APPLICATION OF MACHINE LEARNING TECHNIQUES FOR AUTOMATIC ASSESSMENT OF FRA MEASUREMENTS

J.L. Velásquez Contreras[1*], M.A. Sanz-Bobi[2], S. Banaszak[3], and M. Koch[1]

[1]Omicron Electronics GmbH, Oberes Ried 1 6833 Klaus, Austria
[2] Comillas Pontifical University, IIT, Santa Cruz de Marcenado 26 28015 Madrid, Spain
[3]West Pomeranian University of Technology, ul. Sikorskiego 37 70-313 Szczecin, Poland

*Email: juan.velasquez@omicron.at

Abstract: The Frequency Response Analysis (FRA) is an advanced method for diagnosis of failures in the active part of power transformers. The assessment of FRA results relies on the comparison of a reference FRA curve to an actual curve and based on the deviations between the two curves and the experience of a human expert, an assessment about the integrity of the components of the active part is issued. At the present there is a lack of reliable algorithms for automatic assessment of the results. This motivated to research new methodologies for overcoming this problem. As a contribution to this necessity, this paper summarizes the outcome of a research work in which machine learning algorithms were used for automatic assessment of FRA measurements. Decision tree classifiers were developed using the algorithm C4.5. The results obtained give evidence of the effectiveness of the proposed classifiers.

1 INTRODUCTION

In general, the problem of the assessment of FRA measurements can be formulated as the capability of a human expert or a computational tool of understanding the meaning of the deviations between a pair of FRA traces in such a way that the deviations can be classified to a specific condition of the transformer (e.g. healthy, short-circuit between turns, mechanical deformation, etc.). The typical workflow followed by non-expert users when assessing FRA results is next described.

First, users perform the assessment of the measurements using the assessment algorithms implemented in the software of commercially available FRA instruments (for example, the algorithm of the standard DL/T911-2004). However, the assessment provided by the standard DL/T911-2004 only provides a general diagnosis (severe deformation, obvious deformation, slight deformation and normal winding), but an explanation about the diagnosis is missing and neither information about likely causes or recommendations is provided. For this reason, in the majority of the cases, the questions of the users are not answered by the assessment algorithm and the users contact a human expert in order to ask for assistance in the interpretation of the results.

Machine learning (ML) techniques have been widely applied in the field of diagnosis due to their classification capabilities. Surprisingly, in the field of FRA very few applications of these techniques have been found in the literature. Probably, this is due to the lack of data required for training the algorithms. Among the few attempts to the use of ML for assessment of FRA measurements, the works [1-2] can be quoted. The work presented in [1] illustrates the use of a feed-forward back-propagation ANN consisting of three layers (input, hidden and output) for automatic assessment of FRA measurements. Correlation coefficients of the plot of the magnitude at three frequency sub-bands (LF, MF and HF) and in the whole frequency range were used as input to the ANN. The drawbacks of this work are: the ANN was trained using only 16 samples and only 10 samples were used for validating it and only two classes were defined "deformation", "no deformation". While [2] shows the use of self-organizing maps (SOM) for classification of failures that could arise after subjecting power transformer to the impulse test. This test relies on the fact that the neutral current obtained at full test voltage should be the same as the neutral current obtained at reduced test voltage. In order to avoid any dependence of the applied voltage, the transfer function was used for the analysis of the results. A transformer model was used for simulating deformations on different sections of the winding for creating the required training set for training the SOM.

The main advantage of the use of ML techniques is the possibility of establishing interpretation rules for the assessment of the results. Actually, taking into account that the problem behind the assessment task is not more than a classification task, it is considered that the problem of the interpretation should be addressed with the use of ML techniques. Independently on the kind of indicator used for analysis of FRA results (statistical, physical parameters, etc.), a classifier should be used for allowing the users to understand what a change in the value of an indicator means. In this sense, it is considered that the heuristic should not be totally avoided. All the opposite, because of the complexity of the problem behind the interpretation of FRA measurements, the experience collected world-wide with FRA measurements of real case studies of known failures should be used. In this manner, a combined approach using physical modelling together with ML techniques seems to be an optimal combination.

