DOI: 10.48011/sbse.v1i1.2287

A Stairway Statistical Neural Model for DGA Analysis ⋆

Gabriel de Souza Pereira Gomes *
Daniel Carrijo Polonio Araujo ** **Mateus Batista de Morais** **
Rafael Prux Fehlberg * **Murilo Marques Pinto** ***
Arthur Franklim Marques de Campos ****
Rogério Andrade Flauzino *

* *Esc. Eng. de São Carlos - EESC, Univ. de São Paulo - USP, SP (e-mail: gabriel.gomes@radicetech.com)*
** *Treetech Sistemas Digitais, Atibaia, SP (e-mail: daniel.carrijo@treetech.com.br)*
*** *Radice Tecnologia, Atibaia, SP (e-mail: murilo.marques@radicetech.com)*
**** *CEB Distribuição, Brasília, DF (e-mail: arthur.campos@ceb.com.br)*

Abstract: This paper proposes a new approach for power transformers dissolved gas analysis (DGA) using Statistical Machine Learning Techniques and Neural Networks to compose a stairway model which performs analysis in three levels in order to check the existence of faults and which type it most probably is. The proposed approach shortcuts the problem of lacking reliable data related to the type of fault creating a model with three levels of analysis. The first one uses real data from an energy company and from IEC TC 10 data to classify the DGA samples as faulty or normal. After that, a second one based just on IEC TC 10 takes place to classify three possible types of the fault. The third level is used to classify 5 types of fault in a more detailed analysis. The proposed levels of the model achieved an accuracy in the test set of 100 %, 94 % and 92 % respectively.

Keywords: Power Transformer; Condition Based Maintenance; Fault Prediction; Dissolved Gas Analysis; Neural Network; Classification.

1. INTRODUCTION

Power transformers are one of the most important and expensive assets in energy power systems. As presented by Carrijo (2009), Power transformers are responsible for supplying energy for loads that goes from houses to hospitals and big industries. Therefore, to ensure its reliable and safe operation, besides avoiding incipient faults, monitoring the condition of those assets plays a major role.

Analysis of dissolved gas in oil is considered to provide a reliable diagnostic tool for assessing the condition of power transformers. DGA provides information about the gases formed inside the asset and these can be related to the occurrence of some faults.

Regarding the analysis and interpretation of these tests, lots of methods have been developed, such as Doernenburg, Rogers and Duval methods. The first two methods use gas ratios to detect the existence and the type of faults, but they have the disadvantage of not being closed methods, i.e. they don't provide diagnostic to any gas quantity (e.g. low gas concentration faults). Doernenburg gas ratios can be seen in Table 1. On the other hand, Duval's method presented in Duval (2008) and Duval and Lamarre (2014) use gas percentage and is closed, hence always providing diagnostics as can be seen in Figure 1. There are also some methods that are a combination of the previous ones such

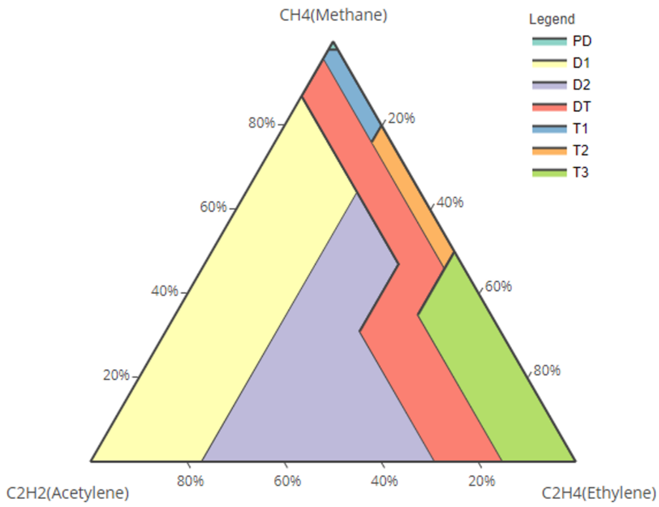

Figure 1. Duval Triangle 1 for DGA Analysis

⋆ This work is supported by PD-5160-1804/2018 ANEEL, developed by CEB and Radice with collaboration of Treetech and SEL-EESC-USP.

as the method presented in Kim et al. (2013) and methods that are based in the thermodynamics of gas formation as Cruz et al. (2015).

