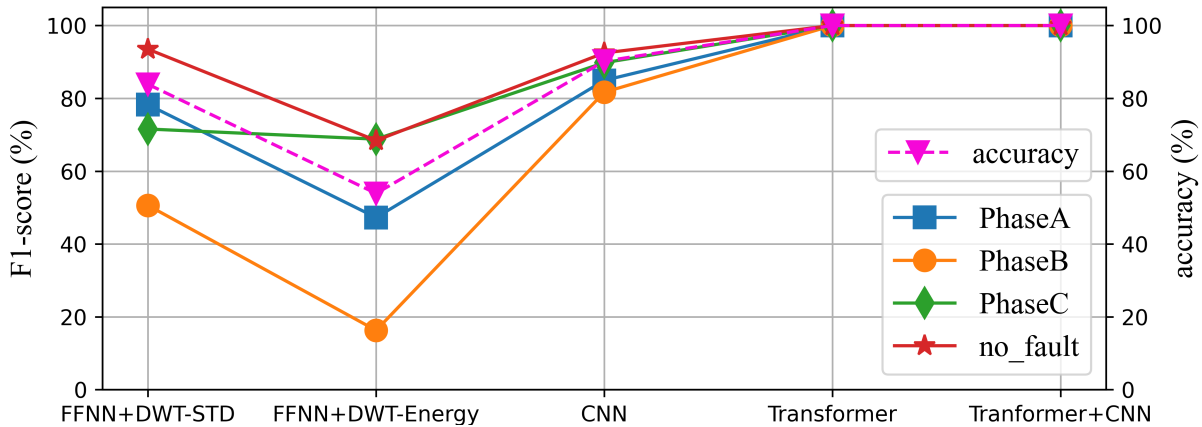
Figure 6.22: Comparison of T-CNN for HIF detection with other approaches.

6.3.5   Effect of Sampling Frequency

As the proposed T-CNN and the transformer network achieved equal performance in experiments from Subsection 6.3.4, here we compare them with respect to the reduced sampling

frequency. To examine the effect of the sampling frequency on HIF detection and classification, the frequency in a range from 1.25 kHz to 10 kHz is considered. Figures 6.23 and 6.24 show that the transformer network and the proposed T- CNN are achieving 100% F1-score for all the classes when the sampling frequency is above 2.5 kHz. For lower frequencies, the proposed T-CNN has performed slightly better: for example, with 1.25 kHz sampling what is reducing the readings in a data window from 332 to 42, the T-CNN achieves about 0.8% improvement over the transformer network. The multi-head attention in the proposed T-CNN and the transformer network enables them to learn complex HIF patterns from even a few samples in a data window enabling HIF detection even with low-frequency signals.
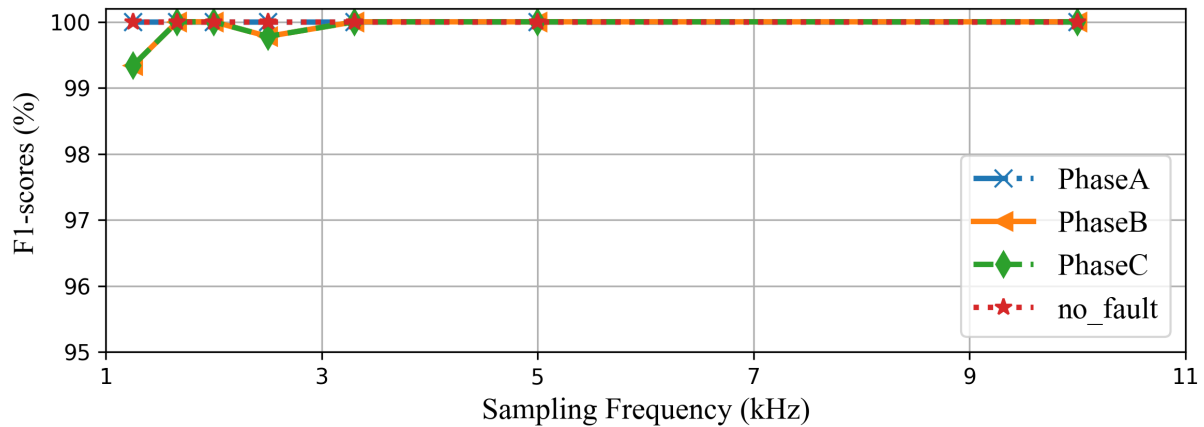
Figure 6.23: Effect of sampling frequency on performance of the transformer network.
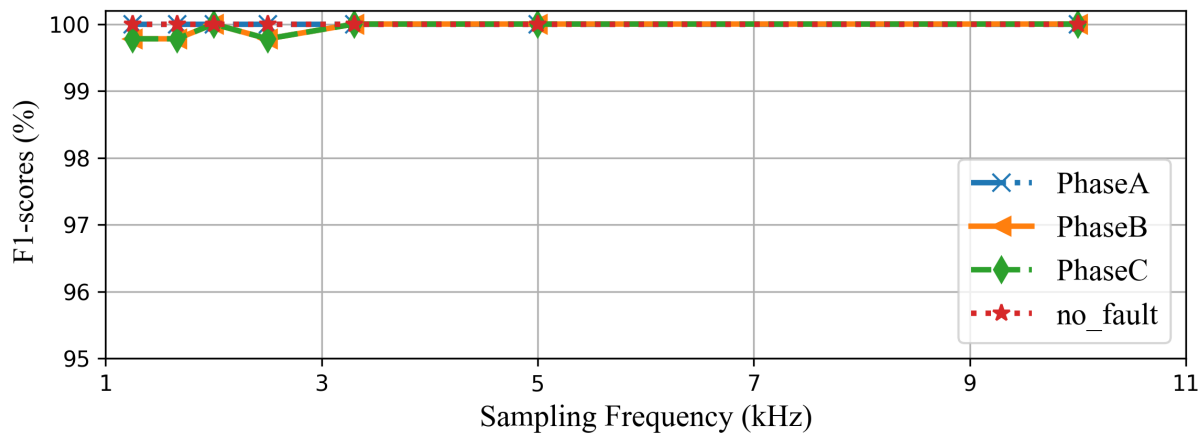
Figure 6.24: Effect of sampling frequency on performance of the proposed T-CNN.

