Machine Learning-Based Electrical Fault Classification: A Comparative Analysis of Logistic Regression and Random Forest

Ashok Ganga[1*], Kanaka Raju Kalla[2], Tamminaina Lokanadha Rao[3], Sontyana Sravya[4], Panchireddi Lakshmi[5], Mandula Jahnavi[6], Madhurakavi Sravani Pravallika[7]

[1,2,3]Assistant Professor, EEE Department, Aditya Institute of Technology And Management, Tekkali.

[4,5,6,7] U G Student, EEE Department, Aditya Institute of Technology And Management, Tekkali.

Abstract:
Fault classification plays a vital role in electrical systems, ensuring reliability and minimizing downtime. This study assesses how well two machine learning models—Random Forest and Logistic Regression—distinguish between fault and no-fault situations. To assess their performance, we analyzed accuracy, precision, recall, and F1-score, both before and after hyper parameter tuning. The results highlight a stark contrast between the two models. Logistic Regression struggled with classification, achieving a peak accuracy of just 62.31% even after tuning. In contrast, Random Forest demonstrated significantly superior performance, reaching an impressive 99.91% accuracy without tuning and a flawless 100% accuracy with optimized hyper parameters. These findings underscore the effectiveness of Random Forest for fault classification and emphasize the critical role of hyper parameter tuning in maximizing model performance.

Key Words: Electrical fault detection, machine learning, fault classification, Random Forest, Logistic Regression, hyper parameter tuning.

1. Introduction:

Maintaining the dependability and safety of electrical systems depends on the detection and classification of electrical faults because it enables possible problems to be addressed before they become serious failures or hazards[1]. Effective fault detection methods contribute to minimizing downtime, lowering maintenance expenses, and optimizing overall system performance [2]. Over time, various techniques have been developed for accurate fault identification, ranging from conventional relay-based systems to cutting-edge approaches that leverage artificial intelligence and machine learning [3]. These cutting-edge technologies facilitate predictive maintenance plans in addition to improving the speed and accuracy of issue detection, hence enhancing system efficiency and resilience [4]. Fault diagnosis is important because it can stop catastrophic failures that could result in costly repairs, severe injuries, or even fatalities[5]. By implementing robust fault detection systems, organizations can maintain uninterrupted operations while preserving the integrity of their electrical infrastructure [6]. This proactive strategy not only reduces downtime but also promotes a safer working environment, ultimately boosting productivity and lowering operational expenses. Electrical faults can take various forms, including short circuits, overloads, ground faults, and open circuits. Each type presents distinct challenges and requires specialized detection methods to ensure timely intervention and resolution [7]. A thorough understanding of these faults how they behave and what risks they pose is essential for developing effective detection strategies tailored to specific electrical systems [8]. The primary objectives of this paper include a comprehensive review of different fault detection techniques, an analysis of the challenges in accurately identifying and classifying faults, and an exploration of how these advancements contribute to improving overall system reliability and safety. By delving into these aspects, this study aims to establish a framework that organizations can leverage to enhance their fault detection capabilities, minimize downtime, and reduce maintenance costs.

2. Literature Review:

Table 1 presents a comparative analysis of multiple research studies focused on fault detection methods in electrical power systems. It outlines the methodologies employed, summarizes key findings, and identifies existing research gaps. This comparison provides insightful information about how well different machine learning and deep learning methods classify and detect faults.

Table 1. Comparative Analysis of Fault Detection Methods in Electrical Power Systems

Author	Method	Results	Research Gap
[9]	Multiple Classifier System (MCS) approaches, META-DES for fault detection in electrical power systems.	MCS approaches outperform single models; META-DES shows resilience to noise.	Monolithic models struggle with adaptability, imbalance, and noise handling. Further exploration of ensemble methods needed.
[10]	Deep learning (LSTM) based fault detection, classification, and fault location estimation.	LSTM-based method demonstrated high accuracy in fault detection and precise fault location estimation.	Modern ML approaches in fault analysis are still developing. Performance under diverse conditions needs further study.
[11]	Random Forest classifier applied to a power grid model in MATLAB-SIMULINK.	Random Forest achieved 100% healthy operation detection and 96% fault detection accuracy.	Study lacks consideration of environmental factors and scalability to larger grid systems.
[12]	Multiple ML models (SVM, Decision Tree, KNN, Random Forest) for fault detection.	All models achieved over 99% accuracy, with SVM reaching 99.6% accuracy.	No discussion on practical application and integration into existing power systems.
[13]	Random Forest model for detecting physical and electrical faults in PV array systems.	Random Forest achieved 98.6% detection accuracy and 94.2% classification accuracy.	Synthetic training data raises concerns about real-world applicability and robustness.
[14]	ML models (SVM, Random Forest, Decision Tree, XGB,	Ensemble techniques (Random Forest, XGB, and Decision Tree)	Implementation challenges, data quality, and model interpretability not addressed.

	Logistic Regression) for detecting electrical irregularities in transmission lines.	improved fault detection accuracy.	
[15]	Support Vector Machine (SVM) classifier trained on real-time electrical system data.	SVM classifier effectively detects faults using real-time sensor data.	Does not explore integration of other ML methods to enhance robustness and accuracy.
[16]	Deep learning (LSTM) model for early fault detection in Electrical Power Transmission Networks (EPTN).	LSTM model achieved 99.65% accuracy, outperforming NN and CNN models.	Challenges of real-time implementation and integration into existing systems not discussed.

