UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

Detección y Clasificación de Cortocircuitos en Sistemas de Potencia utilizando Redes Neuronales Convolucionales Recurrentes

Proyecto de Titulación

Fausto Gonzalo Yugcha Guevara

Felipe Grijalva, Ph.D.

Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito para la obtención del título de Magíster en Inteligencia Artificial

Quito, 02 de Diciembre 2024

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Detección y Clasificación de Cortocircuitos en Sistemas de Potencia

Fausto Gonzalo Yugcha Guevara

Nombre del Director del Programa:	Felipe Grijalva
Título académico:	Ph.D. en Ingeniería Eléctrica
Director del programa de:	Inteligencia Artificial

Nombre del Decano del colegio Académico:	Eduardo Alba
Título académico:	Doctor en Ciencias Matemáticas
Decano del Colegio:	Ciencias e Ingenierías

Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Doctor en Física

Quito, Diciembre 2024

© DERECHOS DE AUTOR

Nombre del estudiante: Fausto Gonzalo Yugcha Guevara

Código de estudiante: 00338526

C.I.: 1716077886

Lugar y fecha: Quito, 02 de Diciembre de 2024.

ACLARACIÓN PARA PUBLICACIÓN

UNPUBLISHED DOCUMENT

DEDICATORIA

Dedicado a quien sembró en mí las ganas de aprender nuevas cosas que me motivaron a empezar estos estudios, para quien ha estado a mi lado durante el proceso de realización de este trabajo siendo siempre un apoyo en todo sentido y para quien desde que escuché sus latidos espero ya llegue el día de poder sostener en mis brazos.

RESUMEN

Las fallas eléctricas en los sistemas de transmisión de energía pueden afectar significativamente la confiabilidad de la red y la seguridad operativa del sistema. Este estudio aborda la necesidad de una detección y clasificación de fallas precisa mediante un enfoque novedoso de aprendizaje automático que utiliza redes neuronales convolucionales recurrentes (CRNN). Se ha construido un dataset a partir de registros oscilográficos de eventos de falla reales en el formato COMTRADE (Common Format for Transients Data Exchange) recopilados de la red eléctrica del Ecuador. Estudios anteriores se han basado en datos simulados, mientras que esta investigación proporciona una representación más realista del comportamiento de sistemas eléctricos utilizando datos de fallas reales. La metodología propuesta implica el procesamiento de datos aplicando remuestreo, detección de un punto trigger, recorte de ventanas de señal y representación de señales de corriente y voltaje como espectrogramas, las cuales son entradas para una arquitectura CRNN. Se aplicó una técnica de sobremuestreo para garantizar un tamaño representativo entre distintos tipos de fallas para el entrenamiento. El modelo desarrollado demostró un desempeño alto con un accuracy de 96.3% al clasificar eventos de falla en tres clases: falla monofásica, bifásica y trifásica. Este método también proporciona información sobre las características espectrales de las fallas eléctricas.

Palabras clave: Fallas eléctricas, Espectrograma, LSTM, CRNN, Sistemas de Potencia, Clasificación de fallas.

ABSTRACT

Electrical faults in power transmission systems can significantly affect grid reliability and operational safety. This study addresses the need for an accurate fault detection and classification by developing a novel machine learning approach using Convolutional Recurrent Neural Networks (CRNN). A dataset is built from real fault events oscillography recordings in the COMTRADE format (Common Format for Transients Data Exchange) collected from Ecuador's Power Grid. Previous studies have relied on simulated data, while this research provides a more realistic representation of the behavior of electrical systems using actual fault event data. The proposed methodology involves data processing applying resampling, trigger point detection, signal window cutting, and current and voltage signal representation as spectrograms, which are input for a CRNN architecture. An oversampling technique was applied to ensure representative training between different fault types. The developed model demonstrated high performance, 96.3% accuracy in classifying fault events into three classes: single-phase, double-phase and three-phase fault. This method also provides insight into the spectral characteristics of electrical faults.

Key words: Electrical fault, Spectrogram, LSTM, CRNN, Power Systems, Fault Classification.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

Detection and Classification of Short-Circuit Faults in Power Systems Using Convolutional Recurrent Neural Networks (CRNN)

Felipe Grijalva, *Senior Member, IEEE,* Fausto Yugcha, *Member, IEEE*

***Abstract*—Electrical faults in power transmission systems can significantly affect grid reliability and operational safety. This study addresses the need for an accurate fault detection and classification by developing a novel machine learning approach using Convolutional Recurrent Neural Networks (CRNN). A dataset is built from real fault events oscillography recordings in the COMTRADE format (Common Format for Transients Data Exchange) collected from Ecuador's Power Grid. Previous studies have relied on simulated data, while this research provides a more realistic representation of the behavior of electrical systems using actual fault event data. The proposed methodology involves data processing applying resampling, trigger point detection, signal window cutting, and current and voltage signal representation as spectrograms, which are input for a CRNN architecture. An oversampling technique was applied to ensure representative training between different fault types. The developed model demonstrated high performance, 96.3% accuracy in classifying fault events into three classes: single-phase, double-phase and three-phase fault. This method also provides insight into the spectral characteristics of electrical faults.**

***Index Terms*—Electrical fault, Spectrogram, LSTM, CRNN, Power Systems, Fault Classification.**

I. Introduction

THE power transmission grid is exposed to external factors that can cause electrical faults. A faulted transmission line can break into many issues in power systems so in order to improve power system reliability a rapid fault detection is needed, to clear affected zones as fast as possible.

A bay in an electrical substation is a key functional unit where transformers, power lines, generation units or loads are connected. A bay allows for the control, protection and isolation of electrical power equipment, facilitating a safe and efficient operation of the power system. In figure 1 there is a one line diagram of an electric bay. For each bay there is a current transformer and a voltage transformer. They provide current and voltage signal to a protection relay which in case of detecting a fault will trip the circuit breaker.