This contribution summarizes the outcome of a research work done by the authors regarding the application of ML algorithms with the aim of extracting knowledge rules under which automatic assessment of FRA measurement can be carried out. In section 2 an introduction to ML is provided and in section 3 the methodology used for the development of the ML algorithms is described. Then, in section 4, the results are presented and in section 5 the application of the extracted knowledge is illustrated by means of real cases. Finally, in section 6 conclusions are provided.

2 MACHINE LEARNING

Machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data, such as sensor data or databases. A learner can take advantage of examples (data) to capture automatically characteristics of interest that they include. Data can be seen as examples that illustrate relations between observed variables. A major focus of machine learning research is to automatically learn for recognizing complex patterns and making intelligent decisions based on data. The difficulty lies in the fact that the set of all possible inputs is too large to be covered by the set of observed examples (training set). Hence, the learner must generalize from the given examples, so as to be able to produce a useful output in new cases.

Machine learning algorithms are frequently called classifiers or pattern recognition algorithms. Classifiers are not more than a machine learning algorithm that syntactically learns simple string rules to guide its performance in an arbitrary environment. This consists in the assignment of some sort of output value (or label) to a given input value (or instance), according to some specific algorithm.

Another term commonly used in the field of machine learning is "data mining (DM)". DM is defined as the process of extracting knowledge from a database for the creation of a knowledge base, which subsequently is used for solving problems. In its basic essence, DM is the application of ML algorithms to find patterns in data. The goal of DM is to find patterns which are not predicted by established theory. In other words, DM can be defined as an approach of extracting full value from data stored in databases [3].

3 METHODOLOGY

The steps to be followed for the design of a classifier using ML algorithms are illustrated in Figure 1. The process consists of five steps: data preparation, pre-processing, training of a ML classifier, analysis of performance and validation. If the performance of the classifier against training and/or against validation is not satisfactory, it could be due to poor data preparation, due to poor feature extraction, or just due to the fact that the patterns of some classes do not distinguish enough from each other. In this late case it is recommend re-defining the classification goals.

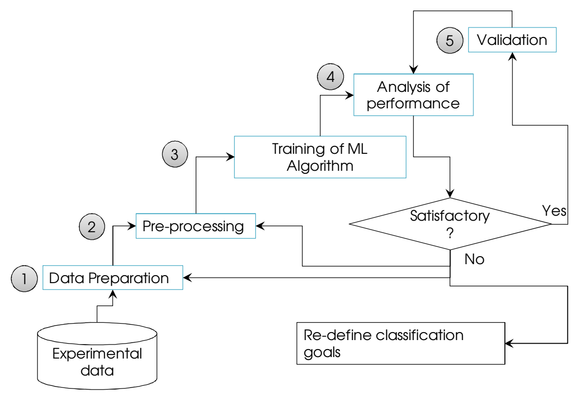

Figure 1: Steps for the development of classifiers

3.1 Experimental data

A database consisting of FRA measurements performed from 2006 till 2011 in more than 500 power transformers of different designs, sizes and different manufactures was used as experimental data for the development of algorithms of automatic assessment. Figure 2 shows an overview of the different types of transformers in the database. Additionally, the FRA data of winding deformations simulated in real transformers was also used.

As part of the data preparation stage, the FRA results of each transformer were analyzed by human experts, who assigned to each transformer a pre-defined output or diagnosis (healthy, short-circuit between turns, mechanical deformations, etc.). In the field of ML these possible outputs used to be called "classes". Particularly, for the research work presented in this paper, only the classes presented in Table 1 were used for training the ML algorithms.

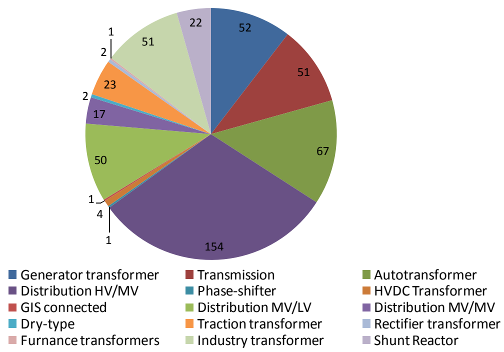

Figure 2: Overview of transformers

Table 1: Description of the training set

Description	Class	N° instances
Healthy winding under the same remanence condition (HESRC)	A	38
Healthy winding under different remanence condition (HEDRC)	B	75
Short-circuit between turns (EFST)	C	30
Mechanical deformation (MEDE)	E	31

