FINGERPRINTING MADE EASY BY MACHINE LEARNING

Sathiswar JAYASEELAN
TU/E – The Netherlands
s.jayaseelan@student.tue.nl

Albert PONDES
Enexis – The Netherlands
albert.pondes@enexis.nl

Armand van DEURSEN
TU/E – The Netherlands
A.v.Deursen@tue.nl

Han SLOOTWEG
TU/E – The Netherlands
J.G.Slootweg@tue.nl

ABSTRACT

Circuit breakers are used to protect electrical networks and components from damages caused by high currents, resulting from overloads or short circuits. Malfunctioning of Medium Voltage (MV) circuit breakers often leads to a longer outage time and the outage spreads to broader areas. Therefore, it is of great importance that the circuit breakers function properly. In order to test the proper working of MV circuit breakers, manual periodic measurements are carried out and the results are interpreted visually by the operator. This process requires much time, money and workforce. This paper discusses the possibility of automating the circuit breaker measurement system and subsequently the data analysis. An economical new tool was developed for automating the measurement system and a novel model was developed using machine learning for automating the data analysis process.

INTRODUCTION

Circuit Breakers (CBs) are one of the most important components in the power system. They are used to isolate faulty parts of the network during overloads or short circuits. There are designated CBs for each area, which will trip whenever there is a fault in that particular area. In case the CB does not trip within the required time, it will result in the backup CB, which covers a larger area, to trip. This will result in the outage spreading to a broader area. Under such circumstances, not only the number of affected customers will increase, but also the time needed for identification and repair of the fault will increase. A large amount of Customer Minutes Lost (CML) is the result.

In order to ensure the proper working of the CBs, workforce is periodically sent to the substation for doing manual measurements. Enexis, one of the major Distribution System Operators (DSO's) in Netherlands, has around 25,000 MV CBs installed in their grid. The sheer amount of CBs makes it very difficult for performing the measurements and analysis frequently with the currently used method. The cycle may be as long as three years, which leaves room for a number of early signs of deteriorating performance to go unnoticed. The workforce carry portable finger printing equipment to the substation, where they perform a secondary injection test and measure the current flowing through the CB. Based on the waveform obtained, the results are interpreted visually. This process is tedious and subject to interpretation and in particular to the expertise of individuals involved [1].

The aforementioned problems led to the idea of automating both the measurement system and data analysis of CBs by creating an efficient and advanced automated tool and model. The feasibility of designing a new tool which is capable of measuring current at high sampling frequency, was explored by developing a prototype using a 10-bit Analog to Digital Converter (ADC) and shunt resistor. The newly designed tool is economical and can be placed at all the substations permanently for automating the measurement. Hence, there is no need of workforce for doing the periodic installation and measurement. The possibility of automating the data analysis for finding the failure modes was explored by developing a novel model based on machine learning with high prediction accuracy. The next sections discuss the following topics: Section 2 will discuss the new tool for automating the measurements. Section 3 will discuss the model for automating the analysis and this papers ends with conclusion

TRIP CURRENT MEASUREMENT AUTOMATION

Existing Device

The fingerprinting equipment currently used for measurements consists of a current sensor, secondary current injector (megger) and a monitor for displaying the results (profiler). Workforce carry this portable equipment to the substation for doing periodic measurements and connect it to the trip coil of the CB to be tested. The CB is disconnected from the grid during the test measurement. The current sensor used is of the Hall effect type and is clamped to the wire near the trip coil. A trip is simulated by injecting current greater than the trip current to the secondary side of the trip coil. The data obtained during the tripping is collected and processed using a profiler. After processing the data, the profiler displays the obtained waveform in graphical format and the results are visually interpreted by the workforce. This process is called fingerprinting measurement, because under normal operation of CBs, the waveform obtained will be similar to that of a standard pattern, whereas under abnormal condition the pattern differs.

New Device

The existing device consists of a Hall effect sensor and a profiler for measuring the current. Enexis has around ten of these measurement devices per province. The Hall effect sensors used are expensive, and hence it is not a wise idea to permanently place these sensors at each substation to automate the current measurements. The circuit diagram of the newly designed tool is given in Fig. 1. A shunt resistor is connected in series with the CB as shown in Fig.1. The main idea of this design is to measure the voltage drop across the shunt resistor during the tripping of CB, which is then converted to current by dividing with resistance, according to Ohm's law. The shunt resistor used for this tool has a resistance of 0.1 and is rated at 10 watts.

