Department of Computer Science and Engineering

FAULT DETECTION AND CLASSIFICATION USING MACHINE LEARNING AND DEEP LEARNING

A report submitted is submitted as a part of Inter/ Intra Institutional Internship

BACHELOR OF ENGINEERING
in
Department of Computer Science and Engineering

by

Mihika Dhariwal
1MS21CS075

Under Supervision of

Mr. MVD Raghava
South Regional Load Dispatch Center
Bangalore

M S RAMAIAH INSTITUTE OF TECHNOLOGY
(Autonomous Institute, Affiliated to VTU)
BANGALORE-560054
www.msrit.edu

September 2023

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the **Fault Classification Using Machine Learning and Deep Learning** project submitted by **Mihika Dhariwal (1MS21CS075)** is work done by her at **South Regional Load Dispatch Center (SRLDC)** and submitted during 2023- 2024 academic year, A report submitted is submitted as a part of Inter/ Intra Institutional Internship

Mr. Pradeep Kumar D		Dr. Anita Kanavalli
Internship Coordinator	**Internal Supervisor**	**Head of the Department**

INTERNSHIP COMPLETION CERTIFICATE

ग्रिड-इंडिया
GRID-INDIA

ग्रिड कंट्रोलर ऑफ इंडिया लिमिटेड
(भारत सरकार का उद्यम)
GRID CONTROLLER OF INDIA LIMITED
(A Government of India Enterprise)
[Formerly Power System Operation Corporation Limited (POSOCO)]
दक्षिण क्षेत्रीय भार प्रेषण केन्द्र / Southern Regional Load Despatch Centre

कार्यालय : 29,रेस कोर्स क्रॉस रोड, बेंगलुरु-560009
Office : 29, Race Course Cross Road, Bengaluru - 560009
CIN : U40105DL2009GOI188682, Website : www.srldc.in, E-mail : srldc@grid-india.in, Tel.: 080-22250047/4525, Fax: 080 22268725

Ref: SRLDC/HR/Internship/2023 Date-27/12/2023

TO WHOM SO EVER IT MAY CONCERN

This is to certify that, Ms. Mihika Dhariwal, 1MS21CS075 has completed her internship for the period 1 month from 11-09-2023 to 11-10-2023 at Grid Controller of India Limited, (A Government of India Enterprise), (formerly known as POSOCO), Southern Regional Load Despatch Centre, Bengaluru in System Logistics- Information Technology Department, under the guidance of Sh. MVD Raghava, Dy. Mgr. (IT) .

During her internship, she has shown great interest and enthusiasm towards gaining knowledge and understanding the subject.

I hereby certify her work was excellent to the best of my knowledge and wish her good luck for all her future endeavours.

27/12/23

Rashmi Ranjan Das
Asst. Mgr. (HR)

सहायक प्रबंधक (मा सं)/Assistant Manager (HR)
ग्रिड कंट्रोलर ऑफ इंडिया लिमिटेड
Grid Controller of India Limited
(Formerly POSOCO)
SRLDC
29, रेस कोर्स क्रॉस रोड/29, Race Course Cross Road
बेंगलुरु-560009/Bengaluru-560009

ACKNOWLEDGMENT

First, I would like to thank **Mr. MVD Raghava, Deputy Manager at South Regional Load Dispatch Center (SRLDC)** for giving me the opportunity to do an internship in their organization, and highly indebted for guiding and making me industry-ready.

I also would like to thank all the people who worked along with me at **SRLDC** for their patience and openness. It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals.

I am thankful to **Dr. N. V. R Naidu**, Principal MSRIT for the facilities provided to accomplish this internship.

I would like to thank our Head of the Department **Dr. Anita Kanavalli** for her constructive criticism throughout my internship.

I would like to thank, Department Internship coordinator **NAME**, and for his support.

I am extremely grateful to my department staff members and friends who helped me in the successful completion of this internship.

Mihika Dhariwal

INDEX

1. ABSTRACT

As the apex body overseeing the power system in the Southern Region, the Southern Regional Load Dispatch Centre (SRLDC) in Bengaluru is entrusted with crucial responsibilities, such as monitoring system parameters, ensuring grid security, and promptly addressing tripping/disturbances through immediate remedial measures. Additionally, their role extends to daily scheduling, operational planning, and overall integration of power system operations in the region. Recognizing the significance of efficient fault detection and classification, SRLDC seeks advanced methodologies to enhance the reliability of their power grid

In alignment with these objectives, this project aims to address the crucial requirement for an efficient and accurate fault detection and classification system in transmission lines, by using various machine learning and deep learning techniques. The goal of this project is to overcome the challenges of timely fault identification and categorization, enhance the reliability and resilience of power grids, minimize downtime and optimize maintenance efforts.

The study explores both supervised and unsupervised techniques to detect and classify faults in transmission lines. Unsupervised approaches included clustering on heatmaps, as well as on time sequence graphs and detailed wavelet coefficients extracted from raw R, Y, B voltage signals. The supervised approach involves training a neural network (VGG16) on heatmaps generated from the raw signal data.

The outcomes reveal that successful classification of distinct fault signatures was achieved through supervised learning, achieving a training accuracy of approximately 88% and a testing accuracy of around 83%. In contrast the results obtained from unsupervised approaches were inconclusive, necessitating further analysis. This study thereby emphasizes the efficacy of supervised learning models in accurately identifying and categorizing various fault types, offering promising insights for enhancing fault management systems in power transmission networks.