6.3.6 Robustness of the Proposed T-CNN Against Noise

Different levels of the Signal-to-Noise Ratios (SNRs) are considered to examine robustness to noise. To include a wide range of frequency spectrum, the Gaussian white noise is added to the current signal [1]. The transformer network and the proposed T-CNN are trained with data containing noise of 40 dB SNR and all experiments in this paper are carried out with those models. Here, we examine how those models behave when the level of noise increases. As observed from Fig. 6.25, although the transformer network is immune to SNR above 50 dB, as SNR decreases to 30 dB, the F1-score for non-HIF drops drastically to 63.3% and the fault type detection F1-score goes as low as 30.3%. On the contrary, even under highly noisy conditions of 30 dB SNR, the proposed T-CNN detects the HIFs and non-HIFs scenarios with an F1-score greater than 89%. Furthermore, as illustrated in Fig. 6.26, the proposed T-CNN accurately detects all the classes through the SNR range of 33 dB to 50 dB; hence, F1-scores remains 100%. The stacking of the transformer network and CNN layers in the proposed T-CNN assisted it to learn and generalize complex patterns of corrupted input comprising of the HIF and the non-HIF data. Although the proposed T-CNN is a better solution than the transformer network in presence of noise, the CNN adds the computational complexity to the T-CNN model. Consequently, there is a trade-off between accuracy in the presence of noise and computational complexity.
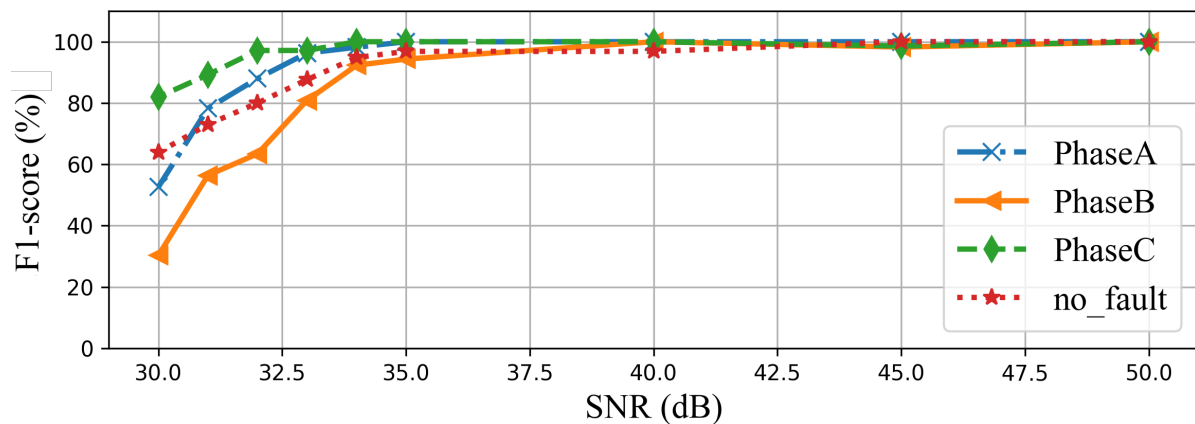
Figure 6.25: Effect of noise on performance of the transformer network

Fig. 6.27 illustrates the response of the proposed T-CNN in presence 33 dB SNR noise.