Table 1. Ratios for keys gases - Doernenburg, Bakar et al. (2014)

Suggested fault diagnosis	Ratio 1 (R1) CH_4/H_2		Ratio 2 (R2) C_2H_2/C_2H_4		Ratio 3 (R3) C_2H_2/CH_4		Ratio 4 (R4) C_2H_6/C_2H_2	
	Oil	Gas space	Oil	Gas space	Oil	Gas space	Oil	Gas space
1. Thermal decomposition	>1.0	>0.1	<0.75	<1.0	<0.3	<0.1	>0.4	>0.2
2. Partial discharge (low-intensity PD)	<0.1	<0.01	Not significant		<0.3	<0.1	>0.4	>0.2
3. Arcing (high-intensity PD)	>0.1 to <1.0	>0.01 to <0.1	>0.75	>1.0	>0.3	>0.1	<0.4	<0.2

The major drawbacks of these methods are the facts that they need some specialist interpretation for the diagnostics without supporting decision making, they just use some gas ratios and they all have been constructed considering only IEC TC 10 database, which is a database that compiles a lot of power equipment post-mortem analysis, relating the failures with DGA tests performed on those assets. In order to overcome the problem of lack of expert personnel, some solutions use fuzzy logic as Abu-Siada et al. (2013), Khan et al. (2015) and Noori et al. (2017) and combinations of Neural Networks and other standard methods, like the method presented in Chatterjee et al. (2019). However, these propositions do not contemplate other points like data reliability and continuous training.

2. GASES GENERATION IN TRANSFORMERS DUE TO FAULTS

During normal operating conditions, it is common for gases to be generated inside the equipment. Under abnormal conditions, such as electrical and thermal stresses, the formation of these gases can vary depending on the type of stress suffered. Typical gases that appear in insulating oil are hydrogen and hydrocarbons such as methane, ethane, ethylene and acetylene. Each of these gases has a specific formation temperature, as can be seen in the Figure 2.

Hydrogen formation starts at around 150 °C and increases as the temperature increases. Methane also starts to be produced at around 150 °C, but after reaching a maximum value it starts to decrease with increasing temperature. The same happens with ethane, which starts to be produced at 250 °C, and ethylene which starts its production at around 350 °C. After reaching their maximum, they all start to decrease as the temperature increases. Finally, acetylene production starts between 500 and 700 °C and increases as the temperature increases.

Between 200 and 300 °C, the methane production rate exceeds hydrogen production. At around 275 °C, ethane production exceeds methane production. At 450 °C, the production of hydrogen begins to be greater than the production of all other gases until around 750 °C, when the production of acetylene starts to increase. Usually, the initial quantity of each gas and the rate of formation varies with several factors such as equipment design, applied load and type of material used for solid and liquid insulation. Figure 3 shows examples of faults in power transformers. The relative concentration of the gases can be seen in the Figure 4.

In this scenario, this paper proposes a new approach based on a three-level stairway analysis, using all gases available