In figure 2 there is a description of the stages in an electrical fault event. First, there is the normal operation state, a steady state. Balanced current flows on each phase of the power line. Ideally this should be the permanent state,

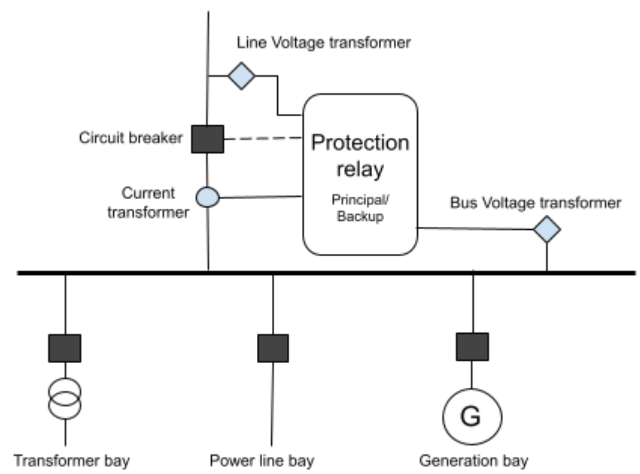

Figure 1. Bay one line diagram.

however, due to external factors electrical faults occur. When a fault occurs there is a increase of current in the faulted phase. Faults can be classified accordingly to the phases involved. In figure 2 we have a single phase event. During fault stage a relay will operate in order to clear the fault. A protection relay receives signals from a current transformer and a voltage transformer. A short circuit is identified by a relay which sends a trip signal to a circuit breaker in order to clear the fault. Ideally this happens as fast as possible, a 100 ms is a reference length of fault duration, although, based on experience, usually a fault lasts for 60 ms. Once circuit breakers have opened there is not any short circuit current so fault has been cleared. Protection relays create events based on a trip to a circuit breaker event which are recorded by in COMTRADE format, sometimes a fault event is also recorded by an automatic disturbance recorder. COMTRADE is a file format for storing oscillography and status data related to transient power system disturbances.

Protective relays can detect that a problem is developing by identifying slight deviations in current, voltage or resistance. According to the protection relay function it will work with a different algorithm to detect a fault that will require to measure instant current or a window with size of quarter or half a cycle and compare to a threshold to identify a fault.

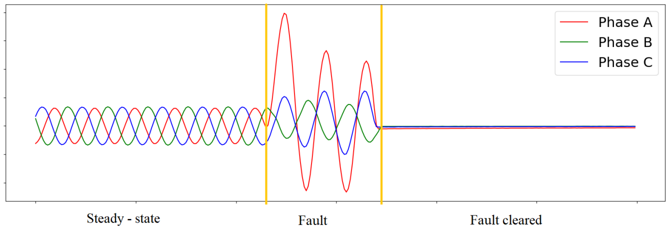

Figure 2. Electrical fault stages.

There is great potential for AI in power systems. AI techniques can further automate and increase the performance of power systems because their ability to compute large amounts of data at a faster speed than numerical optimization methods. [12]

II. Prior works

Article [19] talks about a software implemented called DFR Assistant, that analyzes data files coming from digital fault recorders (DFR) located in substations. The software facilitates the analysis of DFR recordings by automatically classifying all records into groups based on selected criteria, which otherwise could have been a burden to protection engineers to manually examine all DFR Records, identify most important for a given case and then analyze them. The role of the neural net in this solution is to process current signals and identify the one with the largest disturbance. It will try to classify disturbance according to the fault type.

A work done in [13] was based on wavelet transform method to detect and classify faults. It calculated RMS values of the wavelet coefficients of electrical current signals over a varying window. Coefficients were compared with the benchmark values to detect and classify faults.

Study in [14] considered the use of artificial neural network in transmission lines to detect and classify electrical faults. The results realized that the method is efficient in detecting and classifying faults in transmission lines. Faults are simulated by changing fault resistance and fault distance. Results shown in paper are for line to ground fault only and suggest to develop other ANNs for the rest of faults. Likewise, [15] results indicated that ANN can be used for detection, classification and location of faults. Study found that single line to ground faults can be detected and located with the smallest number of neurons in hidden layer, that being 5, while two phase and two phase to ground faults need to have minimum 10 neuron in hidden layer and that three phase fault needs 30 to 35 neuron in hidden layer. Work in [21] suggested to create a separate ANN to identify the affected phase for each fault type.

On [16] a database is created based on Automatic Meter Reading on an electrical secondary distribution network focused on voltage violations. RNNs performance was found to be the best for fault detection and classification so RNN models are compared.

Most of the studies reviewed use simulated data for training and analysis instead of using real events obtained from fault recorders. Some of the studies already use a neural network and identify it as an appropriate tool for fault classification but since there is a wide range of possible architectures different approaches can be tried. In this work, none of the faults used are simulated but they come from real events and an architecture CRNN is used to classify fault events.

III. Materials and Methods

A. Dataset

Power transmission grid operation department collects information related to daily function of the transmission grid. This include information related to short circuit events. Information is stored in NEVT, which is a client-server web application that accesses a relational MySQL database. In this study, data from three years has been used. A fault event is analyzed by reviewing an oscillography of related currents and voltages. Not every event has an oscillography while some have more than one oscillography. This is due to available equipment. Some substations have electromechanical protection relays which function properly detecting and clearing a fault but do not keep track of their operation while some substations have digital relays but lack a dedicated computer that would allow the substation operator to connect to the relay and download oscillographies. Because of redundancy, there is a principal and backup protection relay on each bay. Principal relay will receive current transformer signal and line voltage transformer signal while backup relay will receive current transformer signal and bus voltage signal. See figure 1 for reference. Because of this some events have more than one oscillography. Also, when a fault in a power line happens, there is a record from each end of the line.

An event related oscillography is in a COMTRADE format. In this work, a database has been build by extracting fault related currents and voltages and saving them as numpy arrays. PQDiffractor, a comtrade file viewer utility, has been used to extract a csv file for every short circuit event available due to its ease and reliability compared with software like SIGRA or WaveEv. There is a example on Appendix A.