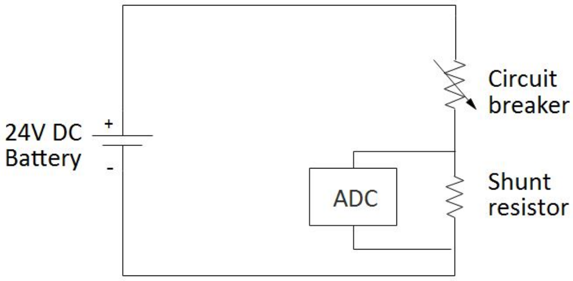

Fig. 1: Circuit diagram for measuring current using the new tool.

The measured data from the ADC (at 20 kHz) is first written to the Sequential Random Access Memory (SRAM) of the ADC and then transferred to a Secure Digital (SD) card by using an interrupt routine [2]. This is done in order to achieve high sampling frequency. The data from the SD card is transferred to a Structured Query Language (SQL) database using a sim800L Global System for Mobile (GSM) module. The architecture of the new tool is given in Fig. 2.

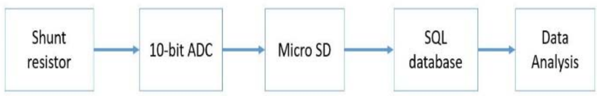

Fig. 2: Automated tool for trip current measurement.

This new tool was validated by performing test measurements at the substation. For performing this measurement both existing and new tool are used. In addition to this a high accuracy 1 GHz, 12-bit Digital Storage Oscilloscope (DSO) is also used. For measuring the current from the trip coil side all the devices (DSO, profiler and the new tool) are connected near the trip coil. The results obtained from all the three devices are given in Fig. 3, which shows that the waveform obtained from the new tool is almost similar and accurate as compared to DSO with little fluctuation in the new tool waveform. This noise which is very negligible is expected to be because of the induced current from the components used in the new tool and also the quantization noise of the ADC itself. The newly designed tool is very economical and can be placed at the substations permanently for automating the measurements.

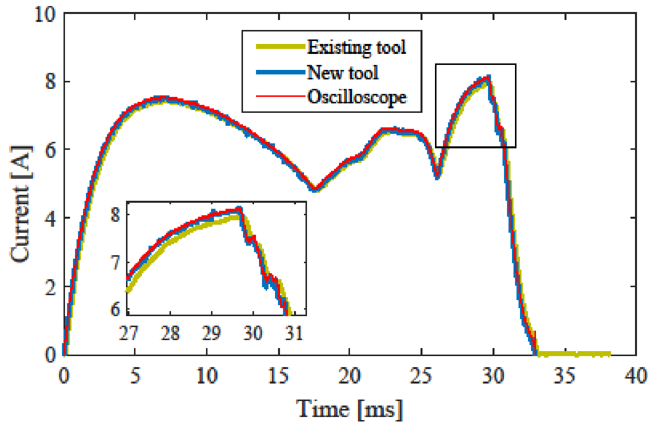

Fig. 3: Measurements recorded near trip coil. The inset shows the small differences in waveforms.

DATA ANALYSIS AUTOMATION

For automating the data analysis, various features are extracted from the existing datasets obtained with the currently used fingerprinting method and processed with different machine learning algorithms. The algorithm with highest prediction accuracy is chosen for creating the model and the results are tested with new datasets. A two-step model is proposed for getting the highest accuracy for analysis. The next subsection discusses about the various failure modes in the circuit breaker, followed by machine learning for analysis automation.

Failure Modes

The standard pattern of the trip current waveform obtained under normal working condition of the CB is given in Fig.4. During the time period T1, the current starts to flow in the trip coil and this exerts a force on the core of the coil. During the next time period T2, the roller bearings starts rolling and during the time period T3 the contact opens. In case of any abnormal operation, the pattern of the waveforms differs. There are seven known failure modes that are taken into account while creating the model but only two will be explained in detail.

.

Roller bearing problem: In order to move the core during the tripping action, roller bearings are used. The roller bearings used contain oil and grease. The oil comes out of the roller bearing over time due to oxidation and this results in hardening of the grease. The movement of the roller bearing will no more be smooth and if left untreated the CB might not trip within the stipulated time. A waveform associated with this phenomenon is shown in Fig. 5.

Relay contact problem: This problem occurs when the spring attached to core of the trip coil gets corroded or worn out. During the switching action of the trip coil, the contact is actually made and broken as shown in Fig. 6, which has a medium contact problem. If the medium contact problem is left untreated, it leads to heavy contact problem. If the heavy contact problem is left untreated, it might increase the time taken for the tripping of the CB.