2. LEARNING AND INTERNSHIP OBJECTIVES

Learning Objectives:

- Gain proficiency in applying machine learning algorithms for fault detection and classification in power transmission systems.
- Understand the differences and applications of supervised and unsupervised learning techniques in fault analysis.
- Learn to preprocess raw signal data (R, Y, B voltage signals) for effective utilization in machine learning models.
- Obtaining time sequence graphs and detailed coefficients from raw signal data to be used in clustering techniques.
- Develop skills in feature extraction and representation using techniques like wavelet analysis for fault diagnosis.
- Enhance knowledge of clustering algorithms and their applicability in analyzing heatmaps, time sequence graphs and detailed wavelet coefficients of voltage signals.
- Learn to implement neural network architectures (VGG16) for fault classification tasks.
- Acquire proficiency in evaluating model performance metrics and interpreting results in the context of fault identification accuracy.

Internship Objectives:

- Apply machine learning techniques to analyze real-world transmission line data to detect and classify faults accurately.
- Contribute to enhancing the company's fault management systems by implementing advanced algorithms for accurate fault identification.
- Innovate and experiment with supervised and unsupervised learning approaches to address challenges in fault analysis, ensuring robustness and efficiency.
- Evaluate the applicability of machine learning models within the company's operational framework for potential integration into real-time systems.
- Document and present findings, including model performance evaluations and insights gained from the study.
- Collaborate with team members to discuss methodologies, results, and potential improvements in fault management systems for power transmission networks.

3. OVERVIEW OF INTERNSHIP ACTIVITY

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
1st WEEK	11/09/23	Monday	Introduction to grid system, transmission lines, faults, and types of faults
	12/09/23	Tuesday	Introduction to python libraries: NumPy, Pandas, Matplotlib
	13/09/23	Wednesday	Introduction to machine learning
	14/09/23	Thursday	Core ML Algorithms: Linear Regression, Classification, Clustering
	15/09/23	Friday	Exercise: House Price Prediction using Regression model
	16/09/23	Saturday	Introduction to deep learning and neural networks

	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
2nd WEEK	18/09/23	Monday	Introduction to Feed Forward Neural Networks and Convolutional Neural Networks
	20/09/23	Wednesday	Working with ML Libraries: TensorFlow, Keras, Scikit Learn
	21/09/23	Thursday	Understanding the problem statement, its requirements and formulating the program flow for the problem statement
	22/09/23	Friday	Clustering heatmaps generated from raw signal data
	23/09/23	Saturday	Continued our work

3rd WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	25/09/23	Monday	Continued our work: Tuned our learning parameters
	26/09/23	Tuesday	Generation of time sequence graphs from raw signal data
	27/09/23	Wednesday	Clustering the time sequence graphs
	28/09/23	Thursday	Analyzing the results, tuning learning parameters
	29/09/23	Friday	Continued our work

4th WEEK	DATE	DAY	NAME OF THE TOPIC/MODULE COMPLETED
	30/09/23	Saturday	Introduction to wavelet transforms and wavelet coefficients
	3/10/23	Tuesday	Extracting detailed wavelet coefficients from raw signal data and clustering based on extracted coefficients
	4/10/23	Wednesday	Clustering based on norm values of extracted detailed coefficients
	5/10/23	Thursday	Supervised Learning using VGG16 like neural network
	6/10/23	Friday	Tuning learning parameters, regularization to minimize overfitting, analyzing results obtained

4. INTRODUCTION

Detecting and classifying faults within electrical transmission lines holds paramount significance due to its multifaceted impact on safety, reliability, and operational efficiency. Faults, abnormalities or disruptions in the electrical system can stem from equipment failure, environmental conditions, human error or unseen circumstances. These faults manifest in various forms, such as line-to-ground (RG, YG, BG), line-to-line (RY, YB, RB), double line to ground (RYG, YBG, RBG), triple line (RYB), and triple line to ground (RYBG) faults.

The timely identification of faults holds immense importance as it helps avert potential hazards to personnel and the public, minimizing accidents, fires, and electrical risks. Swift fault detection also ensures minimal disruption to power supply, safeguarding system reliability and preventing widespread outages that could negatively impact industries, businesses, and essential services. Accurate fault classification enables targeted maintenance strategies, preserving infrastructure integrity and optimizing grid operations.

The provided raw data comprises real-time measurements of the R, Y, B phase voltage signals sampled over a 40-millisecond time interval. This dataset represents the electrical characteristics of the transmission lines during operation, crucial for analyzing if a fault has occurred in the transmission line.

The scope of this internship encompasses the application of supervised and unsupervised machine learning methodologies to significantly advance fault detection and classification within transmission lines.

The first method involves clustering heatmaps generated from R, Y, B phase voltage signals. Initially, the VGG16 neural network was utilized to extract characteristic features unique to each type of fault from these heatmaps, resulting in a distinct feature vector for each heatmap. To streamline the analysis and mitigate computational complexity, Principal Component Analysis (PCA) was employed to reduce the dimensionality of these feature vectors. The reduced feature vectors were subsequently inputted into the K-means clustering algorithm to group similar data points into clusters. The determination of the optimal number of clusters (k) was achieved by analyzing an elbow graph, and the heatmaps for each cluster were visually displayed.

The second method involves clustering time sequence graphs obtained from the raw signal data. The idea behind this approach is that VGG model, being a deep convolutional neural network architecture, is better suited at capturing critical signal characteristics—such as edges, voltage sags, and fault signatures—compared to heatmaps, aligning well with the nature of time sequence graphs. After extracting relevant features and subsequent PCA dimensionality reduction, K-means clustering was applied to group similar signals based on their feature representations.