This comparative analysis demonstrates that while machine learning and deep learning models generally achieve high accuracy in fault detection, several challenges remain. Key research gaps include improving adaptability to real-world conditions, enhancing model robustness against environmental variations, and ensuring seamless integration with existing power system management frameworks. Addressing these issues in future studies will be essential to improving the reliability and scalability of fault detection systems.

3. Proposed Approach

The Figure 1 illustrates Basic power transmission system with a fault section including key components.

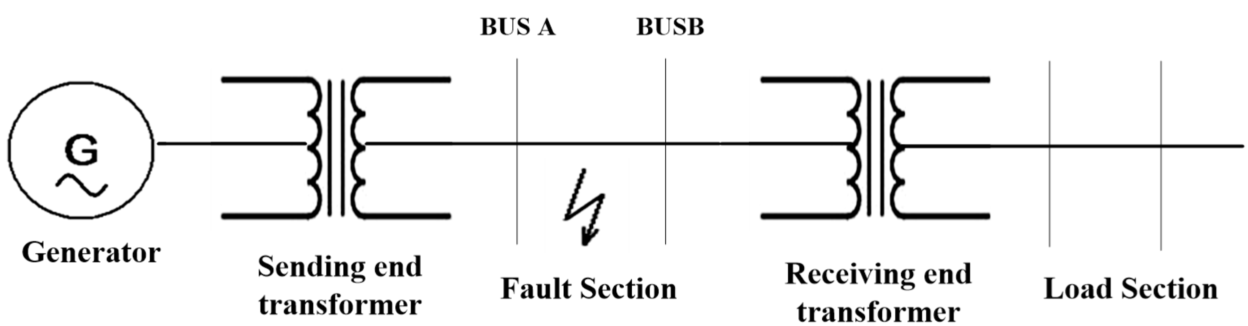

Figure 1. Basic power transmission system with a fault section

3.1. Data Source & Data Specification

The dataset used for this study is sourced from [17] [18], the dataset consists of 7,861 entries and 11 features, primarily focusing on electrical parameters related to power system faults. The columns include four categorical or binary variables (G, C, B, A), three-phase current measurements (Ia, Ib, Ic), and three-phase voltage measurements (Va, Vb, Vc). The target variable, "Fault Type," is binary, indicating whether a fault occurred (1) or not (0). Notably, there are no missing values, making it suitable for machine learning applications.

3.2 Flow chart

The Figure 2. Shows flowchart, this flowchart represents the pipeline for developing a machine learning model. Data preparation, which includes data distribution analysis, correlation analysis, and standardization, comes after data collection. Next, machine learning models, such as Logistic

Regression and Random Forest, are trained and tested. After testing, the model undergoes evaluation to determine its performance. If the model is satisfactory, it is finalized as the Final Model. However, if the model does not meet expectations, hyper parameter tuning using GridSearchCV is performed, and the model is retrained and re-evaluated until it reaches an acceptable performance level.