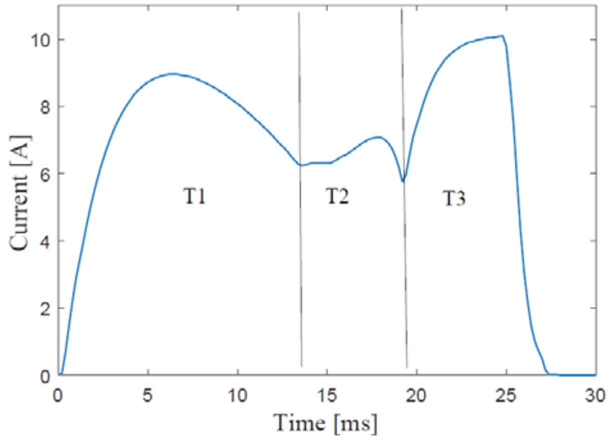

Fig.4 Trip current pattern under normal operation of CB.

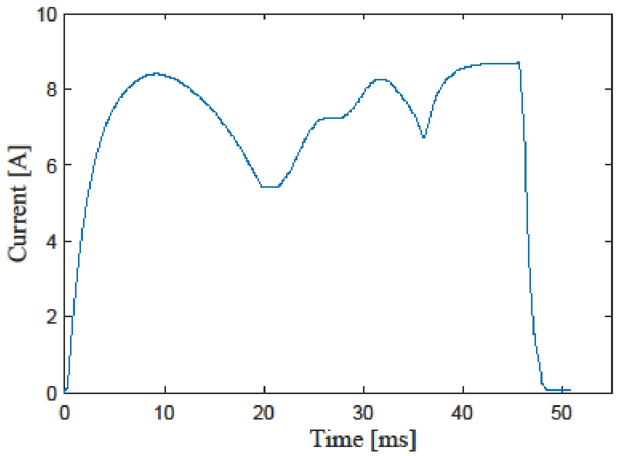

Fig.5 Trip current waveform during roller bearing problem. The pattern and duration of the time period T2 differs when compared to Fig. 4.

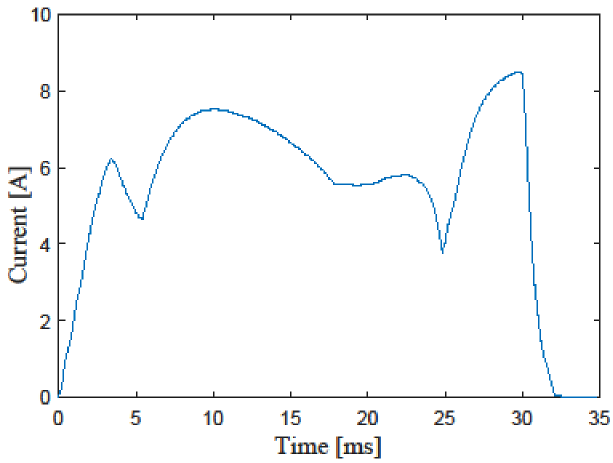

Fig.6 Trip current waveform during relay contact problem. The current rise during the time period T1 is not linear as compared to Fig. 4.

Machine learning for analysis automation

Machine Learning (ML) is defined as an automated process that extracts patterns from data for doing the predictive analysis [3]. In this paper, a novel model is designed using supervised classification learning. Supervised machine learning techniques automatically learn a model of relationship between a set of descriptive features and a target response based on the existing datasets. The working of the supervised learning algorithm is given in Fig. 7.

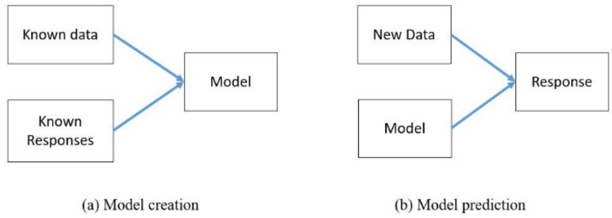

Fig.7 Supervised Machine Learning

The main challenge in ML is selecting the correct features from the datasets. The prediction accuracy of the model is directly proportional to the correct features selected [4]. Once the features are selected along with their target response, the ML model is created. New data is then fed into the ML model for doing the prediction. The working of the machine learning algorithm is given in Fig. 8. High prediction accuracy can be obtained by ML algorithms if large datasets are used for creating the model.