It has been noticed that the order of the csv file obtained from PQDiffractor changes, sometimes the first columns are currents while some other times the first columns are voltages. This information has been recorded on column *'order'* of table I so it can be corrected during processing. Not necessarily every event is related to a short circuit event. Some are disconnections due to a systemic protection operated in order to keep balance between generation and load. Some events don't come from a steady-state but it was a grid element that was disconnected and it was energized into a fault. Some other are transformer energization events.

1) Labeling: Every event available as an oscillography has been reviewed in order to identify which phase is involved in the electrical fault. An ID number has been given based

Table I
Samples of event labeling table.

ID	order	Label _1	Label_2	Label_3	Obs.
21001	iv	cg			More
21133					Invalid
22187	vi	bc			More
22195					N exist
23533	iv	ag	ag, bg	ag, bg, cg	Fault
23535					Disc
23598	iv	bg	ag, bg		More

on the report number on NEVT and the year of occurrence. That way, the first two digits are the year and the three last digits the report number. Some electrical faults begin as a single phase event and later have more phases involved before the circuit breaker clears the fault, so more label columns have been added as shown in table I.

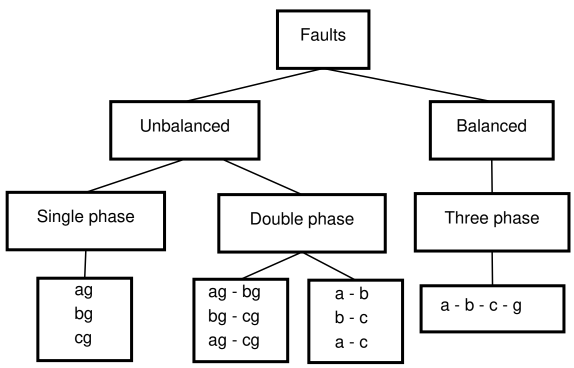

Figure 3. Fault classification.

Faults could be balanced (symmetrical) or unbalanced (unsymmetrical) in an electrical system. A balanced fault will involve all 3 phases, while an unbalanced fault may include a line to ground, a line to line or a double line to ground.

Figure 3 shows the classification used in this study and table I the labels assigned to example events. Label in column 'Label_1' on table I was given a reference label according to figure 3 and also a encoded label as '0' for single phase, '1' for double phase and '2' for triple phase. The classification treated here uses discrete labeling of classes.

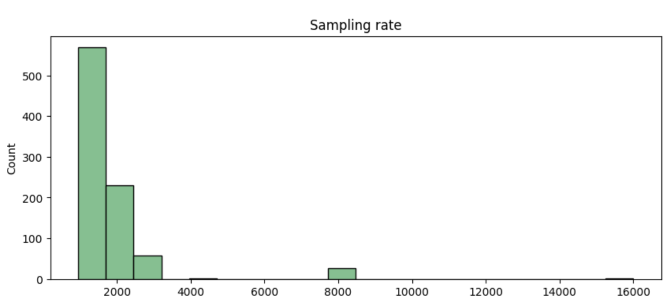

Figure 4. Data sample rate distribution.

B. Preprocessing

Oscillography is obtained from relays or automatic recording of disturbances instruments. They come from different brands and settings, so it is necessary to standardize all data, starting from sample rate which would the most basic feature of the equipment.

1) Resampling: Highest sample rate on the study base is 16000 samples per second. In figure 4 shows a histogram for sampling rate. In this study a sample rate of 1200 has been chosen since it is the most frequent.

C. Finding a trigger

Protective relays monitor current and voltages to protect an electrical system from abnormalities. Some relays require to monitor a window of half a cycle in the current and voltage signals to calculate a phasor and define if there is an issue on the protected element. Phasor addition of phase A, B and C currents is zero in steady state when all currents are balanced. When there is a fault unbalance occurs.

As a means of detecting a change in the wave shape, a couple methods where considered.

1) RMS: the rms value of a set of values is the square root of the arithmetic mean of the squares of the values, or the square of the function that defines the continuous waveform. To calculate it signal was squared, a rectangular window for averaging with defined length was created, moving average of the squared signal was computed with *np.convolve*, which performs convolution between the squared signal and the normalized window. 'Valid' mode was used so it only returns elements that don't require zero-padding, and finally square root of the result was calculated. So, RMS of the signal over a sliding window of defined length was calculated. A length of 20 samples was used in this study which is a cycle of 16.66 ms.

2) Hilbert: Hilbert function is available in scipy library. The envelope of a signal using the Hilbert transform was calculated. The Hilbert transform converts a real signal into an analytic signal. In signal processing this is useful for finding the instantaneous amplitude and phase. Absolute value (magnitude) of the complex analytic signal from Hilbert function is calculated to obtain the envelope of the original signal.

In this study, RMS is applied for each current signal and the maximum of these values is chosen as a reference. The purpose of this is to find a trigger point.

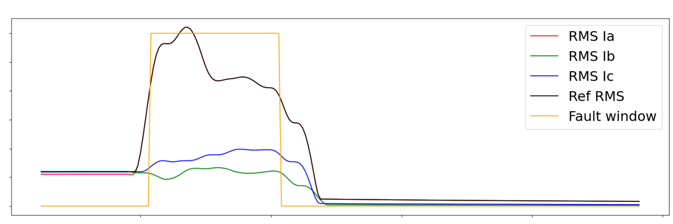

Figure 5. RMS Current representation

D. Cutting samples

The purpose of finding a trigger point is to have a reference point in order to have same length on every signal array. From this reference there will be a fault part to the right and a steady state to the left.

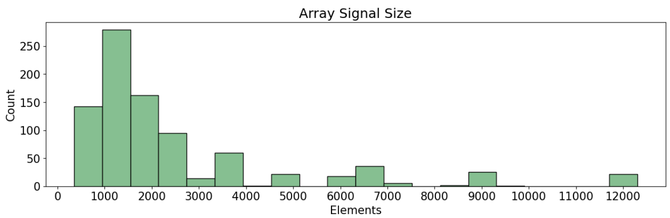

Figure 6. Signal array size distribution.