The kurtosis value is below the threshold before HIF inception at 0.05 s, Fig. 6.27 (b). After HIF inception, the phase C fault has highest output probability (Fig. 6.27 (c)), consequently, the proposed protection system issues the trip signal, Fig. 6.27 (d).
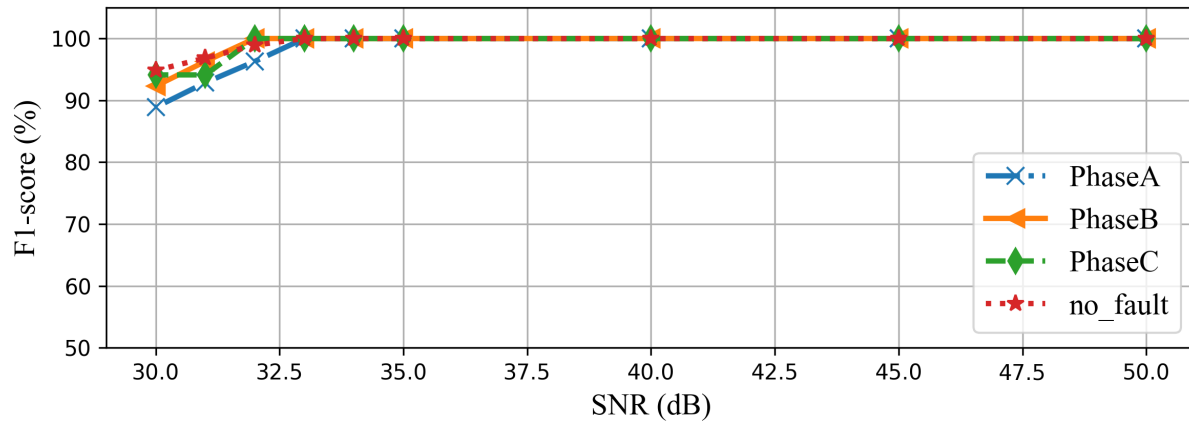
Figure 6.26: Effect of noise on performance of the T-CNM.
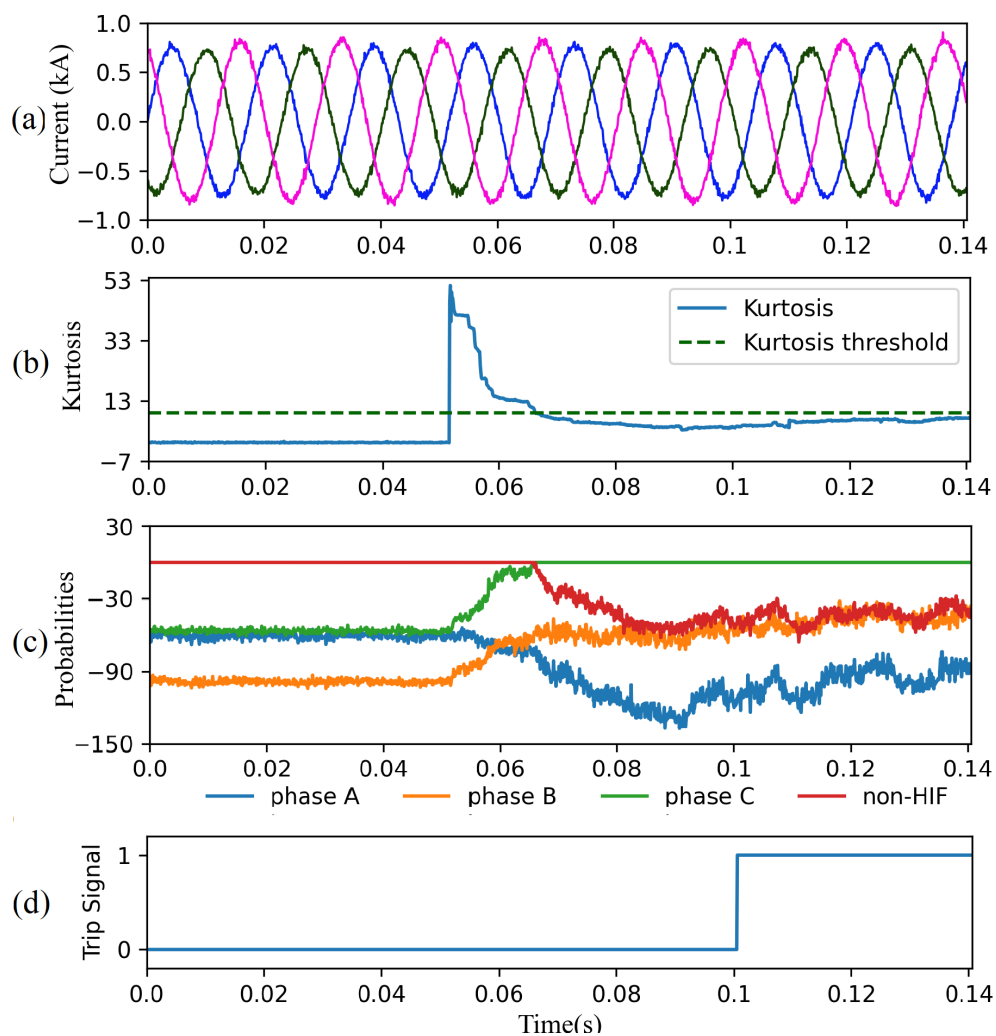
Figure 6.27: T-CNN performance under 33 dB SNR: (a) three-phase input currents, (b) kurtosis, (c) output probabilities, and (d) trip signal.

