HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Fault Detection in Electrical Grids: Harnessing Machine Learning for Enhanced Reliability

William Jiang[1], Alfred Renaud[#] and Guillermo Goldzstein[#]

[1]Winston Churchill High School, USA
[#]Advisor

ABSTRACT

This study focuses on enhancing the reliability and redundancy of electrical systems through machine learning-based fault detection. This project's objective is to create an inexpensive system for early detection of faults in electrical systems. This study proposed a methodology combining machine learning techniques and feature engineering. A dataset containing 12,001 sensor readings each consisting of six values was examined and analyzed by machine-learning techniques. Results showed the system's success in detecting faults with an accuracy of 99% from only six readings. In economically underserved regions, a common trend is the presence of less sophisticated electrical infrastructures. These systems translate into more frequent and longer power outages, a serious concern when essential institutions like hospitals rely on them to function. This research contributes to the field of fault detection by offering a practical and effective way to improve fault detection in underserved regions.

Introduction

In every aspect of life, there are problems that arise. It can be agreed upon that the most challenging and tedious problems to solve are those that are both the most frequently occurring and most severe. This is exactly the case for electrical systems, as the short circuit fault is both the most common and the most dangerous, thus requiring a very accurate and reliable solution. Spanning fields like engineering, manufacturing, and computer science, fault detection plays a vital part in ensuring the safety and reliability of electrical systems, whether they pertain to aircraft, machinery, or software platforms. While basic fault detection suffices in certain areas, industries like aeronautics and extensive power grids demand sophisticated fault detection systems to ensure people don't get harmed.

In underserved regions, a common trend is the presence of less sophisticated electrical systems. While these systems are more economical and consist of fewer components prone to malfunction, those components are of worse quality and the system as a whole lacks substantial redundancy. These factors translate into more frequent and prolonged power outages. This is a concern for important sectors such as healthcare, transportation, and government, where power outages could result in catastrophic consequences. A good fault detection system can not only reduce faults by preemptively solving them, but it can also shorten the length of outages by enabling quicker response time.

This study examined if fault detection in electrical grids could be improved to be faster and more accurate through machine learning techniques. The methods in this study are relatively straightforward. Although more advanced algorithms and their implementation are more accurate and can achieve bigger goals, one limitation they have is that they are more time-consuming. Since this research is for the benefit of underserved areas, a quick improvement in the short-term seemed to be a good first step which is what this study is addressing. To do this, 12001 rows of data values were analyzed through Weka visualization then a simple three-layer neural network was used in conjunction with logistic regression to produce a prediction model to detect faults.

HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 1 (2024)

The subsequent sections of this research paper are organized as follows: The initial section takes a more extensive look and explanation of the dataset used. Following this is a specific explanation of the methodology used. Next, the results of the model pertaining to accuracy and precision are provided. Finally, the last section summarizes the findings and raises further discussion to be made in the future.

Data Analysis

This study used historical sensor data from publicly available power grid transmission lines. The data consisted of 12001 rows of six features(Ia, Ib, Ic, Va, Vb, Vc) and one label(Output(s)).

Table 1. 4 examples from the original dataset

Output(s)	Ia	Ib	Ic	Va	Vb	Vc
0	-90.1615	3.813632	86.34784	0.141026	-0.60528	0.464251
0	-79.9049	2.398803	77.50611	0.156272	-0.60224	0.445963
1	59.22282	-806.627	749.5665	-0.03512	-0.00146	0.036581
1	72.42094	-812.13	741.871	-0.03526	-0.00199	0.037244

As seen in Table 1, the feature's titles were represented by 2 letters. The first letter was either "I" representing current(measured in amps), or "V" representing voltage(measured in volts). Current is the rate and direction of electrical flow. The greater the absolute value of the reading is, the faster the rate is and a negative reading indicates the direction of the current is opposite to what the sensor is set to. Voltage is the electrical pressure or of the electric potential energy at a point. The second letter is either "a", "b", or "c", each representing one transmission line. Ex: Ia: current of a transmission line a; Vb: voltage of transmission line b. These transmission lines all originate from a single substation, therefore having three sets of readings improves the chances of detecting a fault from that substation. The output value is either a 0 or a 1, with "0" being no fault and "1" representing a fault.