As shown in figure 6, array size is wide. Minimum of trigger encountered is 44 and the minimum distance from trigger to total length is 137 for just one event. Also, a data analysis to find a fault window length has been made. Dataset was filtered based on a clearly defined fault window obtaining 762 elements. As shown in figure 7 most faults have a length of 50 to 100 samples which would be 41.66 ms to 83.33 ms, which coincides with a length of 60 ms mentioned before.

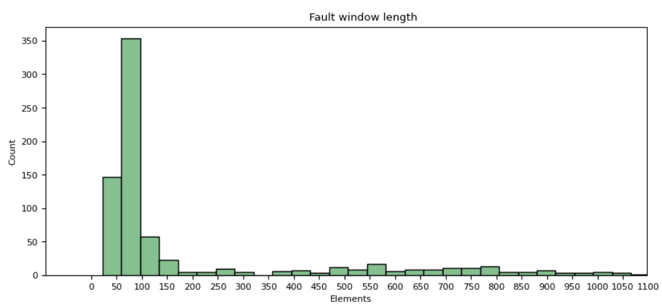

Figure 7. Fault window length distribution.

So, based on this analysis a steady state of length 43 and a 137 samples length for fault state was chosen to have a 180 length array for current and voltage signals.

E. Dataset filters

1) Format filtering: Some csv files obtained from PQDiffractor have twelve columns, they have a time column for each signal of current and voltage, while most csv files have seven columns, one corresponding to time and six for voltage and current signals. Because of this a script is run to identify csv files with twelve columns which would require and additional processing to standardize to seven columns.

2) Event filtering: As mentioned before, not all events are similar and this information was recorded in column *'Observations'* of table I. *'More'* means there are more oscillographies available for that event, *'Invalid'* means oscillography is not relevant to a fault. It may be noise, slight overload, transformer energization among others. *'Not exists'* means there is not any oscillography for that event, *'Fault'* is a normal fault event, *'Energization Fault'* means an element energization from a disconnected state to a fault and *'Disconnect'* means a recording of a disconnection where there was not any fault. A filter was applied to remove *'Energization Fault'* events because they didn't follow a wave shape coming from a flow of current different than zero.

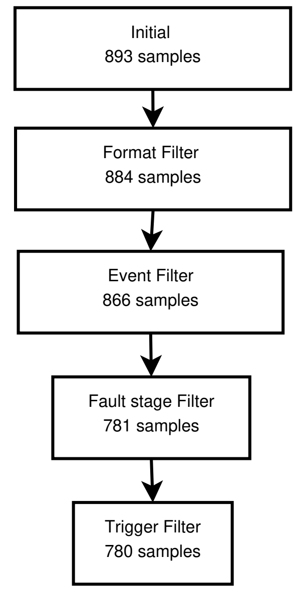

Figure 8. Applied Filter flow.

3) Fault stage filtering: As it can be seen on table I there are three columns for label. They contain the phases involved in a fault, for example 'cg' means a fault between phase C and ground while 'bc' means a fault between phase B and phase C. For event 23598 of table I, the fault initially involved phase B and ground, but later phase A joined in a second stage of the fault named as Label_2. A filter was applied to only work with events of Label_1.

F. Spectrogram

A spectrogram is a tool used to analyze and visualize a signal's frequency content as it varies over time. By transforming time-domain data into frequency-domain data new information about the signal can be found. An electric fault comes from a 60 Hz frequency wave, but due to increase on magnitude frequency content variation should be expected. A signal can be represented in two dimensions using a spectrogram, in which the color or intensity indicates the signal's magnitude at each frequency and time, which are the graph axis. Signal is divided into smaller frames that have some overlap with each other, a window function and Fourier Transform is applied and each of the frames resulting in a series of spectra that can be placed next to one another to create a spectrogram.

1) Windowing the signal: Entire signal is divided into overlapping windows. A window function is employed to reduce the impacts at the beginning and end of the sampled signal because Fourier Transform assumes that the signal is periodic and extends infinitely. Hann window from numpy

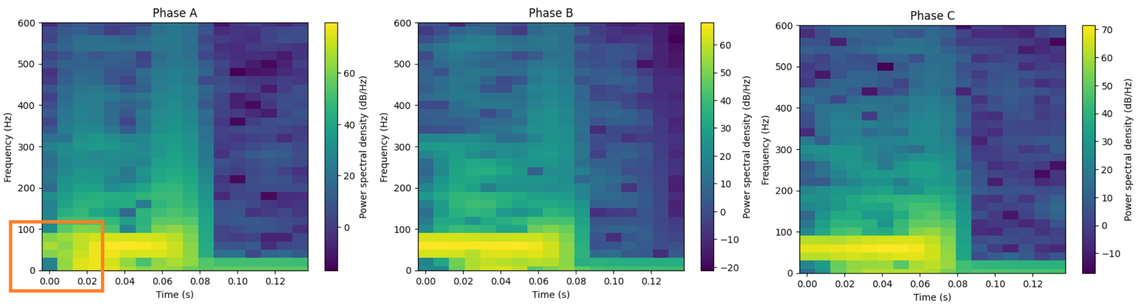

Figure 9. Single-phase fault current Spectrogram sample.

library is used to smooth discontinuities at the beginning and end of the sampled signal. In this work a sample frequency of 1200 samples per second is used, so each cycle is composed by 20 samples. A window of 60 samples is chosen with an overlap of 50. Overlapping is necessary to preserve edge information. To convert each window into the frequency domain Fourier transform is applied.

2) Power spectrum: After applying Fourier transform there is a complex number as a result, the magnitude of these numbers is calculated to get the power of the signal at different frequencies. By using *np.fft.fft* the final element of the output is the complex conjugate of the second element for real input, in function *np.fft.rfft* this symmetry is exploited to compute only the non-negative frequency terms which makes it faster.

Spectrogram is plotted as a heat-map where x-axis represents time and y-axis represents frequency, and the color represents the signal power. In figure 9 there is the spectrogram for each current of figure 2. In this case there was a fault on phase A, as we can see on the highlighted square of figure 9, there is a weaker power spectral density on the faulted phase spectrogram compared to the other phases.