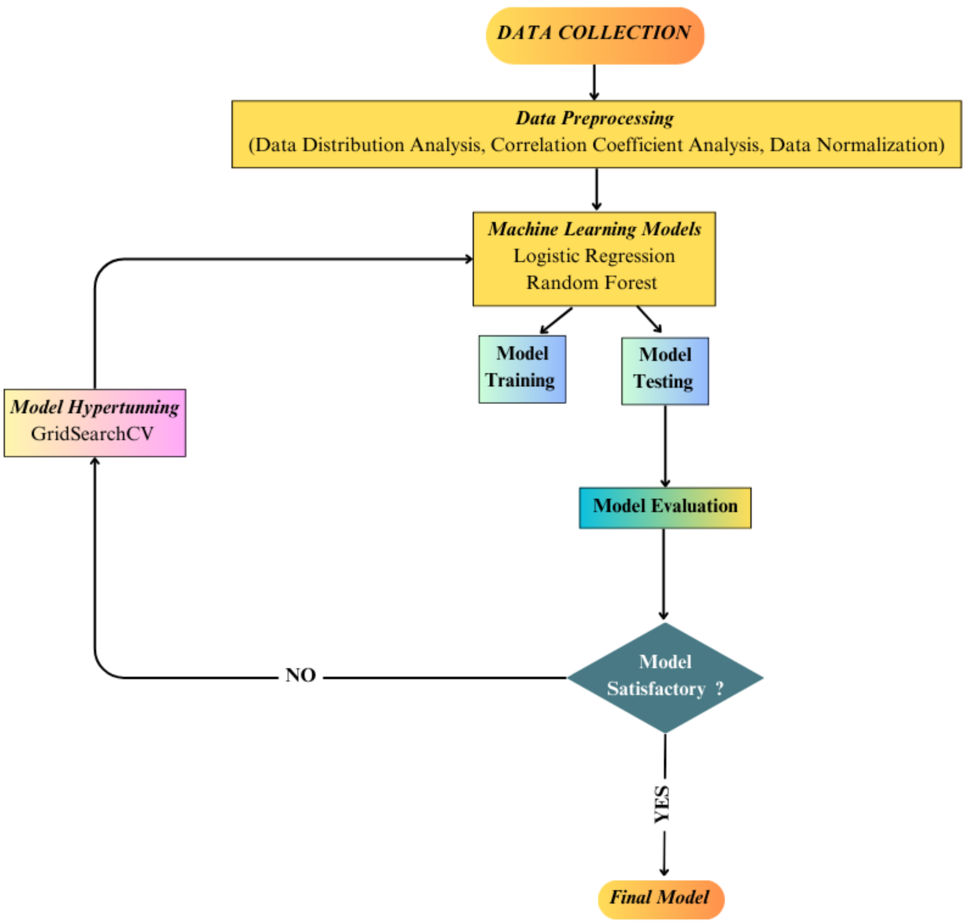

Figure 2. Flowchart of the machine learning model

3.3 Data Preprocessing

The pair plot (scatter matrix) visualization, which illustrates the interactions between several variables in a dataset, is shown in Figure 3. The off-diagonal plots show scatter plots comparing several feature pairings, while the diagonal plots show the distribution of each variable.. The color coding represents different fault types (0 and 1), indicating that the dataset contains labeled instances for classification. The ellipses highlight data clusters and correlations between features. This visualization helps in understanding the feature relationships, separability of fault types, and possible patterns in the dataset. The Figure 4 presents histograms displaying the distribution of current (Ia, Ib, Ic) and voltage (Va, Vb, Vc) values in the dataset. The current distributions (Ia, Ib, Ic) show a bimodal pattern, indicating the presence of two distinct groups corresponding to different fault conditions. The voltage distributions (Va, Vb, Vc) appear more spread out, with multiple peaks, suggesting variations in voltage levels due to system disturbances or faults. These distributions provide insight into how different electrical parameters behave and can help in identifying patterns associated with different types of faults. The Figure 5 shows correlation heatmap illustrates the relationships between various electrical parameters, including line currents (Ia, Ib, Ic) and voltages (Va, Vb, Vc), along with their association with the fault type. Notably, there is a negative correlation between Ia and Ib (-0.37), as well as between Ib and Ic (-0.53), suggesting an inverse relationship among line currents. Similarly, the voltages Va, Vb, and Vc exhibit moderate negative correlations, indicating interdependencies between line voltages during fault conditions. The Fault Type shows weak correlations with individual parameters, with the highest being with Ia (0.03), Ic (0.11), and Vc (0.035).