Chapter 7

Conclusion and Future Work

This chapter first summarizes CAE-HIFD and T-CNN and presents the main finding in the conclusion subsection. Finally, the possible extensions for the work are discussed in the future work section.

7.1 Conclusion

The HIFs occur in a distribution network when a conductor breaks and touches tree branches or ground surfaces. The arcing ignition caused by HIFs can result in fires (including wildfires), injuries, and even fatalities. The conventional relays cannot reliably detect the HIFs because of the low current magnitude. Recently, various machine learning-based methods have been proposed to detect and classify HIFs. However, these methods utilize supervised learning; thus, they are prone to misclassification of HIF or non-HIF scenarios that are not present in the training data. Furthermore, most of the reviewed approaches are sensitive to noise and non-fault disturbances, such as capacitor or load switching. Classification of the HIFs is another challenging task that a majority of ML-based approaches did not consider.

Consequently, this thesis proposes the CAE-HIFD, a novel deep learning-based approach for HIF detection capable of reliably discriminating HIFs from non-HIF signals including diverse disturbances. The convolutional autoencoder in CAE-HIFD applies unsupervised learn-

ing to train from the fault data only, which eliminates the need of considering all possible non-HIF scenarios for the training process. Typically, the autoencoder is trained by using normal data, and during the training, it optimizes the weights according to the MSE between the input and output and determines a threshold, which is utilized to separates the normal data and the anomalies in the testing set. However, in the proposed CAE-HIFD, the commonly used MSE is replaced by cross-correlation in order to discriminate HIFs from steady-state non-HIF scenarios. The non-fault disturbances differ from the HIFs in terms of the probability distribution of differentiated input signals. Therefore, to distinguish the transient events from HIFs, the CAE-HIFD employs kurtosis analysis.

To achieve the HIF classification, this thesis proposes the T-CNN, a novel supervised learning-based framework that utilizes the transformer network and the CNN to reliably detect and classify the HIF scenarios. The transformer network, state-of-the-art deep learning architecture, learns complex patterns in the time-series data, and the CNN further improves generalizability and enhances robustness against noise. Furthermore, as in CAE-HIFD, kurtosis analysis is utilized to provide security against false detection of non-fault disturbances as HIFs.

The results show that CAE-HIFD achieves 100% performance in terms of all five metrics of protection system performance, namely accuracy, security, dependability, safety, and sensitivity. The proposed CAE-HIFD outperforms supervised learning approaches, such as the SVM with DWT and the ANN with DWT, as well as the unsupervised GRU-based autoencoder. The CAE-HIFD performance is demonstrated on case studies including steady-state operation, close-in and remote HIFs, capacitor switching, non-linear load, transformer energization, intermittent faults, and frequency deviations. Moreover, the case studies conducted with the proposed CAE-HIFD establish the importance of the kurtosis analysis in providing security against false identification of non-fault disturbances. The studies on the effect of different noise levels demonstrate that the proposed CAE-HIFD is immune against noise even for SNR levels as high as 40 dB and provides acceptable performance for higher noise levels.

The performance evaluation studies demonstrate that the proposed T-CNN exhibits a 100% success rate in terms of classification metric, the F1-score. Case studies including close-in HIF, capacitor switching, and nonlinear load are conducted to show the performance of the proposed T-CNN. The proposed T-CNN outperforms other approaches for the HIF classification, such as the FFNN with DWT energy, the FFNN with DWT standard deviation, the CNN, and the transformer network. The robustness of the proposed T-CNN against noise and the sampling frequency is demonstrated by conducting studies with different levels of noise and sampling frequency, respectively. The results show that the proposed T-CNN is immune against an SNR of 33 dB and exhibits an acceptable success rate with 30 dB SNR. In addition, a case study on the effect of sampling frequency illustrated that the proposed T-CNN achieves 99.7% accuracy at a low sampling frequency of 1.25 kHz, which is equal to 42 samples available in a window instead of 332 samples with 10 kHZ frequency.

7.2 Future Work

The proposed CAE-HIFD accurately detected the HIFs and illustrated the advantage of unsupervised learning in successfully identifying the non-HIF scenarios without requiring the diverse non-HIF cases during the training process. Furthermore, the transformer network successfully detected and classified HIF scenarios, which confirmed that the deep learning models can be leveraged to achieve the challenging tasks of fault type detection. This work has opened the doors for further investigation. Consequently, future work will include the following:

- Evaluating the proposed models on different distribution networks: This paper utilized data obtained from comprehensive studies conducted on the IEEE 13-node test feeder that simulated a distribution network operating at 4.16 kV. Different types of power distribution networks around the world can be simulated by designing custom test feeders. Therefore, a new dataset of HIFs and non-HIFs scenarios can be generated by changing the type of test feeder. Consequently, the next step will be training and testing the

proposed approaches on different test feeders.