G. Data Balance

On table II we can see the proportion of fault types, as expected most faults involve only one phase. Balance is important to prevent bias since unbalanced classes can cause the neural network to become biased towards the majority class. To evaluate minority classes a poor performance can be expected because of bias.

Table II
FAULT TYPE PROPORTION.

Label	Count Unbalanced	Count Balanced
Three phase	17	629
Two phase	140	630
One phase	623	623

A simple replication of existing examples is used as oversampling technique. The dataset is rebalanced to reflect more equal misclassification costs across all outcomes by oversampling the minority class cases. Considering that false negatives, or failing to discover the minority class, are far more expensive than false positives, or incorrectly classifying a sample as belonging to the minority class, it is important to accurately detect the minority classes when working with imbalanced datasets. Column Count Balanced on table II reflects the new proportion of events.

H. Neural Networks

A spectrogram is easy to process and represents the signal characteristics and it will be the input to a neural network. The input to the neural network will be 150 ms or 180 elements array. In this study a recurrent neural network is used. RNNs have a memory that will influence future predictions. LSTM network is a type of recurrent neural network (RNN) appropriate for the study of sequences in series data of time. The LSTM layer consists of gates that control the flow of information and memory cells, in this way processes sequential data and captures temporary dependencies, so LSTM network can learn long term dependencies between time steps of a sequence. CNN potential to learn various spectro-temporal patterns has made them perfectly suited for classification while LSTM has shown efficiency in time dependency learning. [20]

I. Modules

1) Dataset Module: On Dataset Class, we have 'initialization' that reads an imported dataframe previously prepared from which will read a label and apply spectrogram function to each signal of current and voltage for later stack them on an iterative manner for all items in dataframe.

From dataloader we can see a dimension of first layers of neural network. We have 6 channels, 31 bins and 12 frames.

2) Data Lightning Module: We deliver dataset to be trained. We define batchsize and num_workers. A division of 80 for training, 10 for validation and 10 for testing is applied so there is a dataset for each subset. Dataset is passed to dataloader

3) Architecture: CRNN is an architecture that combines CNNs and RNNs. Convolutional are useful to analyze image patterns. Feature extraction is the way CNNs recognize key patterns of an image in order to classify it. With spectrogram reduces dimension from 180 to 12. First layer receives 6 channels, 16 out channels since there is 16 filters and a kernel 3x3 with a padding of 1. Takes input

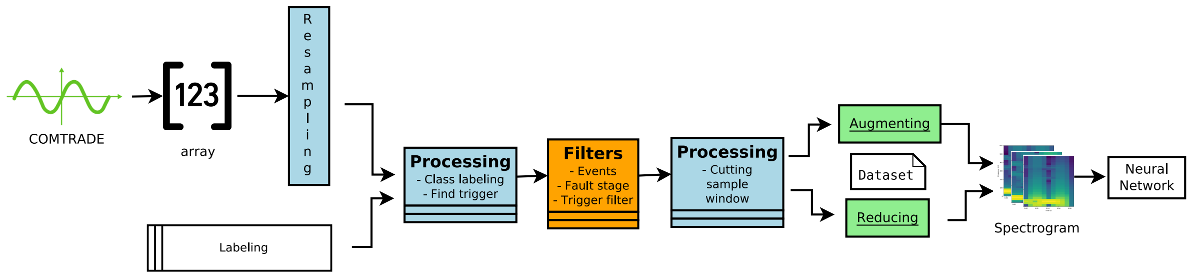

Figure 10. Experimental pipeline.

applies 16 filters, analysis input and feature map same size as input. Receives spectrogram and label, calls first convolution layer uses reLu as activation function and applies Maxpooling that extracts from each 2x2 patch the maximum value. ReLu has several benefits compared with conventional units such as effective gradient propagation and quicker computation rather than sigmoid units [20]. Second convolution layer receives 16 channels and output is 32 channels. Same kernel and padding as before is applied to keep dimensionality. Because of previous pooling the feature map is divided by 2. LSTM receives a one dimension sequence so a reshape is applied. Two LSTM layer are used with bidirectional. Output layer has activation function softmax because of multiclass classification consisting of 3 nodes. Last step of sequence has more information because gets information from previous steps.

4) Architecture Lightning Module: Metrics for Accuracy, precision, recall, f1 and Confusion Matrix are defined. Adam, Adaptive Moment Estimation, is used as optimizer because it performs better with RNNs like LSTM than gradient descent with momentum (SGDM). [1]

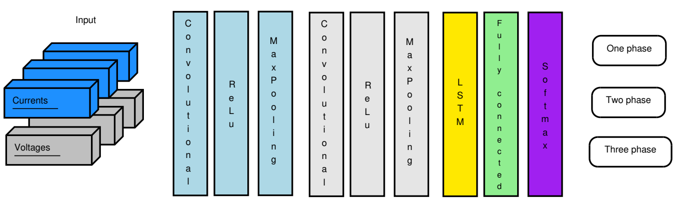

Figure 11. Applied Neural Network

A batch size of 16 is used, a learning rate of 0.0001, num workers of 2 and it is executed for 70 epochs taking into account an early stopping. Early stopping is a regularization technique that stops training if it detects overfitting or if there has been no significant progress in several consecutive epochs. The model was executed on Google Colab with T4 GPU hardware accelerator.

Project files have been uploaded to: https://github.com/faustoyg/Fault_classifier

IV. Experimental setup

On figure 10 there is a flow of the applied process. COMTRADE files are represented as arrays and joined to a labeled set through a common ID name to form a dataframe. The dataframe items are processed, and some filters are applied. Two experiments were performed, one by augmenting data by replicating samples and another by reducing samples of the class with the highest number of samples available. Each signal is represented as a spectrogram and will run through a CRNN to be identified as a single-phase, double-phase or three-phase event.

V. Results and Discussion

The evaluation metrics used to assess the model's performance included Accuracy, Precision, Recall and F1 score. Accuracy represents the number of correct predictions divided by the total number of predictions made by the algorithm. Accuracy treats all classes as equally important. A value of 0.96 was obtained for this metric. Precision measures how often a model's positive predictions are correct. It is calculated as the number of true positives divided by the number of true positives plus the number of false positives. F-score is calculated from the precision and recall of the test. Recall, is computed as the ratio of correctly predicted outcomes to all predictions.