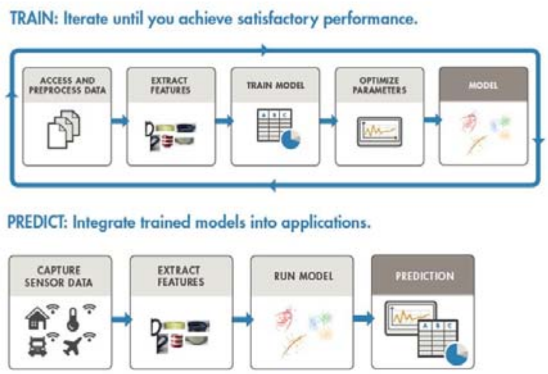

Fig.8 Machine Learning Process [5]

Analysis model

In most literature, a single step approach is used for creating the ML model [6-8]. This depends on the number of responses. If the number of responses are less like classifying the Electro Cardiogram (ECG) signal into normal or abnormal, then a single model approach is sufficient. For mechanical type CBs we have identified seven failure waveforms along with one normal waveform, which makes the number of responses to eight. Hence, a single model approach will have less prediction accuracy as given in Table 1. In this paper, a novel two step approach is proposed using three models (i,j,k) as given in

Fig. 9. The failure forms are categorized into two types: type 1 and 2. The type 1 category consists of failures because of the following reasons: magnetizing problem, rubber bouncing and less space between coil and yoke. The type 2 category consists of failures because of the following reasons: long time taken for tripping, roller bearing problem, medium and heavy contact problem in relay. Whenever there is a new dataset, features are extracted from it and fed into the model i, which will categorize it as normal, type 1 or type 2. If the response is type 1 or type 2, then the data is fed to model j or model k respectively for finding the exact reason for failure.

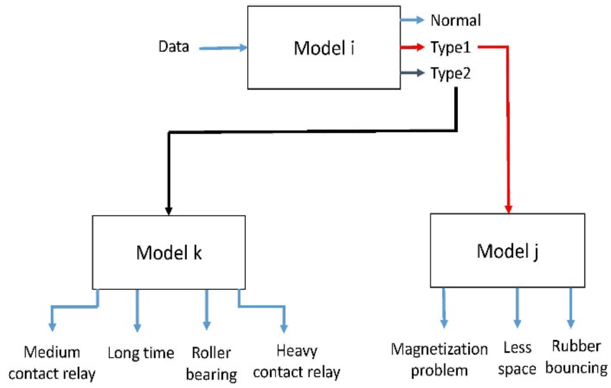

Fig9. Automated two step analysis with three models i, j and k.

Feature Extraction

Feature extraction is one of the most important steps in creating the machine learning model. Selecting correct features from the datasets enhances the prediction accuracy of the model. Same features are used for all the three models but different algorithms are used. This is because of the working mechanisms of the various algorithms considered for creating the model. The various features that are used for creating the model are [9]:

(i) Mean and Median
(ii) Standard Deviation and Mean Absolute Deviation
(iii) Percentiles at 25, 50 and 75
(iv) Skewness and Kurtosis

Algorithms considered

The preprocessed data with the extracted features is imported to the classification learner toolbox for creating the model. The dataset is compared with different ML classification algorithms and the one with highest prediction accuracy is selected for creating the model. Cross validation technique is used for determining the prediction accuracy. This validation scheme will give feedback on how the model will perform, while predicting the new datasets. The various algorithms considered for creating the model are:

(i) Support Vector Machine
(ii) Bagged Ensembles
(iii) K nearest neighbor
(iv) Decision Trees

The working principle of the above mentioned algorithms can be found in [3, 10].

Results of single step approach

A single step approach as given in the literature [6-8] was modelled for comparing the results with the two step approach. For designing this model, the parameters used are: 176 Observations with the same 250 features as used for two step approach. Eight responses, which are categorized into 36 normal samples, 20 samples each for all the other categories. After computation using cross validation, the prediction accuracy obtained for different machine learning algorithms is shown in Table 1.

Algorithm	Prediction accuracy
Cubic SVM	**85.8%**
Bagged Ensembles	85.5%
KNN	80.1%
Decision Trees	77.7%

Table 1. Prediction accuracy of single step model

Results of novel two step approach

The ML models are created based on the prediction accuracy obtained using cross validation. Cross validation signifies how the model will perform with the new datasets. The results discussion starts first with the results of model i and is followed by model j and k.

Model I**:** For designing model i, the parameters used are: 176 observations with 80 type 2 samples, 60 type 1 samples and 36 normal samples. 250 Feature predictors and 3 responses are used for this model. After computation using cross validation, the prediction accuracy obtained for different machine learning algorithms is shown in Table 2.