- Generalizing the models on multiple test feeders: Developing the model in such a way that it can learn on one distribution network and be applied to a different network will save efforts and cost to develop multiple models. The data comprising diverse non-HIF scenarios can be generated by using multiple test feeders. In the future, it would be a good idea to test our proposed frameworks on the large number of non-HIF scenarios.

- Deploying the proposed models: The CAE-HIFD and the transformer network-based HIF classification frameworks will be deployed on an embedded system to enhance the usability of the proposed approaches. In the real world power distribution network, there could be diverse HIF and non-HIF scenarios which adds risk of using simulated data for ML training and evaluation. Furthermore, the simulated data may cause overfitting of ML models used in the proposed frameworks. Consequently, evaluating the trained model on a real system as opposed to simulated will lead to new insights that can assist us to further improve the proposed framework.

- Reducing the model size: The trainable parameters in the proposed CAE are approximately 200,000, which may cause memory issues when deployed over the embedded systems with storage and computation limitations. Therefore, reducing the number of trainable parameters in the proposed models is the next step.

- Locating the HIF: It is imperative to find the node location to ensure the timely removal of the detected HIFs. Thus, the next step will be developing an ML-based framework that can reliably detect the HIF and its locations.

- Applying the proposed models on other types of faults: Apart from the HIFs, other types of faults, such as low-impedance faults and transmission line faults, should be timely detected to maintain the power system integrity. It would be interesting to utilize the CAE and the transformer network to detect the diverse faults occurring in the electrical

power system.

The proposed deep learning-based CAE-HID and T-CNN performed well for detecting and classifying the HIFs in a distribution network without the use of resource-intensive signal processing techniques. However, there is still room for improvement, as discussed in this section.

Bibliography

[1] B. Wang, J. Geng, and X. Dong. High-impedance fault detection based on nonlinear voltage–current characteristic profile identification. *IEEE Transactions on Smart Grid*, 9(4):3783–3791, 2018.

[2] S. Gautam and S. M. Brahma. Detection of high impedance fault in power distribution systems using mathematical morphology. *IEEE Transactions on Power Systems*, 28(2):1226–1234, 2013.

[3] Q. Cui, K. El-Arroudi, and Y. Weng. A feature selection method for high impedance fault detection. *IEEE Transactions on Power Delivery*, 34(3):1203–1215, 2019.

[4] Amin Ghaderi, Herbert L Ginn III, and Hossein Ali Mohammadpour. High impedance fault detection: A review. *Electric Power Systems Research*, 143:376 – 388, 2017.

[5] H. Yeh, S. Sim, and R. J. Bravo. Wavelet and denoising techniques for real-time hif detection in 12-kv distribution circuits. *IEEE Systems Journal*, 13(4):4365–4373, 2019.

[6] B. K. Chaitanya, A. Yadav, and M. Pazoki. An intelligent detection of high-impedance faults for distribution lines integrated with distributed generators. *IEEE Systems Journal*, 14(1):870–879, 2020.

[7] Richard Gonzales. Pge power lines blamed for northern california wildfires, Jun 2018.

[8] B. Wang, J. Ni, J. Geng, Y. Lu, and X. Dong. Arc flash fault detection in wind farm

collection feeders based on current waveform analysis. *Journal of Modern Power Systems and Clean Energy*, 5(2):211–219, 2017.

[9] X. Wang, J. Gao, X. Wei, G. Song, L. Wu, J. Liu, Z. Zeng, and M. Kheshti. High impedance fault detection method based on variational mode decomposition and tea-ger–kaiser energy operators for distribution network. *IEEE Transactions on Smart Grid*, 10(6):6041–6054, 2019.

[10] D. C. Yu and S. H. Khan. An adaptive high and low impedance fault detection method. *IEEE Transactions on Power Delivery*, 9(4):1812–1821, 1994.

[11] I Lee. High-impedance fault detection using third-harmonic current. final report. Technical report, Hughes Aircraft Co., Malibu, CA (USA), 1982.

[12] Wook Hyun Kwon, Gi Wen Lee, Young Moon Park, Man Chul Yoon, and Myeong Ho Yoo. High impedance fault detection utilizing incremental variance of normalized even order harmonic power. *IEEE Transactions on Power Delivery*, 6(2):557–564, 1991.

[13] José Rubens Macedo, José Wilson Resende, Carlos Augusto Bissochi Jr, Daniel Carvalho, and Fernando C Castro. Proposition of an interharmonic-based methodology for high-impedance fault detection in distribution systems. *IET Generation, Transmission & Distribution*, 9(16):2593–2601, 2015.