Table III
Metrics of best model on test.

Label	Augmented	Reduced
Accuracy	0.96296	0.88095
Precision	0.96609	0.89815
Recall	0.96474	0.88888
F1 score	0.96538	0.88776

Figure 12 indicates that both training and validation accuracy increase over time and converge to high levels, suggesting the model is learning effectively on training data and generalizing well to the validation data. Also, both the training and validation loss decrease over time, with validation loss converging to a low level, suggesting the model's strong performance on the validation data.

The confusion matrix evaluates the performance of the classification model by comparing its predicted classifications against the true classifications. It has been encoded as 0 for single phase, 1 for double phase and 2 for three phase fault.

Confusion matrix in figure 14 show model is making relatively few mistakes with the majority of predictions falling along the diagonal. Per-class precision and recall metrics are quite high suggesting the model is performing well across different classes.

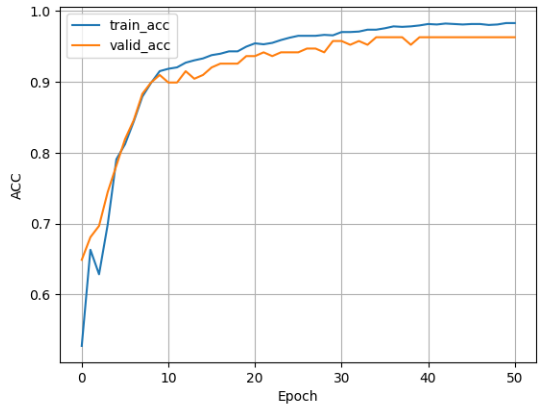

Figure 12. Accuracy vs epoch for augmented dataset

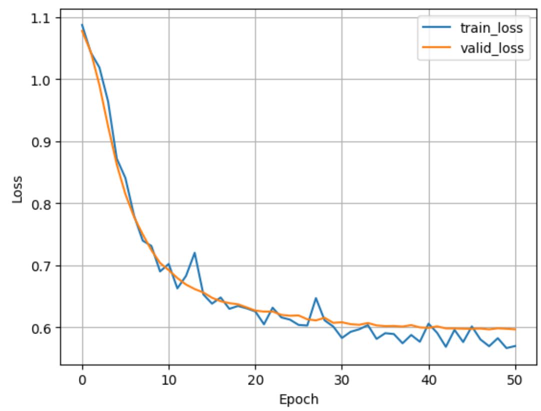

Figure 13. Loss vs epoch for augmented dataset

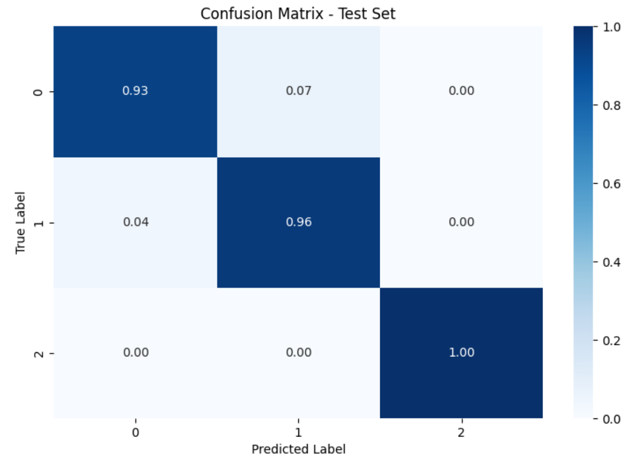

Figure 14. Confusion Matrix in augmented dataset - Test.

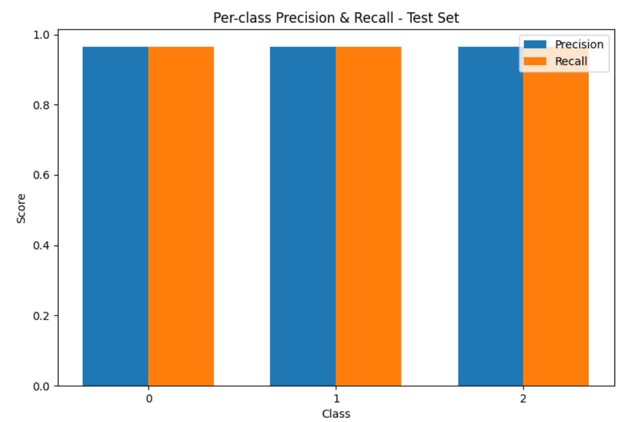

Figure 15. Per-class Precision & Recall - Test.

A simple replication oversampling consists of duplicating minority class examples that already exist rather than creating new ones, so the new samples may only increase noise in the training data rather than offering more insightful information about underrepresented classes. In this case the minority is a three phase fault which is a balanced fault. Because of this it is considered in this study that there is not a wide range of variation for this type of fault. Other data augmentation techniques could be implemented as masking spectrogram to evaluate results variation.

The model was executed again for a reduced dataset. Single phase elements were reduced to same size as double phase elements, having 140 for each class. This because of applied replication that could potentially have same elements on training set as well as in testing set. Resulting confusion matrix is shown in figure 16.

VI. Conclusion

In this work a dataset has been built and an artificial neural network has been applied for the detection and classification of faults on a three phase power transmission line system. Good information coverage with low redundancy can be obtained with a dataset with sufficient and good quality training data. Using real data compared to simulated data allows to replicate reality on a more accurate way. For example, fault duration was intuitively known, by examining the dataset fault duration now has a quantitative value and proportion compared to other fault lengths. This could avoid generalizing values or making wrong assumptions. Fault classification method used required a neural network that determines it from the patterns of pre fault and post fault voltages and currents and working with arrays facilitated this task because of flexibility to change size. Resizing by cutting the waveforms not only standardized the data but also reduced training computation time.

Oversampling is used to avoid the classification bias that occurs because of most faults being single phase, however, a proper technique must be implemented to avoid replicating too many samples.