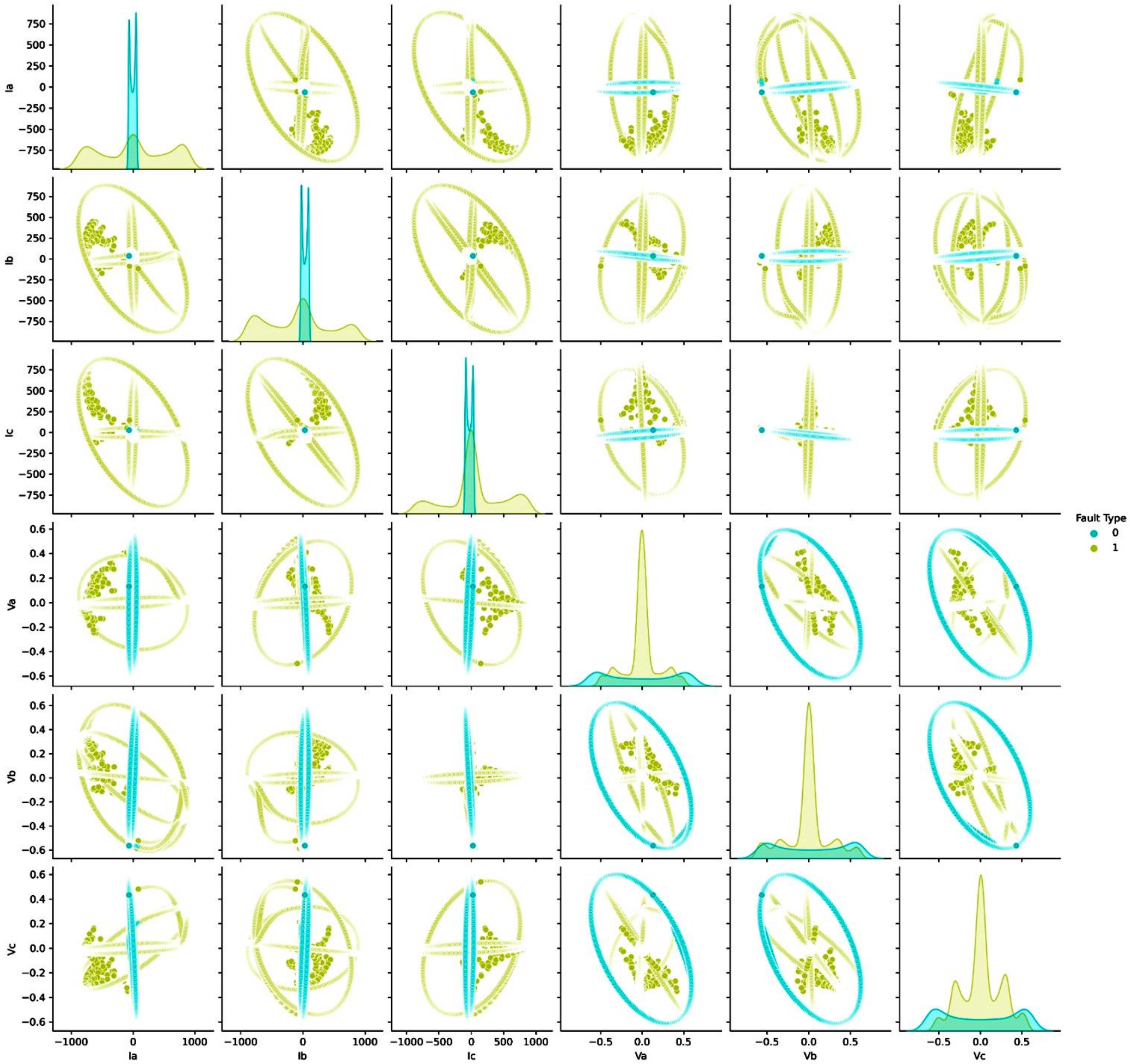

Figure 3. Pair Plot Visualization

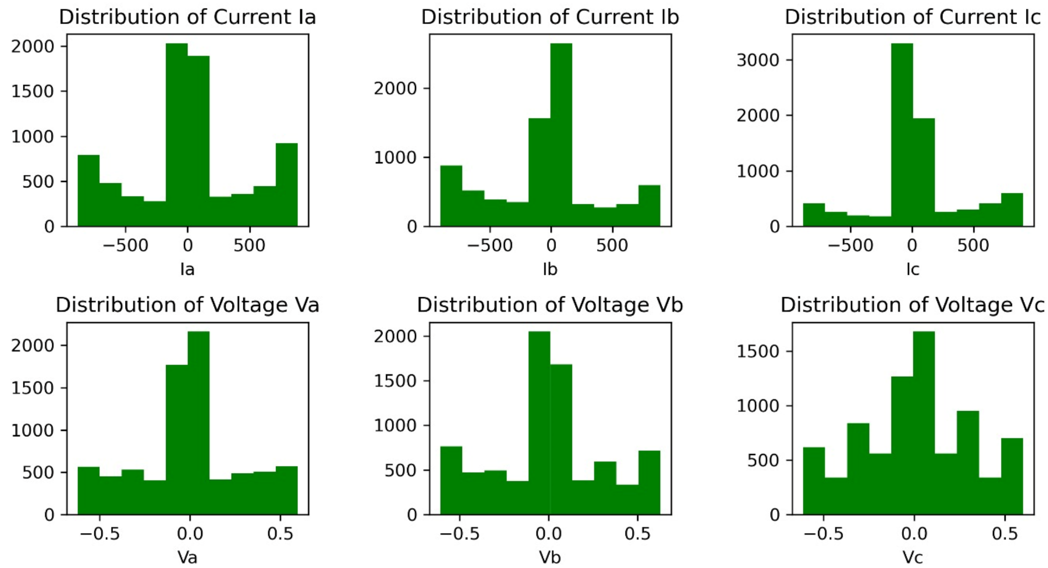

Figure 4. Distribution Analysis

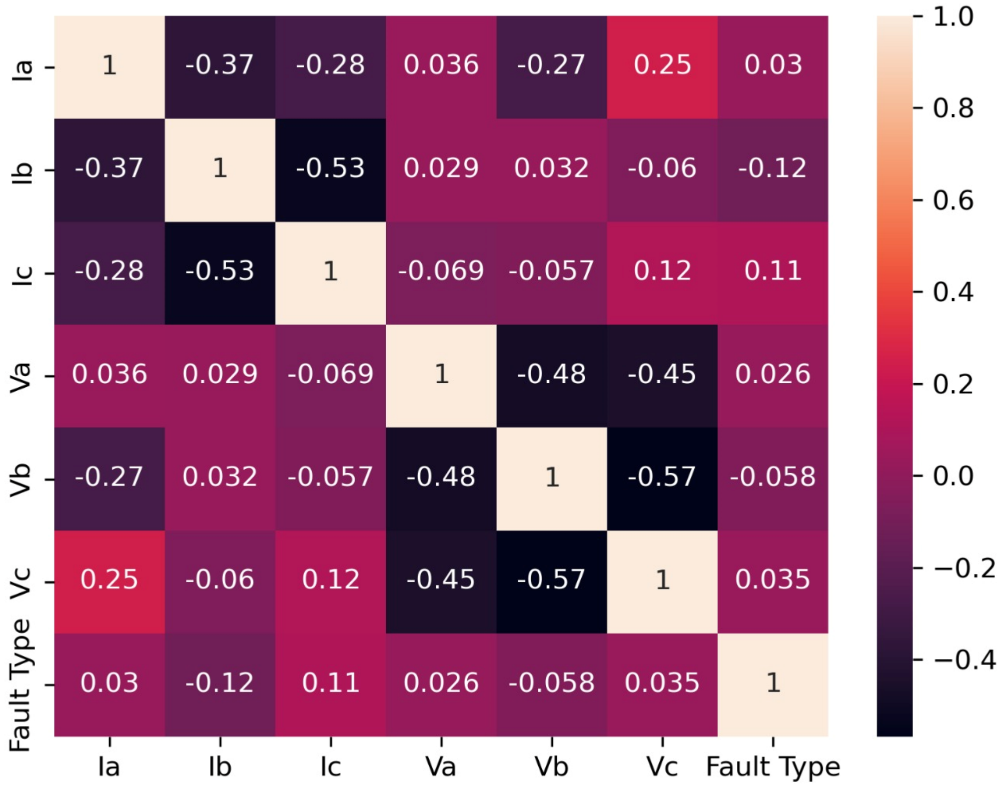

Figure 5. Feature Correlation Heatmap

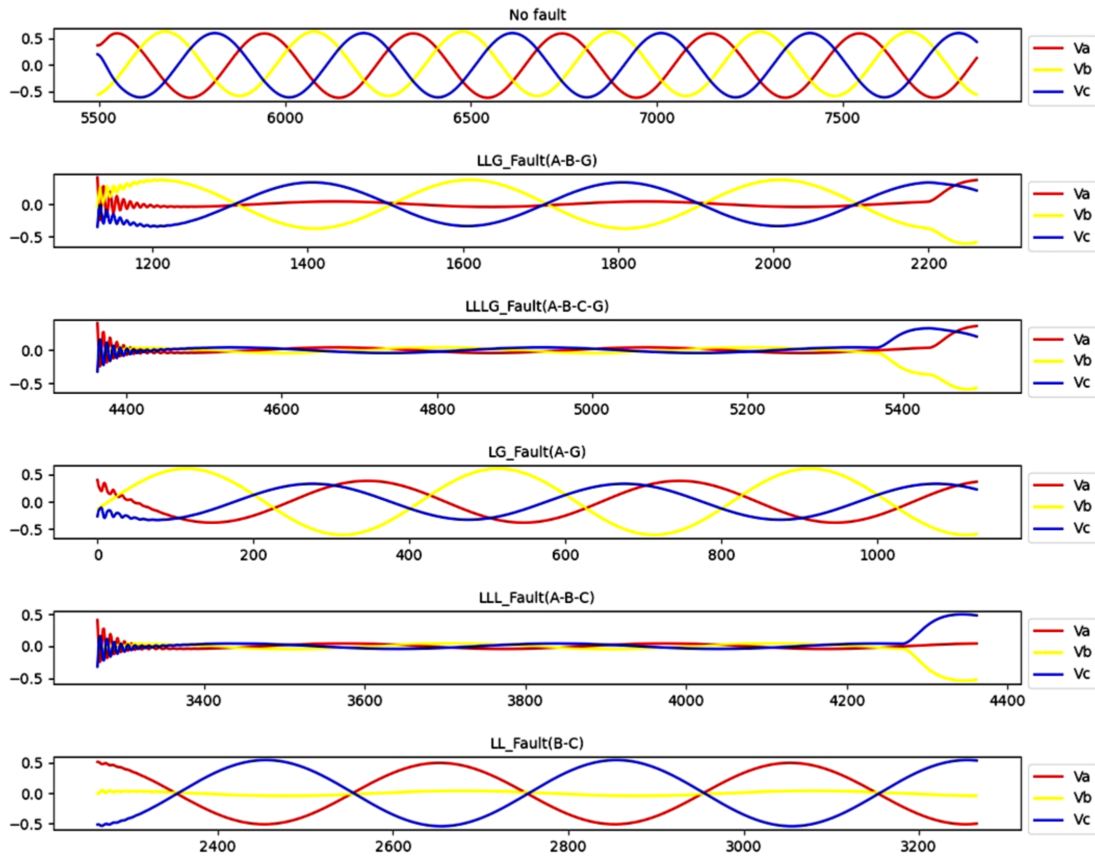

Figure 6. Voltage Waveforms under Different Fault Conditions

The Figure 6. shows Voltage Waveforms under Different Fault Conditions, it displays the line voltages (Va, Vb, Vc) under different fault conditions in a power system. The top subplot represents the normal operating condition, showing balanced sinusoidal waveforms for all three phases. In contrast, the subsequent plots illustrate various fault types, such as Line-to-Line (LL), Line-to-Ground (LG), and combinations like LLG (Line-Line-Ground) and LLLG (Three-Line-Ground). Each fault type introduces disturbances, leading to distortions in voltage waveforms. For instance, in LLLG faults, all three line voltages are significantly reduced, indicating a severe fault. LG faults show asymmetrical distortions due to unbalanced fault currents. Such visual analysis aids in understanding fault behavior and validating machine learning models for fault classification.