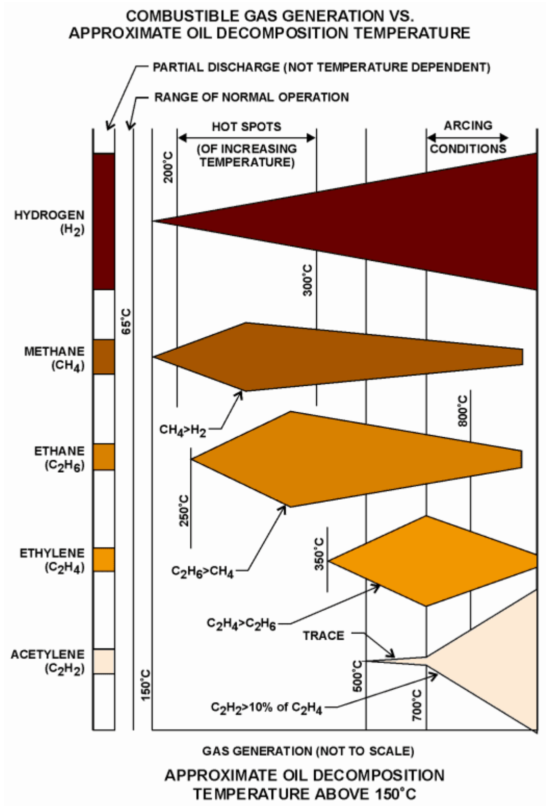

Figure 2. Faults and generated gases, Rogers (1978)

in a DGA test, except for nitrogen and oxygen, aiming to overcome the complexity and strong dependence of experts that other methods present. In the first level of the stairway, it only identifies the existence of a fault. Since IEC TC 10 has few examples, this approach also uses practical data collected from real cases analyzed by transformers specialists. The reliability of this first analysis is high because it is easier for specialists to analyze the existence of a fault than to identify its type. The second level of the stairway analysis performs an identification among three main types of faults and is trained, validated and tested using only IEC TC 10, because there is still a lack of reliable data classifying the fault type. The third level goes deeper, identifying one among 5 types of faults, also based on IEC TC 10.

These serialized models provide a good support for decision-making, allowing it to be performed in a three level, requesting the specialist analysis only when really necessary. Besides, it provides a straightforward way to reinforce model training without losing reliability, since the training process of one stairway level does not interfere on the other.

3. THE STAIRWAY MODEL

When performed by a specialist, a dissolved gas analysis usually follows a diagram like the one presented in Figure 5. First, the specialists classify the transformer as faulty or normal. This part of the process has high reliability, once it is easy to verify if there is a fault or not, e.g. if gas concentrations or rate of gas increase are above typical values for the equipment. After that, the specialist uses some standard method, like Duval's Triangle, to analyze the type of the fault. Then, if the fault diagnostic encountered is dangerous and require immediate action, those will be taken in order to preserve the asset, like shutting down the transformer or scheduling some corrective inspection or maintenance.

Figure 3. Fault examples (from the left to the bottom: D2, D1, T3 and T1), Cigre - JWG D1/A2.47 (2019)

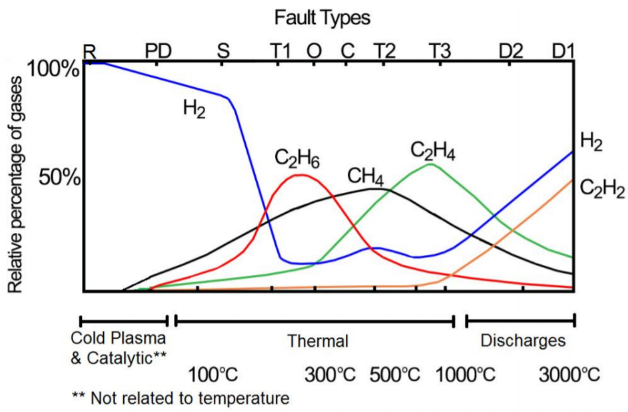

Figure 4. Faults and generated gases, Cigre - JWG D1/A2.47 (2019)

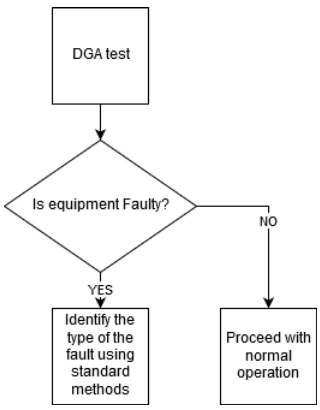

Figure 5. Simple DGA analysis procedure

One way to address this issue is to create a serialized expert analysis in the same way that a human expert would in a real case. The serialized model consists of a three-step analysis, in degrees of data reliability and decision-making support. Figure 6 presents the serialized model.