Training a model using feature extraction has good classification performance. The representation of the spectrum of a signal in the frequency domain can help to better understand its content than with a representation in the time domain and also is a good representation of the characteristics.

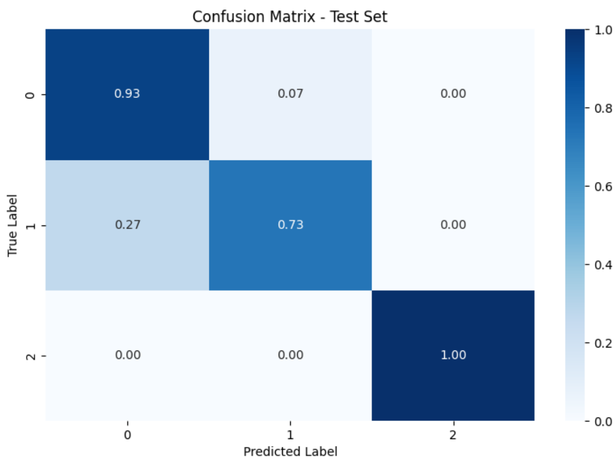

Figure 16. Confusion Matrix in reduced dataset - Test.

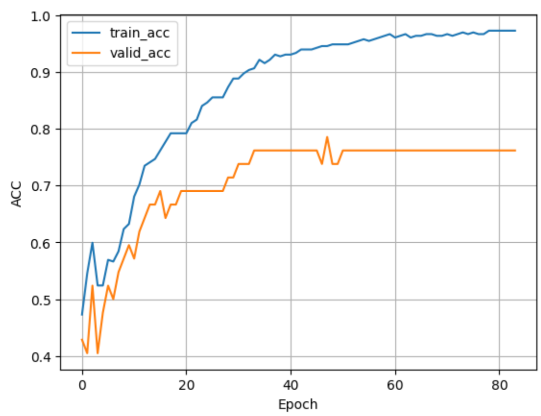

Figure 17. Accuracy vs epoch for reduced Dataset

Early stopping saves computational resources and training time and avoids overfitting by focusing on epochs that truly benefit the model.

Python has many libraries which increase their application capabilities. One identified library that may be helpful is pyComtrade, which could make simpler the method to review oscillographies.

A more accurate way to identify fault duration can help analyze based on data the current relay performance. This is possible by having a dataset with arrays that can be handled and processed.

References

[1] MATLAB & Simulink. Classify ECG Signals Using Long Short-Term Memory Networks. [Online]. Available: https://www.mathworks.com/help/signal/ug/classify-ecg-signals-using-long-short-term-memory-networks.html

[2] CELEC EP Transelectric, NEVT

[3] Rosero Karen, CRNN usando Pytorch ligthning notebook, 2024

[4] Iqra Muhammad, Medium. Time Series Forecasting using LSTM: An Introduction with Code Explanations. [Online]. Available: https://medium.com/@iqra1804/time-series-forecasting-using-lstm-an-introduction-with-code-explanations-c5c2e8ca137d

[5] Espectrograma - Wikipedia, la enciclopedia libre [Online]. Available: https://es.wikipedia.org/wiki/Espectrograma

[6] George V Jose, Towards Data Science. Useful Plots to Diagnose your Neural Network. [Online]. Available: https://towardsdatascience.com/useful-plots-to-diagnose-your-neural-network-521907fa2f45

[7] Scicoding, How to do Spectrogram in Python. [Online]. Available: https://www.scicoding.com/how-to-do-spectrogram-in-python/

[8] NumPy, Discrete Fourier Tranform numpy.fft.rfft, [Online]. Available: https://numpy.org/doc/stable/reference /generated/numpy.fft.rfft.html

[9] Geitgey Adam, Medium. Machine Learning is Fun Part 6: How to do a Speech Recognition with Deep Learning. [Online]. Available: https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a

[10] Ashraf Abdallah, Medium. Oversampling - Handling Imbalanced Data. [Online]. Available: https://medium.com/@abdallahashraf90x/oversampling-for-better-machine-learning-with-imbalanced-data-68f9b5ac2696

[11] Sathya Prakash, Github repository, [Online]. Available: https://github.com/KingArthur000/Electrical-Fault-detection-and-classification

[12] Utkarsh Pandey, Anshumann Pathak, Applications of artificial intelligence in power system operation, control and planning: a review, Clean Energy, 2023

[13] Chakraborty Sriya, Wavelet transform based fault detection and classification in transmission line, IJREAS, 2012

[14] Jamil Majid, Fault detection and classification in electrical power transmission system using artificial neural network, Springer Plus, 2015.

[15] Karic Almin, Power System Fault detection, classification and location using Artificial Neural Networks, 2017

[16] Mnyanghwalo Dausi, Deep learning approaches for fault detection and classification int the electrical secondary distribution network: Methods comparison and recurrent neural network accuracy comparison.

[17] Tirnovan Radu, Advanced techniques for fault detection and classification in electrical power transmission systems: An overview, 2019

[18] Esen Vedat, The determination of short circuits and grounding faults in electric power systems using time-frequency analysis, Journal of Energy in Southern Africa, 2015

[19] Kezunovic Mladen, Detect and Classify Faults using Neural Nets, IEEE Computer Applications in Power, 1996

[20] Das Joy, Ghosh Arka, Urban Sound Classification using convolutional Neural Network and Long Short Term Memory based on Multiple Features, IEEE, 2020

[21] Ferreira Leonardo, Barrozo Viviane, Short Circuit Classification using the Discrete Fractional Fourier Transform and Artificial Neural Network, JAERS, 2021

[22] Belagoune Soufiane, Bali Noureddine, Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems, Elsevier, 2021

Appendix A
Exported Waveform

Figure 18 shows exported waveforms resulting from an oscillograph using different software. PQDiffractor is more stable compared to other software. As it can be seen Sigra added a DC component and scaled current and voltage signals for a single phase fault. For a double phase fault

Sigra and WaveEv give a different scale to faulted phase currents. PQDiffractor was found to export waveforms better.