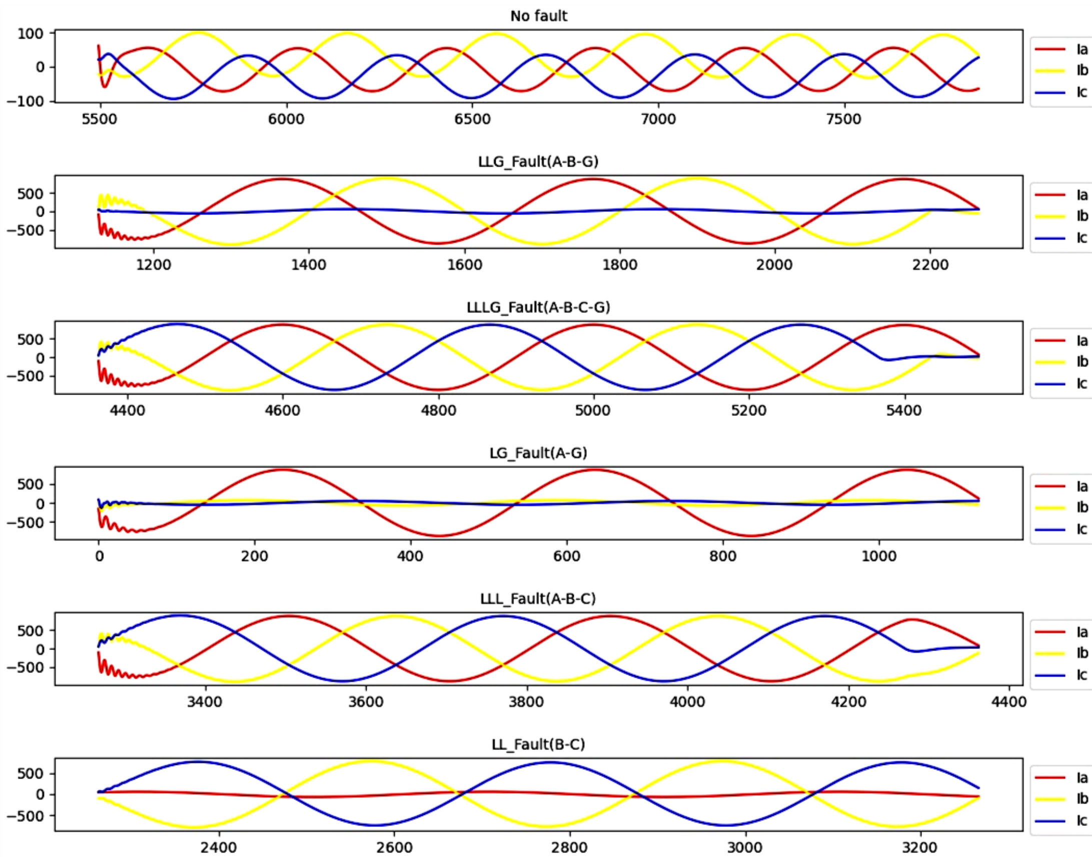

Figure 7. Current Waveforms under Different Fault Conditions

The Figure 7. illustrates different fault conditions in a three-phase electrical system by showing the variations in line currents (Ia, Ib, and Ic) over time. The first plot represents the normal operating condition where all three-phase currents are balanced and sinusoidal. As faults occur, the waveforms become distorted, indicating abnormal current flows. In a Line-Line-Ground (LLG) fault (A-B-G), the currents in the affected phases (Ia and Ib) show significant disturbances, while Ic remains relatively unchanged. A more severe Line-Line-Line-Ground (LLLG) fault (A-B-C-G) affects all three phases, leading to high fault currents. In a Line-to-Ground (LG) fault (A-G), only one phase current (Ia) is affected, while the others remain stable. A Line-Line-Line (LLL) fault (A-B-C) causes symmetrical disturbances in all three phases, making it one of the most severe fault types. Lastly, a Line-to-Line (LL) fault (B-C) impacts two phases (Ib and Ic), while the unaffected phase (Ia) remains stable. In power system analysis, these defects are essential because they aid in problem diagnosis and preventive action implementation.

4. Results and Discussion

The Figure 8 presents four confusion matrices evaluating Logistic regression and Random Forest models. Poor performance is shown by the first model's large number of false positives and false negatives, which result in considerable misclassifications. The second matrix, representing Logistic Regression with GridSearchCV, performs slightly better but still exhibits notable misclassification errors. In contrast, the third model demonstrates near-perfect classification, with only two false positives. The fourth matrix, representing a Random Forest model with GridSearchCV, achieves perfect classification with no misclassifications. This comparison highlights the superior performance of the Random Forest model after hyper parameter tuning, emphasizing the importance of model selection and optimization for achieving high accuracy.