3.2 Pre-processing

Pre-processing covers three main tasks: noise removal, feature extraction and normalization. In this section only a brief overview to feature extraction is provided. One of the main objectives of feature extraction is to transform the input data into a set of features, also called indicators. If the extracted indicators are correctly chosen, it is expected that the indicators will extract the relevant information from the input data in order to perform the desired task using such reduced representation instead of the full size input.

There are different kinds of indicators that can be used for condensing information of FRA data. For more details about indicators, see reference [4]. The authors have research the use of different kinds of indicators, but in this paper only the results corresponding to the use of correlation coefficients (CC) are presented.

Before the calculation of CC coefficients, the FRA plots were divided into 5 frequency sub-bands (LF1, LF2, MF, HF1 and HF2), as depicted in Figure 3, according to the algorithms presented in [5]. Then, in each of these frequency sub-bands, CC coefficients were calculated for both the magnitude and phase plots. As a result, a total of 10 indicators were extracted. The indicators were numbered from 1 till 10, where the indicator 1, 2, 3, 4 and 5 corresponds to the CC coefficients calculated for the plot of the magnitude in the frequency sub-bands LF1, LF2, MF, HF1 and HF2, respectively. In a similar way, the indicators 6, 7, 8, 9 and 10 corresponds to the CC coefficients calculated for the plot of the phase in the five frequency sub-bands.

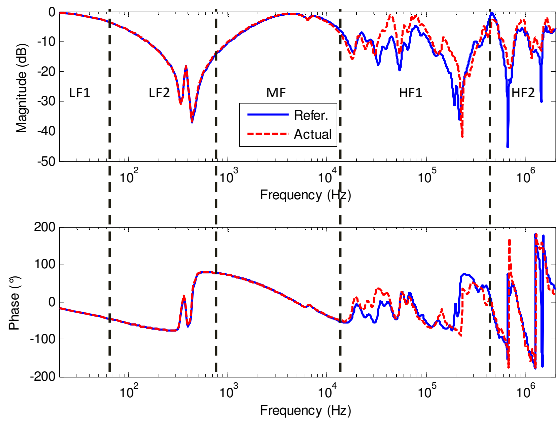

Figure 3: Segregation of FRA plots in sub-bands

For the selection of the indicators to be used as input to the ML algorithm, different statistical methods can be used. As example, the use of principal components analysis (PCA) is here illustrated. From the loading plot of PCA shown in Figure 4, clusters of indicators having a high linear dependency can be identified. So for example, the indicators 2 and 7 provide the same information. Thus, only one of these indicators can be used, reducing in this way the dimension of the input space.

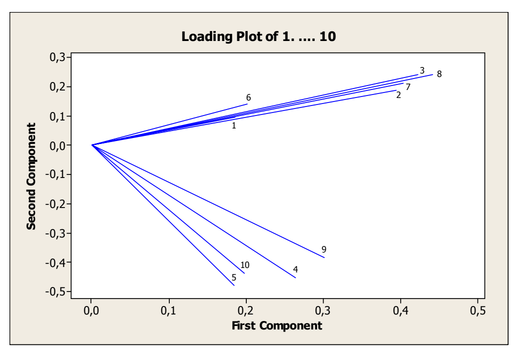

Figure 4: Loading plot of PCA

After reducing the dimension of the input indicators by selecting only one indicator per cluster, the topology of the classifiers shown in Figure 5 was defined. After some trials with different topologies, it was decided to use two independent classifiers: one for the classification of low frequency failure modes and the other one for the classification of high frequency failure modes.

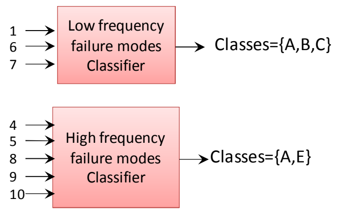

Figure 5: Topology of the classifiers

3.3 Training of the ML algorithm

There are different kinds of ML algorithms that can be used. In this paper only the use of decision tree (DT) classifiers [7] is presented.