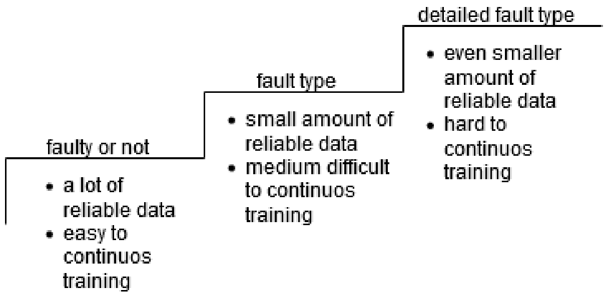

Figure 6. A stairway model for DGA Analysis

In the first step, the model provides information regarding the existence of a fault that can be identified with a DGA test. This first step exists to reduce or eliminate the necessity of an expert analysis to identify if the equipment is faulty or not and can be trained with a lot of high reliable data, once specialists already record such analysis in a proper way. In the sense of decision making, this first analysis give information to the responsible maintenance personnel about which equipment need more attention. In the second step, the model gives information about the type of the fault in three classes: partial discharges (PD), thermal fault (T) and energy discharges (D). In this step, the specialist may be not so precise like in the first one. Even considering that there are other types of tests that can identify the types of faults, like the partial discharge test, they are still expensive and performed only in cases of high necessity. In this sense, it is harder to obtain reliable data to test the models for the second step, but it is still feasible. In the decision making point of view, the second step gives a closer look in the state of the asset, providing to the maintenance personnel information about the severity of the fault. For example, partial discharges damage the equipment, but are not so severe like energy discharges. Hence, step two can provide a general, yet precise classification of fault type, allowing further actions to be planned. In the third level, the model classifies the faults in five types, being: partial discharges (PD), low energy discharges (D1), high energy discharges (D2), thermal faults of low and medium temperature (T1 or T2) and thermal faults of high temperature (T3). The Stairway model decision-tree is shown in Figure 7.

The idea was not to create new classifications of faults, for the existing classification is widely accepted and consolidated. Even considering that the third level of the stairway gives a much better information about the fault and consequently about the state of the equipment, finding reliable practical data to train this third step of the model is even harder. That being said, the third particular level needs more expert analysis, both to obtain relevant data to train the model and to confirm further diagnostics provided on this level. This step can be later improved by considering the equipment's behavior over the time on the training process.

4. STAIRWAY MODEL IMPLEMENTATION

4.1 Dataset challenges

In this work was used the public available IEC TC 10 database. This database contains reliable data of DGA

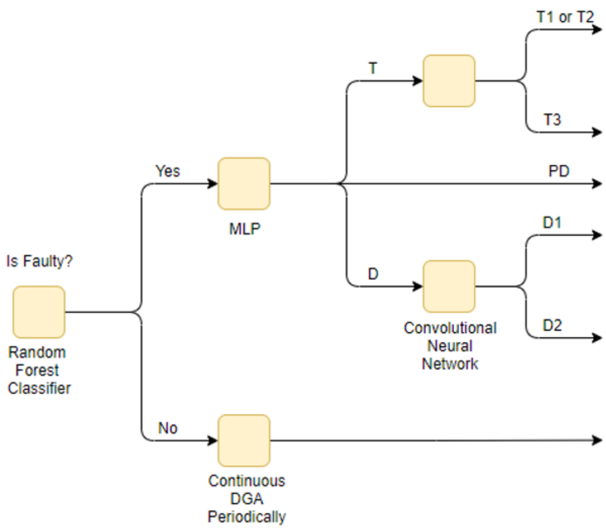

Figure 7. Stairway model decision tree

analysis of faulty equipment inspected in service. The faulty equipment presented by it were removed from service, visually inspected by experts, and the faults were clearly identified for each one. Relevant DGA results were available in all of these cases for correlation purposes. The faults identified by the specialists can be divided into 5 types as show in Duval and DePabla (2001):