Wavelet transforms plays a crucial role in fault detection and classification by decomposing signals into various frequency components across time. This method generates coefficients that represent signal characteristics at different scales. In the context of fault analysis, these coefficients provide a comprehensive view of the transient and dynamic features present in the electrical signals, allowing for a detailed examination of signal variations associated with different fault types. The unique advantage of wavelet analysis lies in its ability to capture both high-frequency (such as sudden voltage spikes) and low-frequency changes (such as gradual voltage fluctuations), enabling the identification of fault signatures that might be obscured in the raw signal data. Thereby, wavelet analysis enhances feature extraction for precise fault detection and classification.

In the third method, discrete wavelet transform was performed on each phase voltage of the time series data. Different wavelet types (Haar, Db1) with different decomposition levels (1,2) were used. The intuition behind this approach was that wavelet coefficients capture certain time-frequency characteristics unique to each type of fault. The set of detailed coefficients obtained for each phase were concatenated into a single feature vector and clustering was performed. In an alternative approach, the norm of detailed coefficients of each phase voltage was calculated, concatenated into a single feature vector, and clustering was performed.

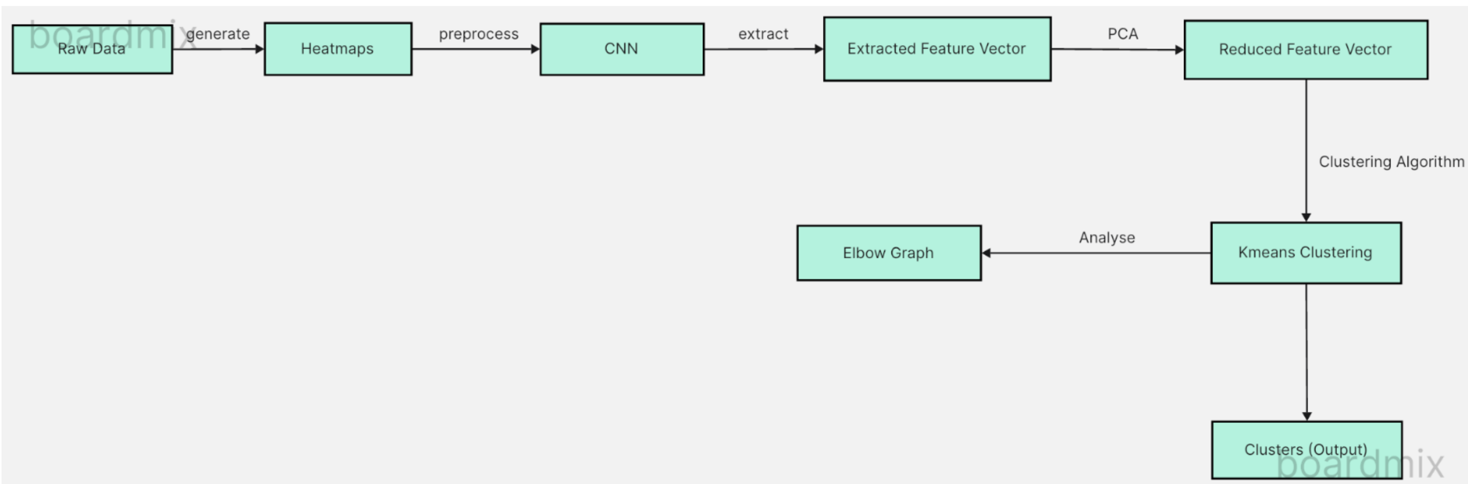
In the final approach, I employed supervised learning using the heatmaps generated from raw signal data. We designed a convolutional neural network based of the VGG16 model, and trained it on the heatmaps to identify 12 different types of faults. (RG, YG, BG, RY, YB, RB, RYG, YBG, RBG, RYB, RYBG, Faultless)