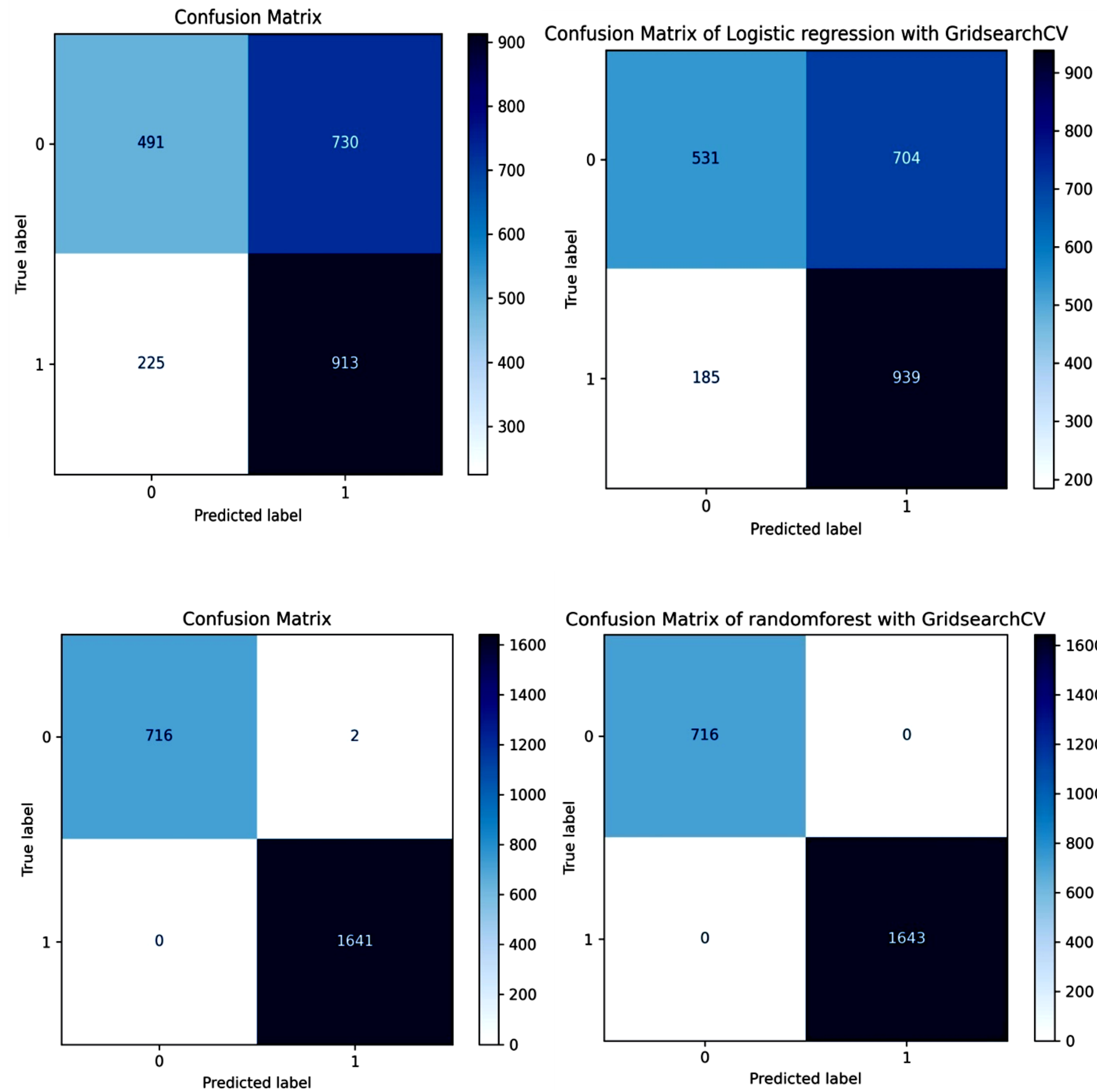

Figure 8. Confusion Matrix

The table 2. Shows Performance Metrics comparison between Logistic Regression and Random Forest models highlights significant differences in classification effectiveness. Logistic Regression, without hyper parameter tuning, achieves a modest accuracy of 59.51%, performing better in detecting faults (F1-score: 0.66) than no-fault conditions (F1-score: 0.51). With hyper parameter tuning, its accuracy slightly improves to 62.31%, with optimized parameters (C=10, penalty='l2', solver='saga'), though it still struggles in classification. In contrast, the Random Forest model demonstrates exceptional performance, achieving 99.91% accuracy without tuning and a perfect 100% accuracy after hyper parameter optimization. The tuned Random Forest model, using parameters like criterion='gini', max_depth=10, min_samples_leaf=1, min_samples_split=5, and n_estimators=40, flawlessly classifies fault and no-fault conditions. This comparison underscores the superiority of Random Forest in fault classification and emphasizes the impact of hyperparameter tuning in enhancing model performance.

Table 2. Performance Metrics

Model (Logistic regression)	**Test Accuracy (%)**	**Fault Conditions**	**Precision**	**Recall**	**F1-Score**
Without Hyper parameter Tuning	59.51	No Fault	0.69	0.4	0.51
		Fault	0.56	0.8	0.66
Best parameters are: {'C': 10, 'penalty': 'l2', 'solver': 'saga'}					
With Hyper parameter Tuning	62.31	No Fault	0.74	0.43	0.54
		Fault	0.57	0.84	0.68

Model (Random Forest)	**Test Accuracy (%)**	**Fault Conditions**	**Precision**	**Recall**	**F1-Score**
Without Hyper parameter Tuning	99.91	No Fault	1.0	1.0	1.0
		Fault	1.0	1.0	1.0
Best parameters are: {'criterion': 'gini', 'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 40}					
With Hyperparameter Tuning	100	No Fault	1.0	1.0	1.0
		Fault	1.0	1.0	1.0

Conclusion:

This study underscores the power of machine learning in electrical fault classification, comparing the performance of Logistic Regression and Random Forest. The findings reveal that Logistic Regression, even with hyper parameter tuning, struggles to achieve high accuracy, peaking at just 62.31%. In contrast, Random Forest demonstrates exceptional performance, achieving an impressive 99.91% accuracy without tuning and a flawless 100% accuracy after optimization. These findings demonstrate how important model selection and hyperparameter tuning are to improving the accuracy of fault identification. The study confirms that Random Forest is a highly reliable approach, delivering robust and precise classification. Looking ahead, future research could explore additional real-world constraints, refine feature selection techniques, and investigate deep learning methodologies to further improve classification accuracy and overall system reliability.