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences. The family of decision trees algorithms are usually called Top-Down Induction on Decision Trees (TDIDT). To this family belongs algorithms such as CART (Classification and Regression Trees), CLS (Concept Learning System), ID3 and C4.5, among others.

The components of a DT are shown in Figure 6. As can be noted, a DT has a root node (RN), intermediate nodes, (IN), also called child nodes, splitters and leaves. Any IN can be a RN of a sub-tree. This leads to the recursive definition of a decision tree. A leave corresponds to a set of instances belonging to a single class. The class of the leave is assigned according to the class corresponding to the majority of the instances. The leaves represent the automatically extracted concepts.

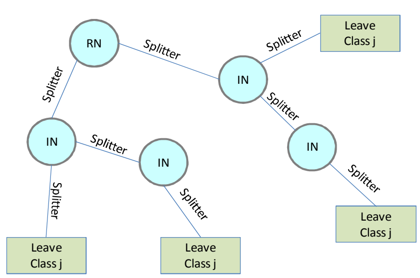

Figure 6: Components of a decision tree

As illustrated in Figure 7, the training set is progressively divided using properly selected splitters. The splitters check the fulfilment of conditions such as: is that value of the attribute X greater/smaller than a threshold? Or is the value of the attribute X in the range between A and B? In Figure 7 the classification process of a training set consisting of elements of four different classes (C1, C2, C3 and C4) is illustrated. The training set to be classified is divided by the splitter S1 into a leave having only elements of the class C1 and a new training set (sub-training set), which is a daughter set of the initial training set. Afterwards, the splitter S2 divides the sub-training set into two leaves (C2 and C3). As can be noted, in contrast to C1, the leaves C2 and C3 are not pure classes because some impurities are present. For example, the leave corresponding to the class C2 has an element of the class C1 and one element of the class C4.

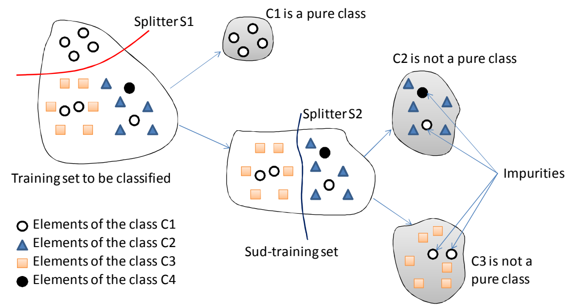

Figure 7: Split of the training set for obtaining pure classes

For the creation of the decision tree classifiers the software Weka was used [6]. The decision tree was created using the WEKA algorithm J48 that is an approximation to the algorithm C4.5, . After running the algorithm, the DT classifiers presented in section 4 were obtained.

4 RESULTS

The DT classifier obtained for the classification of low frequency failure modes (classes A, B and C) is presented in Figure 8. The tree was pruned using a confidence factor of 0.25. The performance of the DT against cross-validation using a partitioning of 10 folds was very good as 84.45% (125 of 148) of the instances were correctly classified. Details on the number instances in the training set (NITS), number of correctly classified instances per class (NCCI) as well as the TP Rate and the confusion matrix are presented in Table 2.

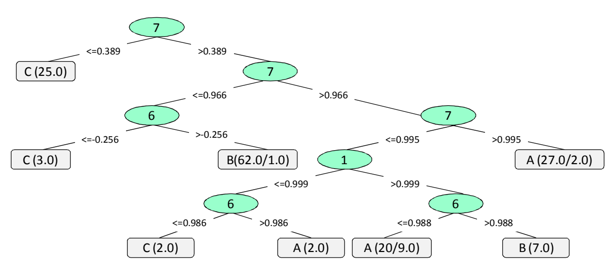

Figure 8: DT for classification of low frequency failure modes

Table 2: Performance of the low frequency DT

Class	NITS	NCCI	TP Rate	Confusion matrix
A	38	31	0.816	31 7 0 | a = A
B	78	67	0.859	10 67 1 | b = B
C	32	27	0.844	3 2 27 | c = C

The DT classifier obtained for the classification of high frequency failure modes (classes A and E) is presented in Figure 9. The tree was also pruned using a confidence factor of 0.25. The performance against cross-validation is also good as 81.63% of the instances were correctly classified (119 of 147). Details on the performance are presented in Table 3.