5. INTERNSHIP WORK

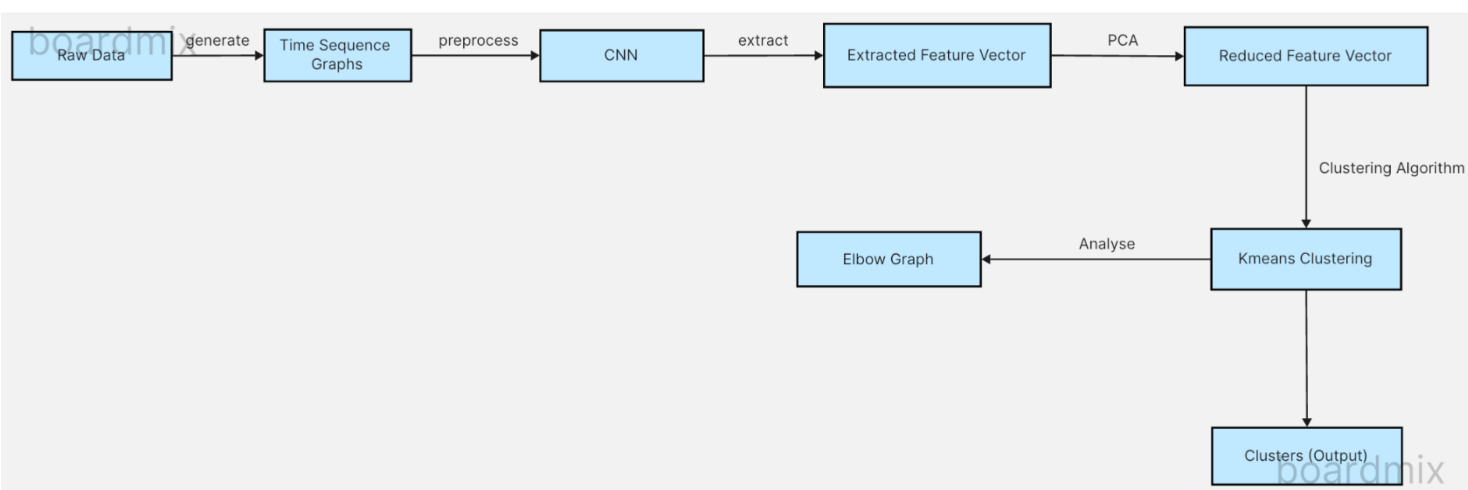
5.1 Architecture Design

Here is a high-level design of each of the methodologies used

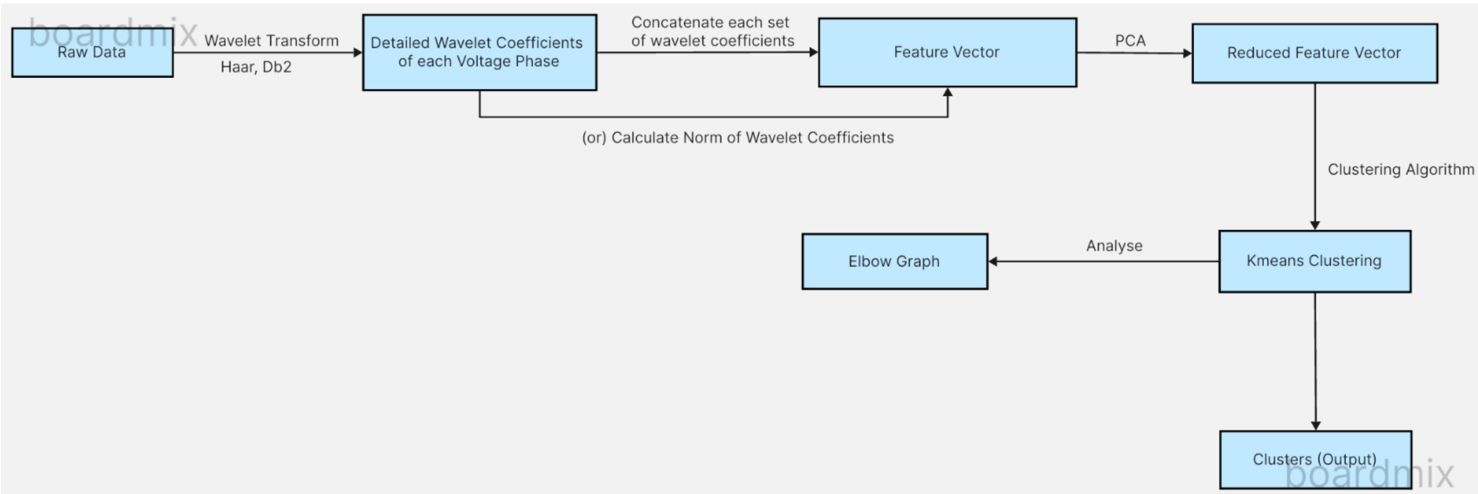
Method 1:
Classifying Faults by Clustering Heatmaps



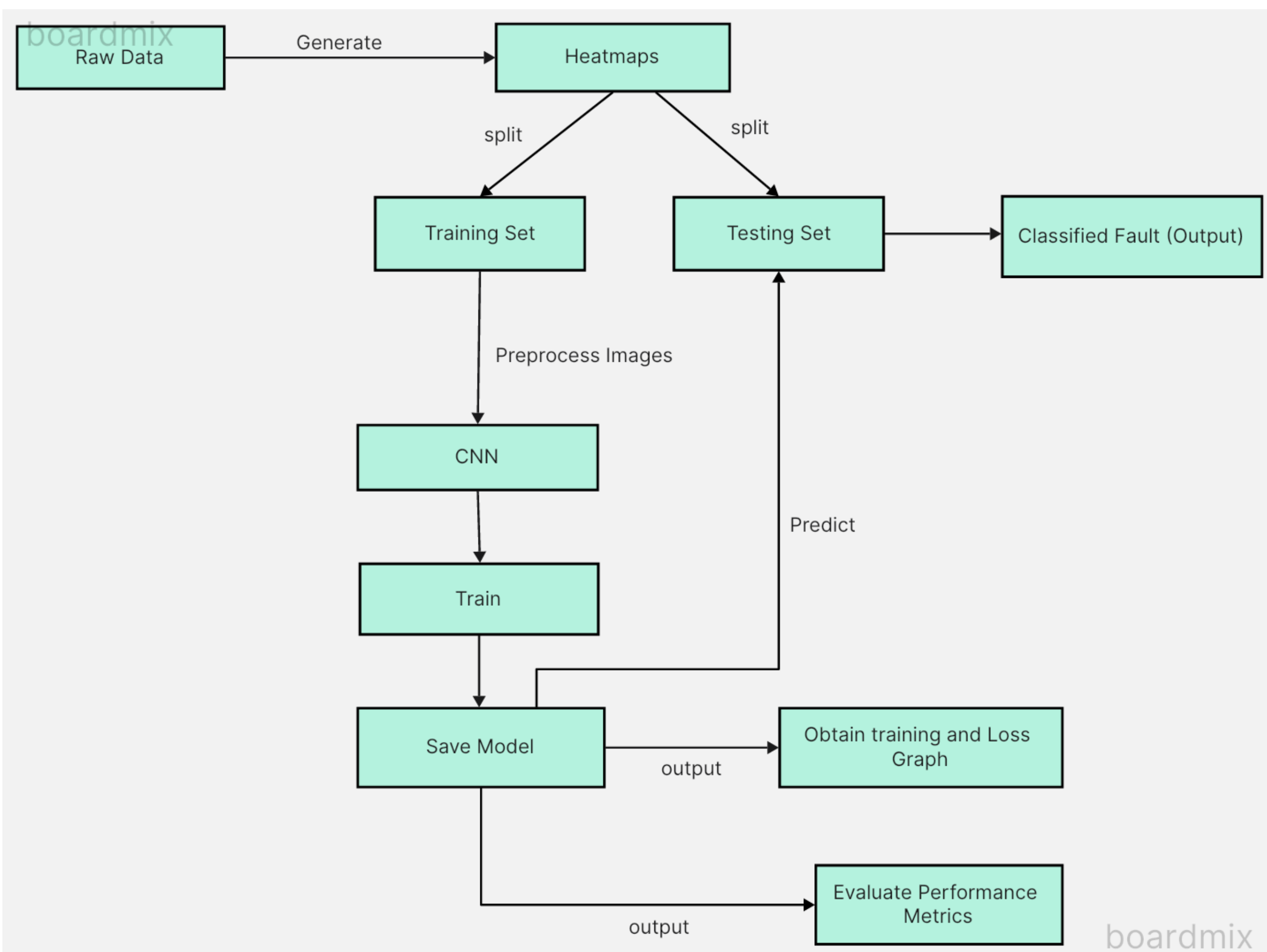
Method 2:
Classifying Faults by Clustering Time Sequence Graphs