References

[1] S. Mandal and A. K. Bohre, “Fault Classification in Electrical Systems Using Machine Learning Algorithms,” Aug. 2022. doi: 10.1109/ICICICT54557.2022.9917976.

[2] H. Mirshekali, R. Dashti, and H. R. Shaker, “A Novel Fault Location Algorithm for Electrical Networks Considering Distributed Line Model and Distributed Generation Resources,” in *IEEE PES Innovative Smart Grid Technologies Europe*, Oct. 2020. doi: 10.1109/ISGT-EUROPE47291.2020.9248755.

[3] S. Heo and J. H. Lee, “Fault Detection and Classification Using Artificial Neural Networks,” *IFAC-PapersOnLine*, vol. 51, no. 2, pp. 115–120, Jan. 2018. doi: 10.1016/J.IFACOL.2018.09.380.

[4] M. Fernandes, A. Canito, J. M. Corchado, and G. Marreiros, “Fault Detection Mechanism of a Predictive Maintenance System Based on Autoregressive Integrated Moving Average Models,” *Springer*, 2019. doi: 10.1007/978-3-030-23887-2_20.

[5] J. De La Ree, Y. Liu, L. Mili, A. G. Phadke, and L. A. DaSilva, “Catastrophic Failures in Power Systems: Causes, Analyses, and Countermeasures,” *Proceedings of the IEEE*, vol. 93, no. 5, pp. 956–964, May 2005. doi: 10.1109/JPROC.2005.847246.

[6] T. Knüppel, M. Blanke, and J. Østergaard, “Fault Diagnosis for Electrical Distribution Systems Using Structural Analysis,” *Int. J. Robust Nonlinear Control*, vol. 24, no. 7, pp. 1320–1336, May 2014. doi: 10.1002/RNC.3080.

[7] P. Peng, T. Schiestel, Y. Zhang, H. Wang, and H. Zhang, “Towards Robust and Understandable Fault Detection and Diagnosis Using Denoising Sparse Autoencoder and Smooth Integrated Gradients,” *ISA Transactions*, vol. 114, pp. 412–423, Jun. 2021. doi: 10.1016/J.ISATRA.2021.06.005.

[8] A. Abdelmoumene and H. Bentarzi, *Electrical Faults in Power Systems*, Hershey, PA, USA: IGI Global, 2016. doi: 10.4018/978-1-4666-9911-3.CH001.

[9] J. W. Oliveira et al., “Improving Fault Detection in Electrical Power Systems Using Multiple Classifier Systems,” Oct. 2024. doi: 10.20944/preprints202410.0239.v1.

[10] A. K. Sahoo and S. K. Samal, “Deep Learning-Based Detection, Classification, and Location of Power System Faults,” *Bull. Electr. Eng. Inform.*, vol. 13, no. 6, pp. 4248–4259, Sep. 2024. doi: 10.11591/eei.v13i6.7239.

[11] P. Tatraiya, H. Priyadarshi, A. Srivastava, and V. Gupta, "Enhancing Power System Reliability: Leveraging Machine Learning for Unsymmetrical Fault Detection and Analysis," *Proc. IEEE SEFET 2024*, pp. 1–6, Jul. 2024. doi: 10.1109/sefet61574.2024.10717482.

[12] H. Liu and X. Liu, "Electrical Fault Detection and Classification Based on Multiple Machine Learning Algorithms," *Appl. Comput. Eng.*, vol. 74, no. 1, pp. 245–250, Jul. 2024. doi: 10.54254/2755-2721/74/20240484.

[13] S. Syed, B. Li, and A. Zheng, "Detection and Classification of Physical and Electrical Fault in PV Array System by Random Forest-Based Approach," *Int. J. Electr. Energy Power Syst. Eng.*, vol. 7, no. 2, pp. 67–84, Jun. 2024. doi: 10.31258/ijeepse.7.2.67-84.

[14] D. Kulshrestha and V. Chaudhary, "Enhancing Electrical Fault Detection in Transmission Lines Through Machine Learning," *Proc. IEEE ICEPES 2024*, vol. 10, pp. 1–7, Jun. 2024. doi: 10.1109/icepes60647.2024.10653619.

[15] M. Singla, R. Chauhan, K. S. Gill, and A. Kapruwan, "Usage of a Support Vector Machine Learning Classifier to Identify and Sort Electrical Faults," *Proc. IEEE OTCON 2024*, pp. 1–5, Jun. 2024. doi: 10.1109/otcon60325.2024.10688000.

[16] S. H. Khaleefah et al., "A Deep Learning-Based Fault Detection and Classification in Smart Electrical Power Transmission System," *JOIV: Int. J. Informatics Vis.*, vol. 8, no. 2, pp. 2701–2713, May 2024. doi: 10.62527/joiv.8.2.2701.

[17] M. Jamil, S. K. Sharma, and R. Singh, "Fault Detection and Classification in Electrical Power Transmission System Using Artificial Neural Network," *SpringerPlus*, vol. 4, p. 334, 2015. doi: 10.1186/s40064-015-1080-x.

[18] E. Sathyaprakash, "Electrical Fault Detection and Classification Dataset," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/esathyaprakash/electrical-fault-detection-and-classification