Algorithm	Prediction accuracy
Cubic SVM	**96.6%**
Bagged Ensembles	89.3%
KNN	89.4%
Decision Trees	83.4%

Table 2. Prediction accuracy of model I

Model J: For designing model j, the parameters used are:60 observations with 20 magnetization problem samples, 20 rubber bouncing samples and 20 less distance between core and yoke samples. 250 Feature predictors and 3 responses are used for this model. After computation the prediction accuracy obtained for different machine learning algorithms is shown in Table 3.

Algorithm	Prediction accuracy
Cubic SVM	83.3%
Bagged Ensembles	**91.6%**
KNN	80%
Decision Trees	86.3%

Table 3. Prediction accuracy of model J

Model K: For designing model k the parameters used are: 80 observations with 20 medium relay contact problem samples, 20 large operation time samples, 20 roller bearing problem samples and 20 heavy relay contact problem samples. 250 Feature predictors and 4 responses are used for this model. After computation using cross validation, the prediction accuracy obtained for different machine learning algorithms is shown in Table 4.

Algorithm	Prediction accuracy
Cubic SVM	95%
Bagged Ensembles	**96.3%**
KNN	80%
Decision Trees	94.3%

Table 4. Prediction accuracy of model K

Overall accuracy and testing: The total overall accuracy of the model is 92.3%. The maximum prediction accuracy as achieved in the literature and single step approach is around 80-85% [6-8]. Using a novel twostep approach and extracting correct features, an accuracy of greater than 92% is achieved. In the next step the models are exported and compared with the new datasets. Overfitting occurs when the prediction model selected by the algorithm is so complex, such that the model fits to the dataset too closely and becomes sensitive to the noise in the data. Hence, it will result in a decreased prediction accuracy on predicting the new datasets. For checking the newly created model for overfitting, a collection of 25 new random samples are preprocessed. After extracting the features the data is fed into model i and then to model j and k respectively, depending upon the results of model i. The overall model is able to predict 23 datasets correctly. From model i, one dataset of type 1 magnetizing problem is classified as type 2. From model j, one rubber bouncing is classified as magnetization problem. The model k predicted all the results correctly. This proves that the automated model does not suffer from the problem of over-fitting.

CONCLUSION

This paper presented a new tool for automating the trip current measurements and a novel model for automating the data analysis. The trip current measurement process was automated by designing a new tool with a shunt resistor and a 10-bit ADC. The performance of the newly designed tool was validated by doing test measurements at the substation. A novel two step approach was implemented by using ML for automating the data analysis. Experimental results shows that an accuracy of 6% higher is achieved using the proposed two step approach when compared to the single step approach. The new tool is economical and can be designed at a very low cost, hence it can be placed at all the substations permanently for automating the measurements.

If the new tool and model are implemented in the DSO's operation, it will not only reduce the cost and time taken for the measurements but also reduce the human errors in data analysis. The time interval between the measurements can also be reduced, for detecting a number of signs of deteriorating performance at earlier stage and thus preventing CB failure, saving the DSO's CML, thus money. The new tool will also record the trip current information of the CBs during real time tripping, from which analysis can be done for evaluating the CB performance. Thus the new tool and model will reduce the CML by a greater extent.

REFERENCES

[1] M. Knezev and Z. Djekic, "Automated circuit breaker monitoring," IEEE, 2007.

[2] M. Noor and M. Omar, "Design of data logger with multiple sd cards," IEEE conference on clean energy and technology, 2013.

[3] J. D.Kelleher and B. mac Namee, Fundamentals of machine learning for predictive data analytics : algorithms, worked examples, and case studies, 2015.

[4] M. C. Popescu and L. M. Sasu, "Feature extraction, feature selection and machine learning for image classification: A case study," IEEE, 2014.

[5] Mastering Machine Learning - A step-by-step guide with MATLAB, 2018.

[6] P. Dabul and S. Patil, "Skin cancer detection and classification," IEEE, 2017.

[7] G. Kesavaraju and S.Sukumaran, "A study on classification techniques in data mining," IEEE, 2013.

[8] A. Singh and A. Sharma, "A review of supervised machine learning algorithms," IEEE, 2016.

[9] Y. ben Youseef and J. Zbitou, "Statistical features and classification of normal and abnormal mammograms," IEEE, 2014.

[10] A. Bifet and G. Holmes, Machine Learning for Data Streams, 2017.