Method 3:
Classifying Faults by Clustering Wavelet Coefficients



Method 4:
Supervised Learning on Heatmaps using CNNs



Architecture of CNN Model used in our Solution

The convolutional neural network (CNN) architecture designed in my solution exhibits a hierarchical structure composed of multiple convolutional and pooling layers, culminating in densely connected layers for classification. The architecture follows a classic design pattern for image classification tasks, progressively extracting and abstracting features from the input data.

Convolutional Layer and Max-Pooling:

- The initial convolutional layer employs 32 filters with a (1,3) kernel to process images of dimensions (50, 787, 3), focusing on capturing spatial features along the vertical dimension.
- Subsequent max-pooling layers with a pooling size of (1,2) and strides of (1,2) contribute to spatial downsampling, enhancing the network's ability to learn hierarchical features.

Convolutional Blocks:

- The network iteratively applies multiple convolutional blocks, and with each block, the number of filters is doubled, enabling the extraction of increasingly complex patterns.

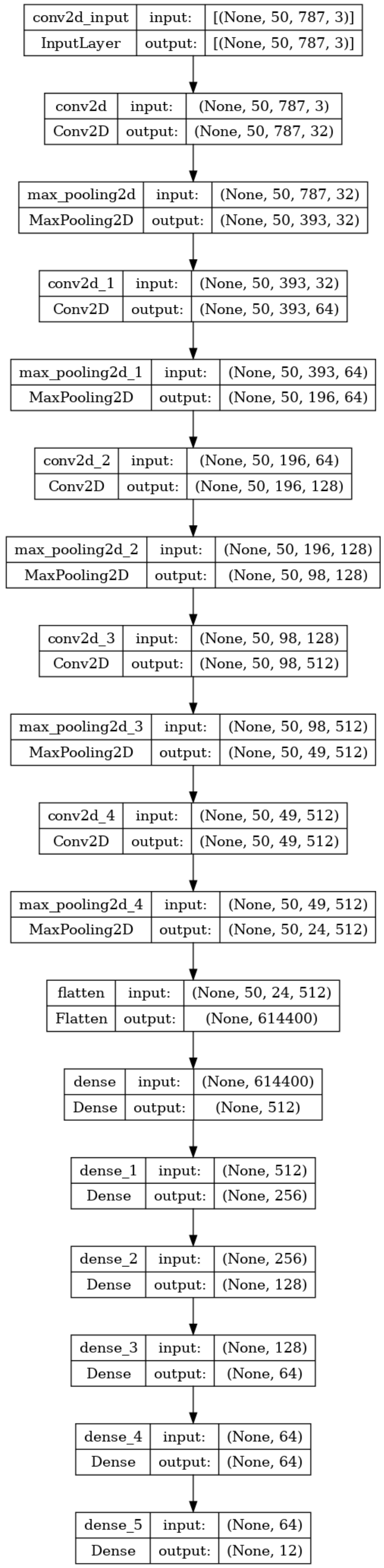
Deeper Convolutional Layers:

- Deeper layers in the network use larger kernel sizes such as (2,3) and (3,3) to capture broader spatial relationships within the data.
- Convolutional layers are followed by max-pooling operations to maintain a manageable spatial resolution.

Flattening and Densely Connected Layers:

- The architecture concludes with a flattening layer, transitioning from the spatial hierarchy to a flattened vector representation.
- This flattened data is then fed into a sequence of densely connected layers.
- The dense layers gradually reduce in size from 512 to 12 neurons, introducing non-linearity to the network.
- This facilitates the extraction of high-level features and ultimately leads to the final classification into 12 different classes using a sigmoid activation function.

Diagramatic Representation of my CNN Model:



5.2 Tools and Technology Used

Python

Python serves as the primary programming language for developing and implementing this project. Its versatility and extensive ecosystem of libraries make it an ideal choice for machine learning projects. In this project, Python played a pivotal role as the programming language offering a range of tools and libraries essential for machine learning implementation. Pandas and NumPy were utilized for efficient data manipulation and preprocessing, allowing for seamless integration with machine learning workflows.

TensorFlow

TensorFlow, an open-source machine learning framework developed by Google, played a central role in building and training the machine learning model. TensorFlow's strong support for deep learning allowed for the implementation of Convolutional Neural Networks, such as the VGG16-like deep learning model used in this project. TensorFlow's flexibility enabled customization of the neural network architecture.

Keras

Keras, an open-source neural network library, served as a high-level API for building and training neural networks. It is seamlessly integrated with TensorFlow in our project. Keras provides an abstraction layer, simplifying the process of defining and training complex neural network models.