Looking at the original dataset through the naked eye, it's almost impossible to tell how each value corresponds to the other and the output. Therefore, Weka was used to visualize both the spread of the data as well as their relationships.

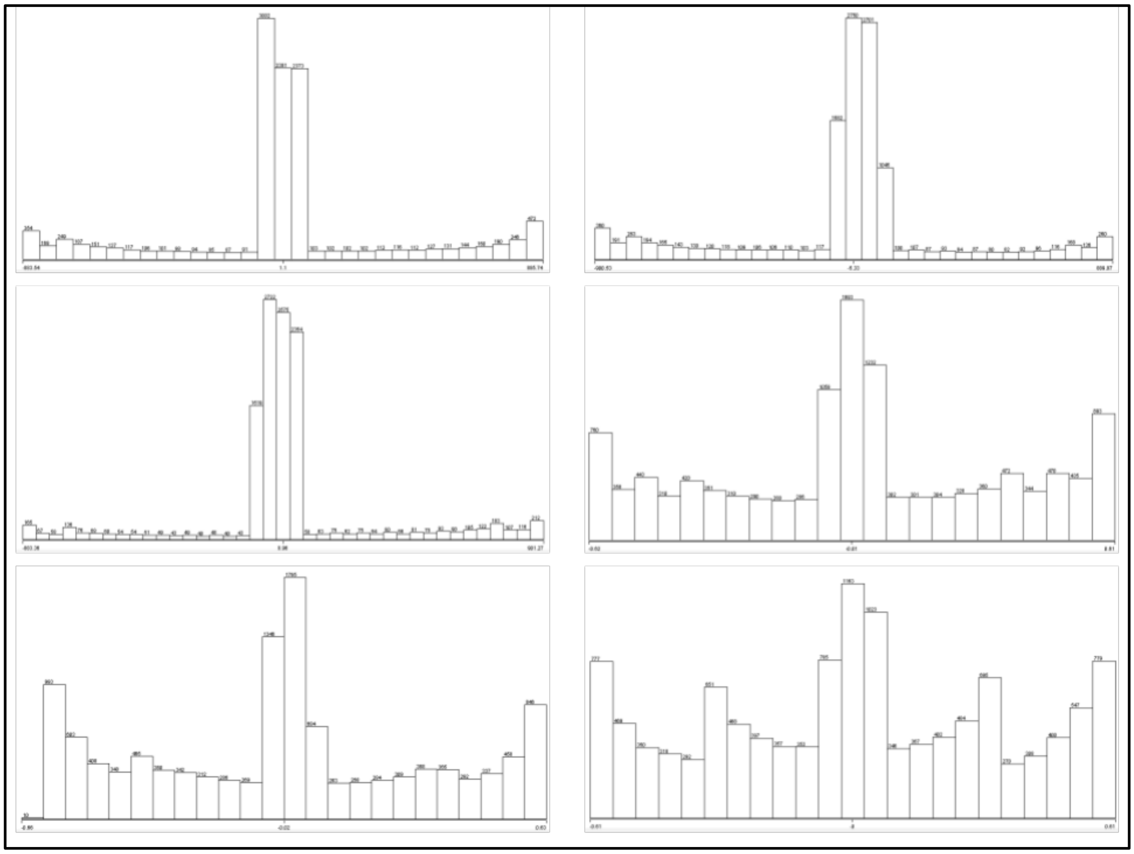

Figure 1. Visualization of data spread of features. From top left to bottom right(Ia, Ib, Ic, Va, Vb, Vc)

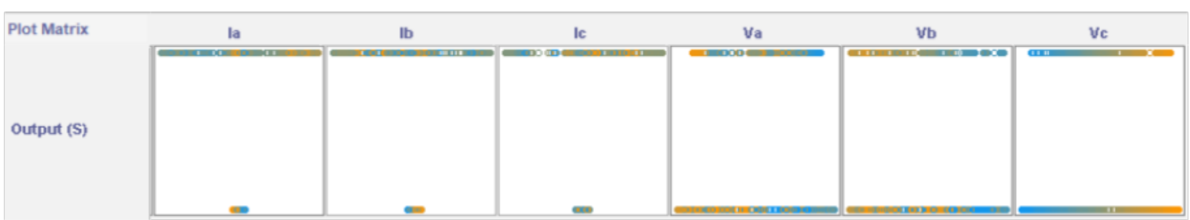

Figure 2. Visualization of the relationship between the features and label.