(1) Partial discharges (PD) of the cold plasma (corona) type with possible X-wax formation, and of the sparking type inducing small carbonized punctures in paper;
(2) Discharges of low energy (D1), evidenced by larger punctures in paper, tracking, or carbon particles in oil;
(3) Discharges of high energy (D2), with power follow through, evidenced by extensive carbonization, metal fusion;
(4) Thermal faults below 300 °C if paper has turned brownish (T1), above 300 °C if paper has carbonized (T2); and possible tripping of the equipment;
(5) Thermal faults above 700 °C (T3), evidenced by oil carbonization, metal coloration, or fusion.

Some challenges can be pointed in IEC TC 10, the main dataset used in this work, namely:

- Missing values for some gases;
- Unbalanced number of faults per classification:
  - · 34 normal
  - · 9 of type PD
  - · 26 of type D1
  - · 48 of type D2
  - · 16 of type (T1 + T2)
  - · 18 of type (T3)
- Small amount of data (only 151 analysis)

All missing values were filled up with the mean value of the class for the missing gas. The unbalanced data problem was solved by creating synthetic data using the probability distribution of each class. The synthetic data was used only for training the models, they were not used for testing. Unfortunately, the small number of samples will always be a drawback when the subject is neural network, causing the resulting trained network to overfit in most cases. This can only be solved by performing more tests and post-mortem analysis in equipment.

4.2 Implementation

To implement the Stairway model, one specialized model was trained and validated for each level as proposed. The validation was performed using a k-fold cross-validation approach with 5 folds. For the first level, a Random Forest composed by 400 Trees was used in the way presented in Figure 8. This Forest was trained, validated and tested using IEC TC 10 database and real data (50 power transformers) from a Brazilian energy company. The model performance was quite remarkable, reaching an accuracy of 100 % in the test data. The results are detailed and discussed in the next section. For the second level, only

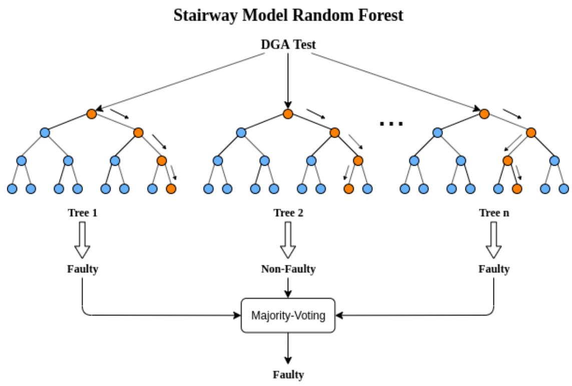

Figure 8. Stairway Random Forest (simplified)

IEC TC 10 was used for training, validating and testing, because there were not enough practical data to do differently. The architecture of the specialized chosen to perform this analysis was a simple Multi-layer Perceptron, with the topology presented in Figure 9 and schema presented in Figure 10. The model reached an accuracy of 94.5 %.

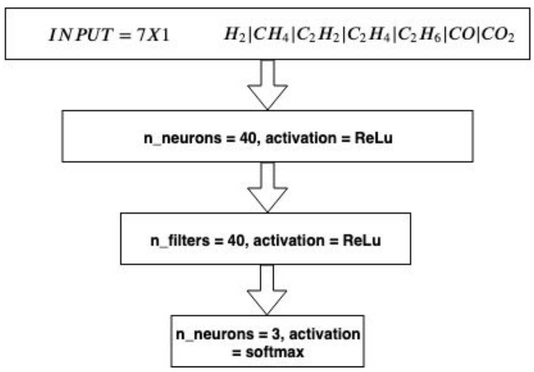

Figure 9. Multi-layer Perceptron Topology

For the third and final level, also IEC TC 10 was used for training, validating and testing. The architecture chosen for this case was a convolutional neural network because a simple MLP did not perform well for this case. The topology of the network can be seen in Figure 12 and the schema in Figure 11. The model achieved an accuracy of 92 %, an expected result when considering the simplicity of the used convolutional neural network.