[14] A. E. Emanuel, D. Cyganski, J. A. Orr, S. Shiller, and E. M. Gulachenski. High impedance fault arcing on sandy soil in 15 kv distribution feeders: contributions to the evaluation of the low frequency spectrum. *IEEE Transactions on Power Delivery*, 5(2):676–686, 1990.

[15] D. I. Jeerings and J. R. Linders. Unique aspects of distribution system harmonics due to high impedance ground faults. *IEEE Transactions on Power Delivery*, 5(2):1086–1094, 1990.

[16] D. I. Jeerings and J. R. Linders. A practical protective relay for down-conductor faults. *IEEE Transactions on Power Delivery*, 6(2):565–574, 1991.

[17] A. Lazkano, J. Ruiz, E. Aramendi, and L.A. Leturiondo. A new approach to high impedance fault detection using wavelet packet analysis. In *Ninth International Conference on Harmonics and Quality of Power. Proceedings (Cat. No.00EX441)*, volume 3, pages 1005–1010 vol.3, 2000.

[18] Flavio B. Costa, B. A. Souza, N. S. D. Brito, J. A. C. B. Silva, and W. C. Santos. Real-time detection of transients induced by high-impedance faults based on the boundary wavelet transform. *IEEE Transactions on Industry Applications*, 51(6):5312–5323, 2015.

[19] Chul-Hwan Kim, Hyun Kim, Young-Hun Ko, Sung-Hyun Byun, R.K. Aggarwal, and A.T. Johns. A novel fault-detection technique of high-impedance arcing faults in transmission lines using the wavelet transform. *IEEE Transactions on Power Delivery*, 17(4):921–929, 2002.

[20] AR. Sedighi, MR. Haghifam, O. P. Malik, and MH. Ghassemian. High impedance fault detection based on wavelet transform and statistical pattern recognition. *IEEE Transactions on Power Delivery*, 20(4):2414–2421, 2005.

[21] T. M. Lai, L. A. Snider, E. Lo, and D. Sutanto. High-impedance fault detection using discrete wavelet transform and frequency range and RMS conversion. *IEEE Transactions on Power Delivery*, 20(1):397–407, 2005.

[22] A. A. Girgis, W. Chang, and E. B. Makram. Analysis of high-impedance fault generated signals using a kalman filtering approach. *IEEE Transactions on Power Delivery*, 5(4):1714–1724, 1990.

[23] S. Xue, X. Cheng, and Y. Lv. High resistance fault location of distribution network based on EEMD. In *9th International Conference on Intelligent Human-Machine Systems and Cybernetics*, volume 1, pages 322–326, 2017.

[24] A.V. Mamishev, B.D. Russell, and C.L. Benner. Analysis of high impedance faults using fractal techniques. *IEEE Transactions on Power Systems*, 11(1):435–440, 1996.

[25] Keng-Yu Lien, Shi-Lin Chen, Ching-Jung Liao, Tzong-Yih Guo, Tsair-Ming Lin, and Jer-Sheng Shen. Energy variance criterion and threshold tuning scheme for high impedance fault detection. *IEEE Transactions on Power Delivery*, 14(3):810–817, 1999.

[26] Marek Michalik, Miroslaw Lukowicz, Waldemar Rebizant, Seung-Jae Lee, and Sang-Hee Kang. New ann-based algorithms for detecting hifs in multigrounded mv networks. *IEEE Transactions on Power Delivery*, 23(1):58–66, 2008.

[27] Ibrahem Baqui, Inmaculada Zamora, Javier Mazón, and Garikoitz Buigues. High impedance fault detection methodology using wavelet transform and artificial neural networks. *Electric Power Systems Research*, 81(7):1325 – 1333, 2011.

[28] M. S. Tonelli-Neto, J. G. M. S. Decanini, A. D. P. Lotufo, and C. R. Minussi. Fuzzy based methodologies comparison for high-impedance fault diagnosis in radial distribution feeders. *International Journal of Electrical Power and Energy Systems*, 11(6):1557–1565, 2017.

[29] Zahra Moravej, Seyed Hamid Mortazavi, and Seyed Mohammad Shahrtash. Dt-cwt based event feature extraction for high impedance faults detection in distribution system. *International Transactions on Electrical Energy Systems*, 25(12):3288–3303, 2015.

[30] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y. Shin. High-impedance fault detection in the distribution network using the time-frequency-based algorithm. *IEEE Transactions on Power Delivery*, 30(3):1260–1268, 2015.

[31] Y. Sheng and S. M. Rovnyak. Decision tree-based methodology for high impedance fault detection. *IEEE Transactions on Power Delivery*, 19(2):533–536, 2004.