These visualizations helped determine if the data had any holes and if it was plausible to run machine learning algorithms with it. From observing the plots, the study concluded that the data didn't have any holes, the distribution of features was normal, and it was compatible with machine learning algorithms.

Data Preparation

The dataset was then split into 2 different categories, a training set(60%) and a validation set(40%). Although the Data wasn't randomly split into the two sets, the natural order of the dataset allowed for an accurate and unbiased representation in each set. The reason for this split was to ensure the accuracy of the model and prevent overfitting in the model. Overfitting is when the model becomes accurate at predicting the training set but not at accurately predicting new data; the validation set prevents this by introducing new data to the model made from the training set to reassess the model. After this, both the training set and validation set were scaled to prepare for machine learning algorithms.

Neural Networks & Error

Neural networks are a type of machine learning model inspired by the structure of the human brain. Consisting of interconnected nodes, or "neurons," organized in layers, all neural networks contain at least two layers; the first layer is the input/feature(s), and the final layer is the output/label(s). Each node has a function built into it that uses the data from the previous node(s) into it to produce an output value. Each layer added in between the two original layers increases the complexity of the model and offers a significant chance of increasing the accuracy as well. Logistic regression is the most simple form of a neural network, with only two layers. With Neural Networks it is possible to modify the model to achieve more accuracy by experimenting with the shape of the neural network and the number of epochs(iterations) the training set was run through the neural network. This study focuses on predicting fault or no-fault, a binary classification problem, so binary cross-entropy error was used to create and evaluate the model.

Model Creation

This study uses a neural network model to detect faults. The previous data visualizations hinted that the model didn't need to be complicated, so the initial model was a Logistic Regression model/two-layer neural network(Figure 3.). This, unfortunately, yielded sub-par results: 75% accuracy, 74% precision, and 59% recall for detecting faults. Various three-layer and four-layer models were experimented with: changing the number of nodes in each hidden layer and the number of epochs that the algorithm took. The hidden layers employed the "relu" function which solves two problems the "sigmoid" function used in the first and last layer. This allowed for a more effective function to be embedded within the nodes, resulting in a more accurate model. After the new models were made each of their results was compared and the one that provided the most accuracy with relatively the fewest layers and nodes was chosen. This resulted in the final model: A three-layer neural network, with the second layer consisting of four nodes(Figure 4.). By trialing different epochs for the model to take, the model arrived at an optimal 100 epochs, where the binary cross-entropy error flatlined(Figure 5.), meaning that the best model was achieved.