Scikit-learn

Scikit-learn was used to implement the PCA module and KMeans algorithm. The KMeans implementation allowed us to identify distinct clusters within the data, revealing patterns associated with fault occurrences . Additionally, scikit-learn's PCA module was instrumental in reducing the dimensionality of our feature space. PCA was applied to streamline the complexity of the data, retaining the most informative components.

OpenCV

OpenCV serves as a pivotal tool in computer vision projects, offering a comprehensive set of functionalities for image processing. In my project, it is used to perform tasks such as reading, resizing, and transforming images. Its broad applications make it a cornerstone in computer vision projects, allowing for efficient manipulation and enhancement of visual data to meet the requirements of diverse applications in image classification.

Wavelet Transform

Wavelet coefficients played a role in extracting relevant features for fault detection. PyWavelets, a Python library for wavelet transform, is used to implement the wavelet-based feature extraction process.

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are highly suitable for image-based data due to their innate ability to capture spatial hierarchies. Their specialized convolutional layers excel in feature extraction by automatically learning and emphasizing relevant patterns, enabling robust image classification. In my project, the a VGG16-like architecture was employed, which featured a sequence of convolutional and max-pooling layers for hierarchical feature extraction. The model was built using the Sequential API from the Keras library. VGG16, short for Visual Geometry Group 16, is a deep convolutional neural network architecture renowned for its simplicity and effectiveness in image classification tasks. Introduced by the Visual Geometry Group at Oxford University, VGG16 consists of 16 weight layers, including 13 convolutional layers and 3 fully connected layers.

5.3 Discussion of Results and Inferences

Clustering based on heatmaps

The clustering results, where only a subset of similar heatmaps were grouped into clusters while others were not, suggest a potential heterogeneity in the fault patterns present in the data. The variation in clustering outcomes may indicate different fault subtypes or diverse manifestations of faults within the dataset. It was be beneficial to explore the dissimilar clusters more deeply, analyzing the unique features that contribute to their separation, and investigate if there are specific fault scenarios or conditions that result in distinctive heatmap patterns.

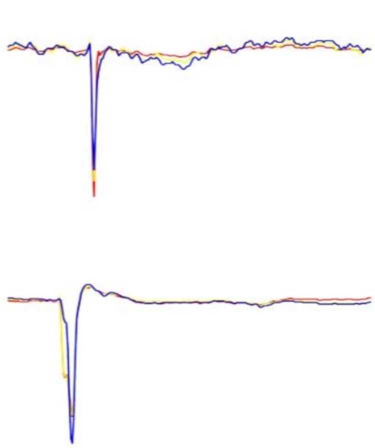


Figure 1: Cluster 1

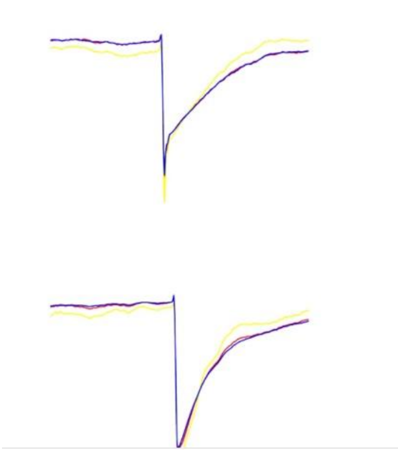


Figure 2: Cluster 2

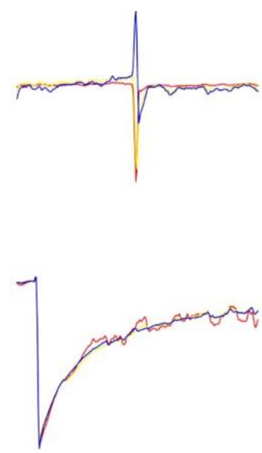


Figure 3: Cluster 3 (Incorrect)

Optimal Number of Clusters (k)

The optimal number of clusters (k) was determined by analyzing an elbow graph, which plotted the variance explained against the number of clusters. The "elbow" point represents the optimal trade-off between model complexity and performance.

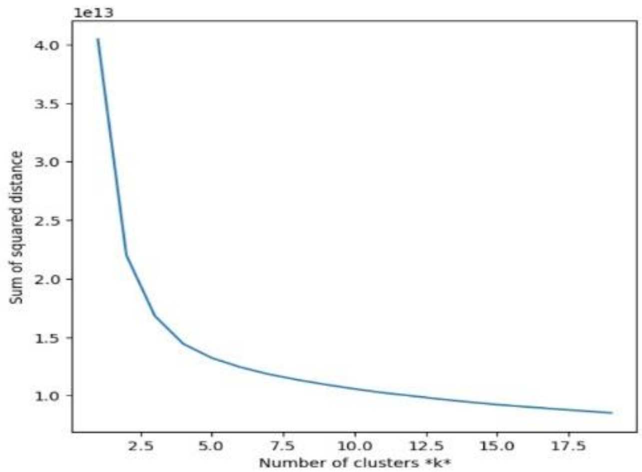


Figure 4: Elbow Graph

Clustering based on time-sequence graph

Results from the clustering based on time sequence graphs using the VGG-like model for feature extraction and K-means clustering revealed a notable success in grouping signals associated with tripping events into a distinct cluster. However, the effectiveness of the clustering approach varied for other types of fault events, suggesting potential challenges in capturing the diversity of fault signatures within the dataset. The hypothesis that the VGG-like model, designed for identifying patterns in sequential data, would outperform heatmap-based clustering in capturing crucial signal characteristics such as edges, voltage sags, and fault signatures held true for tripping events. However, further refinement of the methodology or exploration of alternative models were necessary to improve the clustering performance for other fault types, addressing the unique challenges posed by different fault scenarios within the time sequence graph data.