[32] Veerapandiyan c, Noor Izzri Abdul Wahab, Mohammad Lutfi Othman, Sanjeevikumar Padmanaban, Kavaskar Sekar, Rajeswari Ramachandran, Hashim Hizam, Arangarajan Vinayagam, and Mohammad Zohrul Islam. Lstm recurrent neural network classifier for high impedance fault detection in solar pv integrated power system. *IEEE Access*, 9:32672–32687, 2021.

[33] S. Wang and P. Dehghanian. On the use of artificial intelligence for high impedance fault detection and electrical safety. *IEEE Transactions on Industry Applications*, 56(6):7208–7216, 2020.

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[35] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[36] A. Ghosh and K. Grolinger. Edge-cloud computing for iot data analytics: Embedding intelligence in the edge with deep learning. *IEEE Transactions on Industrial Informatics*, 17(3):2191–2200, 2020.

[37] Cheng Fan, Fu Xiao, Yang Zhao, and Jiayuan Wang. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Applied Energy*, 211:1123–1135, 2018.

[38] Daniel B. Araya, Katarina Grolinger, Hany F. ElYamany, Miriam A.M. Capretz, and Girma Bitsuamlak. An ensemble learning framework for anomaly detection in building energy consumption. *Energy and Buildings*, 144:191 – 206, 2017.

[39] Shuangqi Li, Hongwen He, and Jianwei Li. Big data driven lithium-ion battery modeling method based on SDAE-ELM algorithm and data pre-processing technology. *Applied Energy*, 242:1259–1273, 2019.

[40] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International conference on artificial neural networks*, pages 52–59, 2011.

[41] M. S. Hossain, M. Al-Hammadi, and G. Muhammad. Automatic fruit classification using deep learning for industrial applications. *IEEE Transactions on Industrial Informatics*, 15(2):1027–1034, 2019.

[42] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.

[43] Luca Bondi, Silvia Lameri, David Guera, Paolo Bestagini, Edward J Delp, Stefano Tubaro, et al. Tampering detection and localization through clustering of camera-based cnn features. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 1855–1864, 2017.

[44] Syed Muhammad Anwar, Muhammad Majid, Adnan Qayyum, Muhammad Awais, Majdi Alnowami, and Muhammad Khurram Khan. Medical image analysis using convolutional neural networks: a review. *Journal of medical systems*, 42(11):1–13, 2018.

[45] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Transactions on Industrial Electronics*, 63(11):7067–7075, 2016.

[46] Kejun Wang, Xiaoxia Qi, and Hongda Liu. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Applied Energy*, 251:113315, 2019.

[47] D. Gholamiangonabadi, N. Kiselov, and K. Grolinger. Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. *IEEE Access*, 8:133982–133994, 2020.

[48] Marc Rußwurm and Marco Körner. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:421–435, 2020.

[49] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In *Advances in Neural Information Processing Systems*, pages 5243–5253, 2019.

[50] Ning Zhang. Learning adversarial transformer for symbolic music generation. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[51] A. N. Milioudis, G. T. Andreou, and D. P. Labridis. Detection and location of high impedance faults in multiconductor overhead distribution lines using power line communication devices. *IEEE Transactions on Smart Grid*, 6(2):894–902, 2015.

[52] S. H. Mortazavi, Z. Moravej, and S. M. Shahrtash. A searching based method for locating high impedance arcing fault in distribution networks. *IEEE Transactions on Power Delivery*, 34(2):438–447, 2019.

[53] Adel Soheili and Javad Sadeh. Evidential reasoning based approach to high impedance fault detection in power distribution systems. *IET Generation, Transmission & Distribution*, 11(5):1325–1336, 2017.

[54] Érica Mangueira Lima, Caio Marco dos Santos Junqueira, Núbia Silva Dantas Brito, Benemar Alencar de Souza, Rodrigo de Almeida Coelho, and Hugo Gayoso Meira Suassuna de Medeiros. High impedance fault detection method based on the short-time fourier transform. *IET Generation, Transmission & Distribution*, 12(11):2577–2584, 2018.

[55] Jiang-Yong Cheng, Shyh-Jier Huang, and Cheng-Tao Hsieh. Application of gabor–wigner transform to inspect high-impedance fault-generated signals. *International Journal of Electrical Power & Energy Systems*, 73:192–199, 2015.

[56] S. Chakraborty and S. Das. Application of smart meters in high impedance fault detection on distribution systems. *IEEE Transactions on Smart Grid*, 10(3):3465–3473, 2019.

[57] J. Carr. Detection of high impedance faults on multi-grounded primary distribution systems. *IEEE Transactions on Power Apparatus and Systems*, PAS-100(4):2008–2016, 1981.

[58] Mostafa Sarlak and S. Mohammad Shahrtash. High-impedance faulted branch identification using magnetic-field signature analysis. *IEEE Transactions on Power Delivery*, 28(1):67–74, 2013.

[59] W. Yao, X. Gao, S. Wang, and Y. Liu. Distribution high impedance fault detection using the fault signal reconstruction method. In *IEEE 3rd Conference on Energy Internet and Energy System Integration*, pages 2573–2577, 2019.