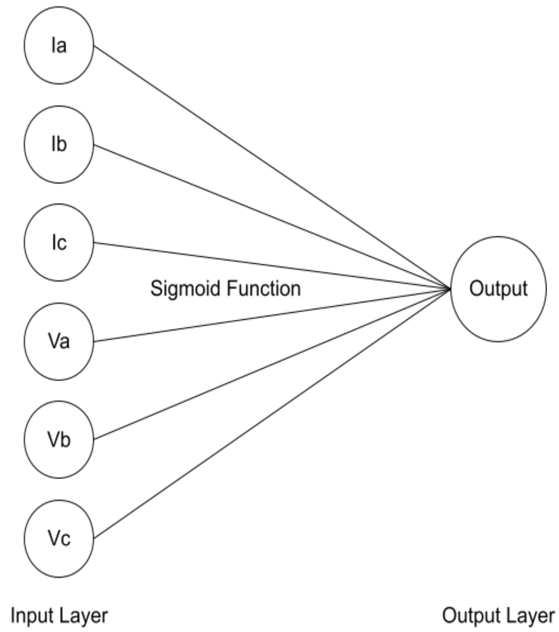

Figure 3. First Model

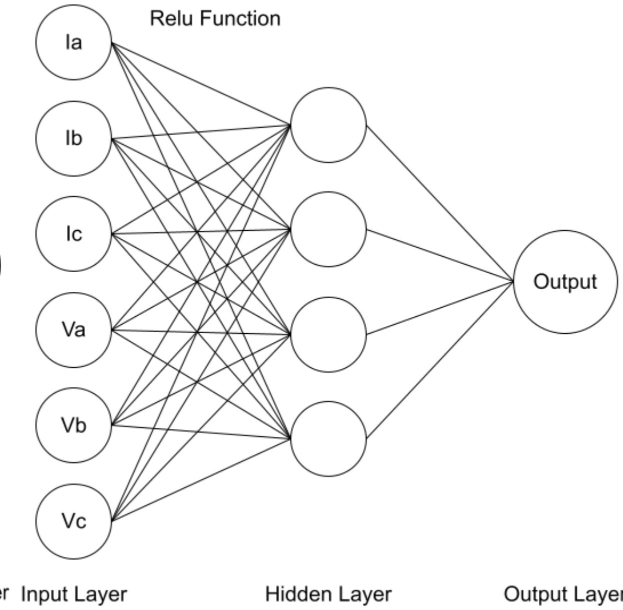

Figure 4. Final Model

Results

With the methods discussed above, the model on the 100th epoch produced a validation set accuracy of 0.99 and a validation loss of 0.042. This indicates that the model's performance in detecting no-fault and faults was very good, having 99% accuracy. Our model also had an average precision as well as recall of 99%, meaning there were almost no false predictions and almost all positives/faults were detected. This is even more important in the field of fault detection as a false negative can produce anywhere from a tiny to a huge consequence. The validation set the model used had an even spread of validating fault or no-fault(0 or 1) so it can be concluded that there is no bias present in the final result.

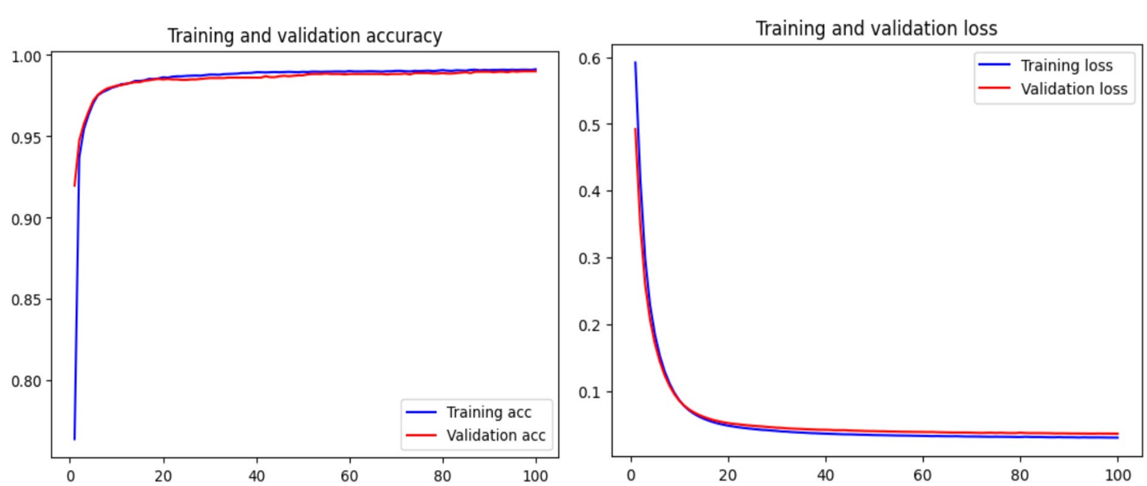

Figure 5&6. Training and validation set accuracy and loss over a range of 0-100 epochs.