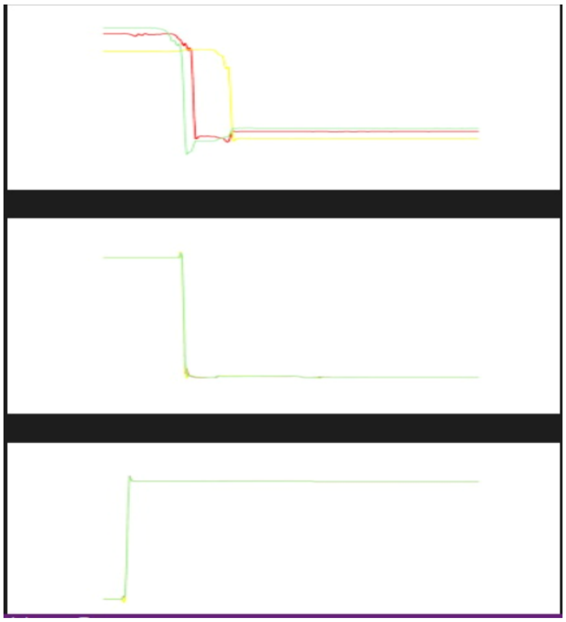


Figure 5: Cluster of Tripping Event

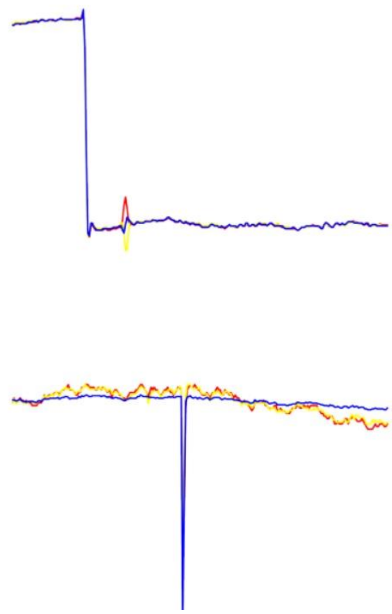


Figure 6: Incorrect Cluster

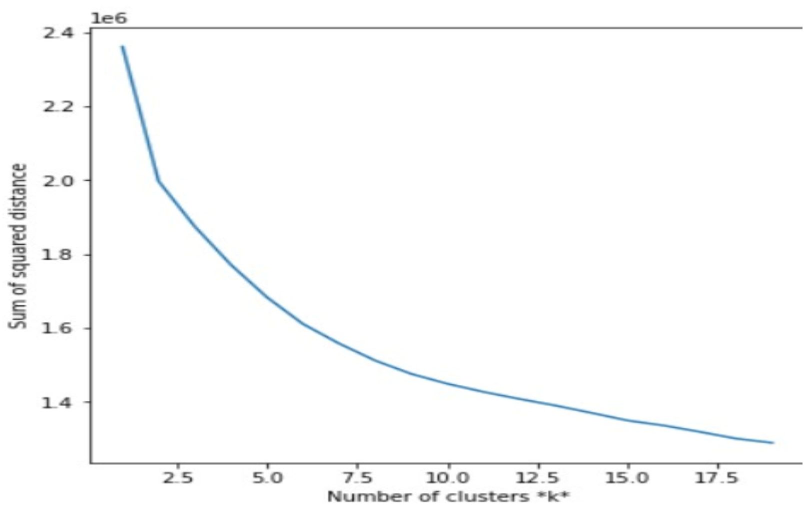


Figure 7: Elbow Graph for Time Sequence Graphs

Clustering based on wavelet coefficients

The results of clustering based on wavelet coefficients in our project yielded unexpected outcomes. Despite the anticipation of capturing unique time-frequency characteristics related to different types of faults, the clustering did not align with fault types as initially anticipated. The concatenation of detailed coefficients from each phase voltage, as well as an alternative approach involving the norm of detailed coefficients, did not lead to distinct fault-based clusters.

This unexpected outcome suggests that the detailed coefficients or their norms might be capturing a different property of the time series data that is currently unknown. The clustering may be influenced by features or patterns in the data unrelated to the types of faults we were aiming to detect. It's crucial to investigate and analyze the resulting clusters to identify the specific time-frequency characteristics or hidden patterns that are contributing to the clustering results.