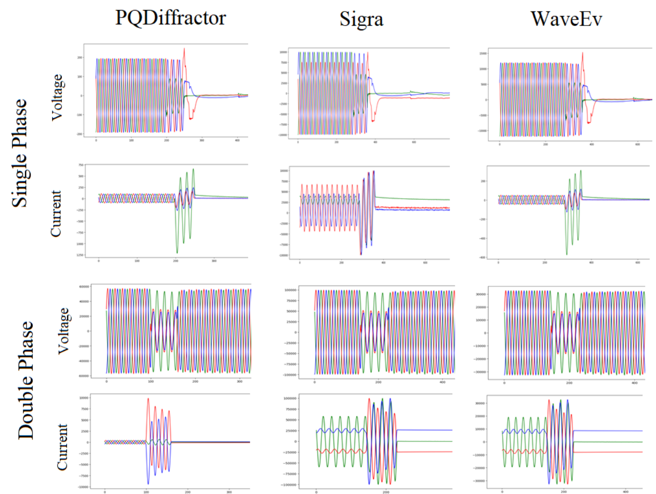

Figure 18. Exported Waveforms

Appendix B
Array length and spectrogram parameters

Different array lengths and the parameters window size and overlap were tested on same single phase fault waveform in order to compare and define which parameter describes better waveform characteristics. In this case phase A is the faulted phase and it can be seen in figure 19 the differences on spectrogram representation compared to healthy phases.

Appendix C
Fault waveform characteristics

Figure 2 was simulated on ATPDraw, a graphical pre-processor to Electromagnetic Transients Program used to create and edit circuit files, in order to compare spectrograms changing fault parameters like duration and fault impedance. Results are shown in figure 20.

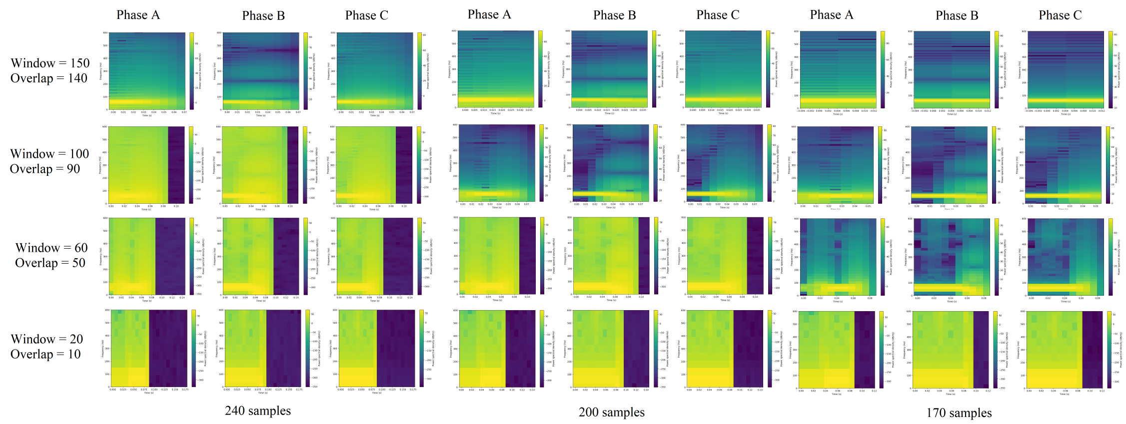

Figure 19. Spectrogram for different sample sizes and window parameters.

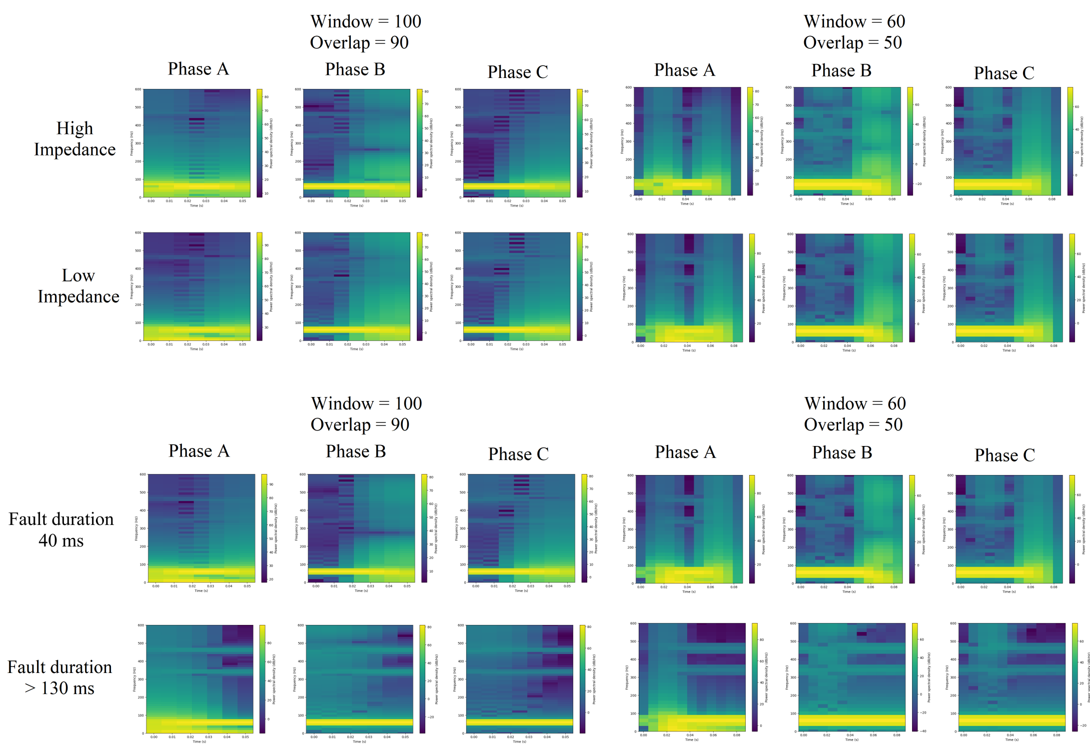

Figure 20. Spectrogram for different fault parameters.