Table 2. Validation set classification report

Validation Set	Precision	Recall	F1-Score	Support
0	0.98	0.99	0.99	1933
1	0.99	0.98	0.99	1668
accuracy			0.99	3601
Macro avg	0.99	0.99	0.99	3601
Weighted avg	0.99	0.99	0.99	3601

Conclusion and Discussion

These results show that through the use of relatively simple machine learning techniques and a common electrical system dataset, a model can be created that very accurately detects faults. In the field of Fault Detection, there are usually four guidelines to evaluate a model: Accuracy, false positives/negatives, quality and quantity of data available, and human intervention. This model meets or exceeds the requirements for all these guidelines. The model has a 99% accuracy which is adequate for most fault detection systems and it has an average precision of 99%, meaning it has next to no false positives or negatives. Electrical substations, where this model will be used, all store continuous data readings identical or easily translated to the ones used, so since data is being generated in real-time at the substations, the quality will be good and the quantity will be infinite. Finally, virtually all modern computing devices have the capacity to run this model with very little stress on their computing power and require no human assistance other than when downloading and starting the model. The limited stress on the computer will also result in less lagging, crashes, or problems that require human intervention. Furthermore, another advantage of this model is that it is tailored toward economically underserved areas and their electrical systems. This shows that those economically underserved systems can utilize this model as easily and efficiently as significantly more advanced systems.

The limitation of the model is that it is not a relatively good option for more expensive systems as those systems will have more unique faults not able to be detected through this model. Furthermore, this study only detects the presence of the fault but doesn't classify what type of fault is present. In future studies, there is the possibility of using these readings to classify the type of fault as well as detect it. Each electrical substation will also have hundreds of other readings available for analysis, so there is also the possibility of using more data values to classify faults in the case that six readings are not enough to classify faults accurately.

The question of how machine learning-based fault detection compares to traditional methods of current and vibration analysis will have to be analyzed in a future study. Also, the use of unsupervised and semi-supervised learning algorithms as opposed to supervised learning algorithms and various other algorithms poses the question of whether this fault detection model can be expanded and enhanced proves to be another interesting topic.

Acknowledgments

I would like to thank Mr. Alfred Renaud and Dr. Guillermo Goldsztein, my teacher and faculty advisors for guiding me through this project. I would also like to thank the entire Scholar Launch team for providing me with the resources required to do this project.

References

Faults in Power System: Statistics and Kinds | Engineering Notes (n.d.). Retrieved August 24, 2023, from https://www.engineeringenotes.com/electrical-engineering/power-system/faults-in-power-system-statistics-and-kinds-electrical-engineering/24433

Jamil, M., Sharma, S.K. & Singh, R. (2015). Fault detection and classification in electrical power transmission system using artificial neural network. SpringerPlus, 4(334). https://doi.org/10.1186/s40064-015-1080-x

Rachna Vaish, U.D. Dwivedi, Saurabh Tewari, S.M. Tripathi. (2021). Machine learning applications in power system fault diagnosis: Research advancements and perspectives, Engineering Applications of Artificial Intelligence,106(104504). https://doi.org/10.1016/j.engappai.2021.104504

Soufiane Belagoune, Noureddine Bali, Azzeddine Bakdi, Bousaadia Baadji, Karim Atif. (2021). Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems, Measurement, 177(109330). https://doi.org/10.1016/j.measurement.2021.109330

Md.Omaer Faruq Goni, Md. Nahiduzzaman, Md.Shamim Anower, Md.Mahabubur Rahman, Md.Robiul Islam, Mominul Ahsan, Julfikar Haider, Mohammad Shahjalal. (2023). Fast and Accurate Fault Detection and Classification in Transmission Lines using Extreme Learning Machine, e-Prime - Advances in Electrical Engineering, Electronics and Energy, 3(100107). https://doi.org/10.1016/j.prime.2023.100107

Michail Cheliotis, Iraklis Lazakis, Gerasimos Theotokatos. (2020). Machine learning and data-driven fault detection for ship systems operations, Ocean Engineering, 216(107968). https://doi.org/10.1016/j.oceaneng.2020.107968

Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, Asoke K. Nandi. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap, Mechanical Systems and Signal Processing, 138(106587) https://doi.org/10.1016/j.ymssp.2019.106587