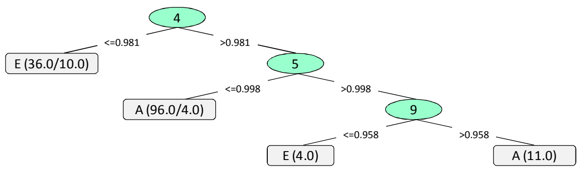

Figure 9: DT for classification of high frequency failure modes

Table 3: Performance of the high frequency DT

Class	NITS	NCCI	TP Rate	Confusion matrix
A	113	101	0.894	a b 101 12 | a = A
E	34	19	0.559	15 19 | b = E

5 CASE STUDIES

In this section the application of the DT classifiers presented in section 4 are illustrated by means of two cases.

5.1 Case 1

This case corresponds to a 40 MVA, 138/13.8 kV, Ynynd11-connected power transformer. In view that an FRA fingerprint was not available, a comparison among phases was done, as illustrated in Figure 10.

The calculated CC coefficients and the outcome of the classification provided by the DT classifiers are shown in Table 4. According to these results, the transformer has short-circuits between turns (class C) with a probability of 1. At the same time, the classification provided by the DT in charge of assessing high frequency failure modes, that is, mechanical deformations, indicates that the windings are healthy. In other words, no mechanical deformations were diagnosed. The diagnosis provided by both DTs was in agreement with the real condition of the transformer, as reported by the owner of the transformer.

The calculated relative factors and the diagnosis provided by the algorithm of the DL/T911-2004 are also presented in Table 4. According to this algorithm, the transformer has a slight deformation.

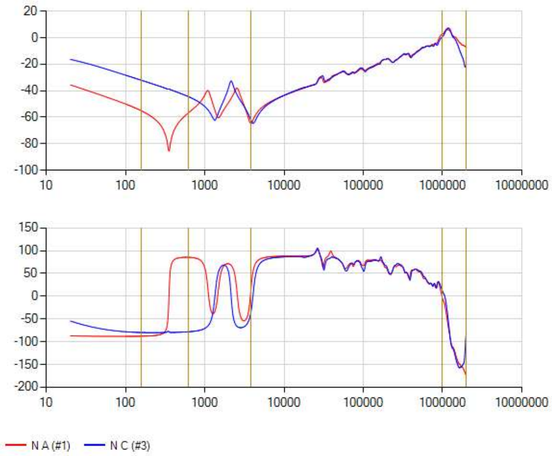

Figure 10: FRA plots: case 1

Table 4: Classification: case 1

CC coefficients 1 = 0.998; 4=0.998; 5=0.883; 6=0.996; 7=0.278; 8=0.841; 9 =0.998; 10=0.985 **Relative factors** R_{LF}=1.69, R_{MF}=2.2, R_{HF}=1.92		
Classification of decision trees		
	Class	Probability
Low Frequency DT	C (25)	1.00
High Frequency DT	A (96/4)	0.96
Assessment according DL/T911-2004 Slight deformation		

5.2 Case 2

This case corresponds to an 800 kVA, 6.3/0.4 kV, Dyn5-connected distribution transformer in which an axial deformation was simulated by inserting spacers between two discs, as illustrated in Figure 11. A comparison of the FRA plots before and after the deformations are shown in Figure 12.