[60] Hassan Khorashadi Zadeh. An ann-based high impedance fault detection scheme: Design and implementation. *International Journal of Emerging Electric Power Systems*, 4(2), 2005.

[61] P. Routray, M. Mishra, and P.K. Rout. High impedance fault detection in radial distribution system using s-transform and neural network. In *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*, pages 545–551, 2015.

[62] Veerapandiyan Veerasamy, Noor Izzri Abdul Wahab, Rajeswari Ramachandran, Mariammal Thirumeni, Chitra Subramanian, Mohammad Lutfi Othman, and Hashim Hizam. High-impedance fault detection in medium-voltage distribution network using computational intelligence-based classifiers. *Neural Computing and Applications*, 31(12):9127–9143, 2019.

[63] Mohammad Navid Fekri, Harsh Patel, Katarina Grolinger, and Vinay Sharma. Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network. *Applied Energy*, 282:116177, 2021.

[64] A. Soheili and J. Sadeh. Evidential reasoning based approach to high impedance fault detection in power distribution systems. *IET Generation, Transmission and Distribution*, 11(5):1325–1336, 2017.

[65] Chong-Yung Chi, Chih-Chun Feng, Chii-Horng Chen, and Ching-Yung Chen. *Blind equalization and system identification: batch processing algorithms, performance and applications*. Springer Science and Business Media, 2006.

[66] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[67] W. H. Kersting. Radial distribution test feeders. *IEEE Transactions on Power Systems*, 6(3):975–985, 1991.

[68] Adel Soheili, Javad Sadeh, and Reza Bakhshi. Modified fft based high impedance fault detection technique considering distribution non-linear loads: Simulation and experimental data analysis. *International Journal of Electrical Power and Energy Systems*, 94:124 – 140, 2018.

[69] M. Kavi, Y. Mishra, and M. D. Vilathgamuwa. High-impedance fault detection and classification in power system distribution networks using morphological fault detector algorithm. *IET Generation, Transmission and Distribution*, 12(15):3699–3710, 2018.

[70] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pages 1015–1021, 2006.

[71] N Narasimhulu, DV Ashok Kumar, and M Vijay Kumar. Lwt based ann with ant lion optimizer for detection and classification of high impedance faults in distribution system. *Journal of Electrical Engineering & Technology*, 15:1631–1650, 2020.

Appendix A

Load Data

The utilized study system, the IEEE 13 node test feeder, is carrying significant levels of imbalance in a steady state. This system is chosen because it contains (i) three phases, two phases and single-phase overhead lines and underground cables and (ii) is heavily loaded and unbalanced. Furthermore, the load data in Table A.1 shows the unbalanced nature of the utilized test system.

Table A.1: IEEE 13 node test feeder Load Information.

Node	Load Model	Phase 1 (kW)	Phase 1 (kVar)	Phase 2 (kW)	Phase 2 (kVar)	Phase 3 (kW)	Phase 3 (kVar)
634	Y-PQ	160	110	120	90	120	90
645	Y-PQ	0	0	170	125	0	0
646	D-Z	0	0	230	132	0	0
652	Y-Z	128	86	0	0	0	0
671	D-PQ	385	220	385	220	385	220
675	Y-PQ	485	190	68	60	290	212
692	D-I	0	0	0	0	170	151
611	Y-I	0	0	0	0	170	80
	Total	1158	606	973	627	1135	735

Table A.2: IEEE 13 node test feeder Line Length and phasing.

Node A	Node B	Length (ft)	Phasing
632	645	500	C, B, N
632	633	500	C, A, B, N
633	634	0	Transformer
645	646	300	C, B, N
650	632	2000	B, A, C, N
684	652	800	A, N
632	671	2000	B, A, C, N
671	684	300	A, C, N
671	680	1000	B, A, C, N
671	692	0	Switch
684	611	300	C, N
692	675	500	A, B, C, N

Curriculum Vitae

Name:	Khushwant Rai
Post-Secondary Education and Degrees:	PEC University of Technology Chandigarh, India 2014 - 2018 B.Tech., Electronics and Communication Engineering University of Western Ontario London, ON 2019 - 2021 M.E.Sc., Electrical and Computer Engineering
Honours and Awards:	MITACS Research Training Award MITACS Internship Grant - ACCELERATE
Related Work Experience:	Teaching Assistant and Research Assistant The University of Western Ontario 2019 - 2021

Publications:

- K. Rai, F. Hojatpanah, F. Badrkhani Ajaei, K. Grolinger, Deep Learning for High-Impedance Fault Detection: Convolutional Autoencoders, Energies, Vol. 14, No. 12, 2021.

- K. Rai, F. Hojatpanah, F. Badrkhani Ajaei, K. Grolinger, Deep Learning for High-Impedance Fault Detection and Classification: Transformer-CNN, IEEE Transactions on Smart Grid [Submitted].