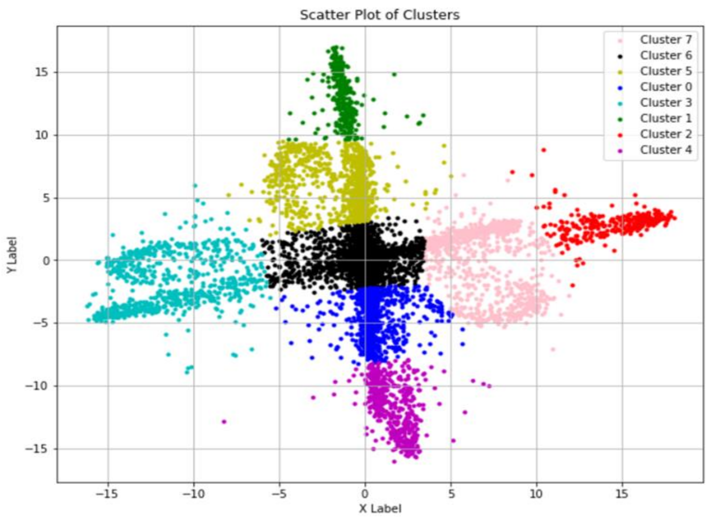


Figure 8: Wavelet: Haar, level: 2

All Detailed Coefficients considered to generate feature vector

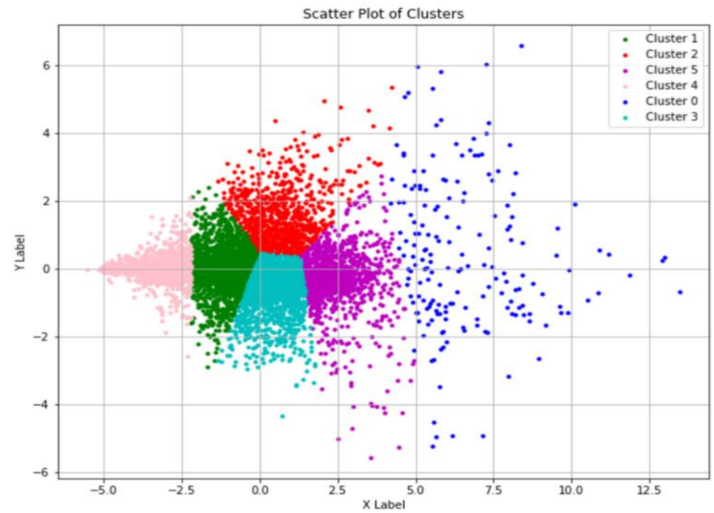


Figure 9: Wavelet: Db1, level: 1

Norm of Detailed Coefficients considered to generate feature vector

Supervised learning

The results from the supervised learning approach utilizing a customized VGG16 model for fault classification surpassed the efficiency of other methods. The model achieved a training accuracy of 88% and a testing accuracy of 83%, indicating a notable ability to generalize to unseen data. The superior efficiency of the supervised learning model suggests that the combination of VGG16 architecture and heatmap image features provided a powerful framework for fault classification. This approach not only demonstrated high accuracy in distinguishing between the 12 different fault types but also showcased a level of generalization that outperformed other methodologies employed in the project.

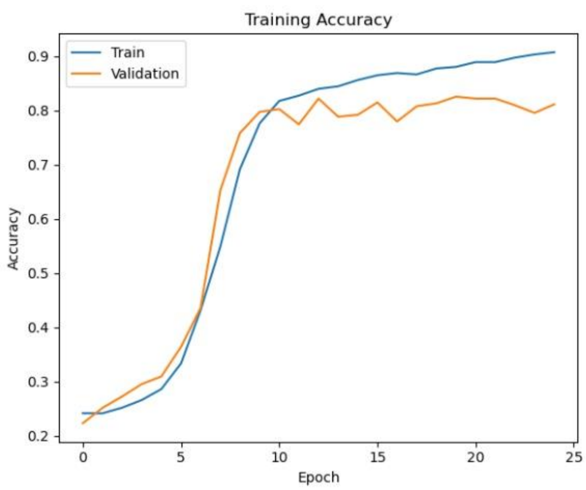


Figure 10: Training Accuracy

Figure 11: Training Loss

6. CONCLUSION

In conclusion, the integration of Machine Learning (ML) and Deep Learning (DL) techniques in fault detection and classification for transmission lines represents a significant advancement in the field of power systems. Throughout this report, we have explored the various methodologies, algorithms, and models employed to enhance the reliability and efficiency of fault detection in transmission lines.

One noteworthy approach involved the application of K-means clustering to classify heatmaps, and time sequence graphs generated from R, Y, B phase voltage signals. While the clustering results provided valuable insights into fault patterns, it is important to note that they were not entirely accurate. Further analysis and refinement of the clustering methodology may be required to enhance its precision. In contrast, a supervised learning approach utilizing a Convolutional Neural Network based on the VGG16 model demonstrated a notable accuracy of 85% in classifying 12 different fault types.

As we advance into the era of smart grids and intelligent power systems, the implementation of ML and DL technologies offers a promising avenue to enhance the overall resilience and reliability of transmission lines. These technologies not only enable swift and accurate fault detection but also facilitate the classification of faults, aiding in the prompt and precise deployment of maintenance strategies.

Despite the remarkable progress made in this domain, it is essential to acknowledge the ongoing challenges and potential areas for improvement. Issues such as data quality and the need for large labeled datasets remain prominent concerns. Future research efforts should focus on addressing these challenges to further refine and optimize the performance of ML and DL models for fault detection and classification in transmission lines.

7. BIBLIOGRAPHY

Research Papers:

[1] Jamil, M., Sharma, S.K. & Singh, R. Fault detection and classification in electrical power transmission system using artificial neural network. *SpringerPlus* **4**, 334 (2015).

[2] D. Paul and S. K. Mohanty, "Fault Classification in Transmission Lines Using Wavelet and CNN," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-6, doi: 10.1109/I2CT45611.2019.9033687.

[3] R. S. Perez, A. C. Oviedo, J. Camarillo-Peñaranda and G. Ramos, "A novel fault classification method using Wavelet transform and Artificial Neural Networks," 2016 17th International Conference on Harmonics and Quality of Power (ICHQP), Belo Horizonte, Brazil, 2016, pp. 448-453, doi: 10.1109/ICHQP.2016.7783476.

[4] H. Kumar, M. Shafiq, G. A. Hussain, L. Kumpulainen and K. Kauhaniemi, "Classification of PD Faults Using Features Extraction and K-Means Clustering Techniques," 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), The Hague, Netherlands, 2020, pp. 919-923, doi: 10.1109/ISGT-Europe47291.2020.9248984.

Websites and Blogs:

https://www.tensorflow.org/

https://keras.io/guides/

https://towardsdatascience.com/how-to-cluster-images-based-on-visual-similarity-cd6e7209fe34

https://franky07724-57962.medium.com/using-keras-pre-trained-models-for-feature-extraction-in-image-clustering-a142c6cdf5b1