Figure 11: View of the simulated deformation

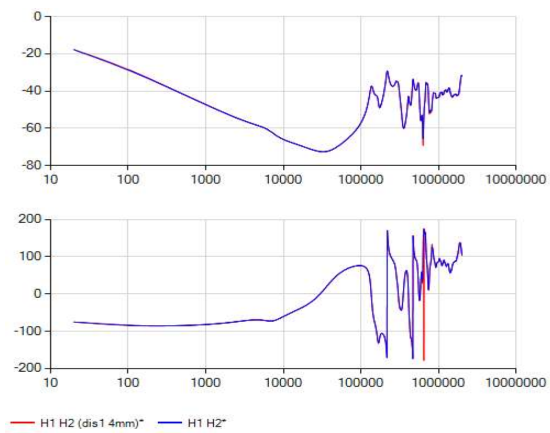

Figure 12: FRA plots: case 1

The outcome of the diagnosis provided by the DT classifiers and by the algotihm of the DL/T911-2004 are presented in Table 5. The low frequency DT classifiers indicates that the transformer is healthy, as expected. At the same time, it can be appreciated that the high frequency DT successfully diagnosed the simulated axial deformation. This illustrates the high sensitiveness of the classification provided by the DT when compared to the algorithm DL/T911 which diagnosed the windings as healthy.

Table 5: Classification: case 2

CC coefficients		
1 = 0.999; 4=0.999; 5=0.999; 6 = 0.999; 7 =0.999; 8=0.999; 9 =0.958; 10=0.999		
Relative factors		
R_{LF}=4.93, R_{MF}=3.6, R_{HF}=2.39		
Classification of decision trees		
	Class	Probability
Low Frequency DT	A (27/2)	0.92
High Frequency DT	E (4)	1
Assessment according DL/T911-2004		
Normal winding		

These cases illustrate the effectiveness of the diagnosis provided by the DT classifiers. The classifiers not only indicates that there is a problem, but also indicate the specific failure that has occurred the in the active part transformer (electrical or mechanical). This kind of algorithms are without doubt an step forward to the enhancement of the reliability of the automatic assessment because of their capabilities of diagnosing both electrical and mechanical failures and due to its enhaced sensitiveness to the diagnosis of mechanical deformations (as demonstrated by the case 2). An additional benefit of these algorithms is that the outcome of the diagnosis is provided with a probability that allows the user to get a good idea about the reliability of the diagnosis.

6 CONCLUSION

It can be concluded that the application of ML techniques is very promising for solving the biggest challenge of the FRA method, namely, reliable automatic interpretation of results. It was illustrated how from a database of FRA measurements together with simulation of deformations in real transformers, patterns can be extracted for assessing in a more reliable way FRA results. The obtained decision trees are quite simple and of easy implementation. The benefits of the classification using the proposed decision trees against the algorithm DL/T911 was illustrated by means of two cases. The DT classifiers allow diagnosing with high sensitivity not only mechanical deformations, but also short-circuit between turns. Further research works include: expansion of the training set (from real FRA data of transformers and from simulation of failures using physical models) and development of an expert system using as knowledge base the knowledge extracted from ML algorithms.

7 REFERENCES

[1] J. Zhijian, L. Jingtao, Z. Zishu, "Diagnosis of Transformer Winding Deformation on the Basis of Artificial Neural Network", Proceedings of The 6th International Conference on Properties and Applications of Dielectric Materials, Xi'an Jiaotong University, Xi'an, China, 2000.

[2] T. Kato, K. Inoue, S. Ishii, "Automatic Diagnosis of Faults and Their Locations in Power Transformer Impulse Test by SOM", XIIIth International Symposium on High Voltage Engineering, Rotterdam, Netherlands, 2003.

[3] A.J. McGrail, et al., "Data Mining Techniques to Assess the Condition of High Voltage Electrical Plant", CIGRÉ Session 2002, Paper 15-107, 2002.

[4] J.L. Velásquez, A. Hedgecock, M.A. Sanz-Bobi, S. Galceran, "Reliable and Automatic Assessment of the Frequency Response Analysis of Power Transformers: Is it feasible?", ARWtr2010, Santiago de Compostela, Spain, 2010.

[5] J.L. Velásquez, et al., "Identification of Transformer-specific Frequency Sub-bands as Basis for a Reliable and Automatic Assessment of FRA results", CMD Conference 2010, Tokio, Japan, 2010.

[6] M. Kramer, "Machine Learning with Weka: An Introduction", http://www3.stat.sinica.edu.tw/stat2005w/download/weka_050930.pdf

[7] L. Rokach, O. Maimon, “Top-down induction of decision trees classifiers-a survey". IEEE Transactions on Systems, Man, and Cybernetics, Part C 35: 476–487 2005.