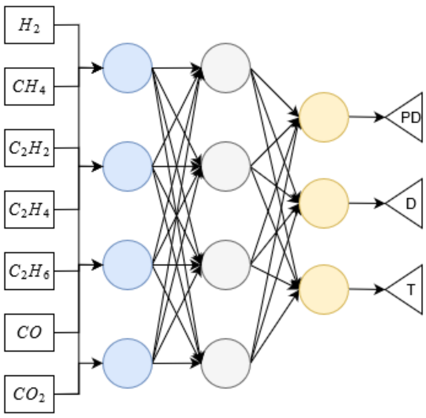

Figure 10. Multi-layer Perceptron Schema

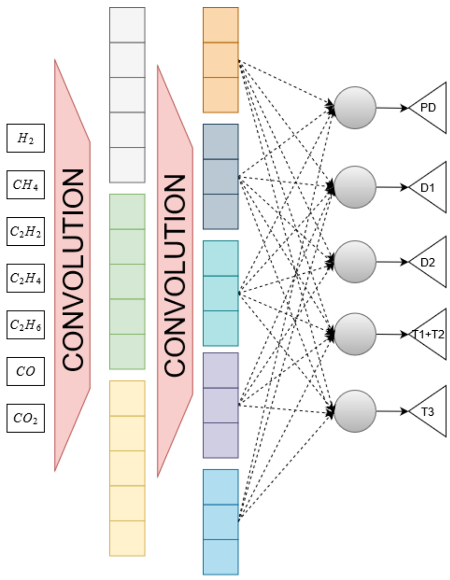

Figure 11. Convolutional Neural Network Schema

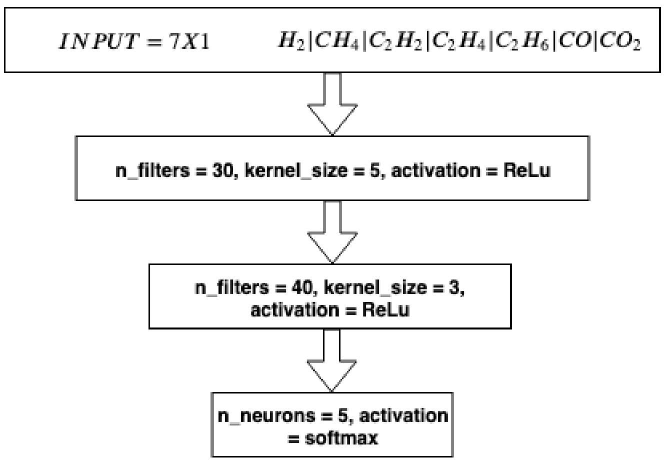

Figure 12. Convolutional Neural Network Topology

5. RESULTS AND DISCUSSIONS

The results for the first level of analysis is presented in Table 2.

Table 2. Results for the first level of analysis

	Train	Validation (kfold)	Test
Accuracy	100 %	96 $\pm$0.09%	100%

As it can be seen, the model performed quite well in identifying the existence of a fault. Nevertheless, due to the small amount of data, the model is a little over-fitted, which can be seen in the validation analysis performed through a cross-validation approach using 5 folds (four for training and one for testing). In order to solve this problem, more data is needed. Even so, the mean accuracy obtained was 96 %, which proves this fact, but, the standard deviation of the cross-validation was small (around 0.09 %) which means that the model is consistent.

An analysis of feature importance was performed based on the average influence of each gas on the impurity decrease of the random forest trees. This was done in order to identify which gases are more relevant when classifying the equipment as faulty. This method is in general biased for continuous or high-cardinality categorical variables. In our case we only have continuous variables, thus the use of this method is quite effective, mainly considering that it runs fast. The result is shown in Figure 13.

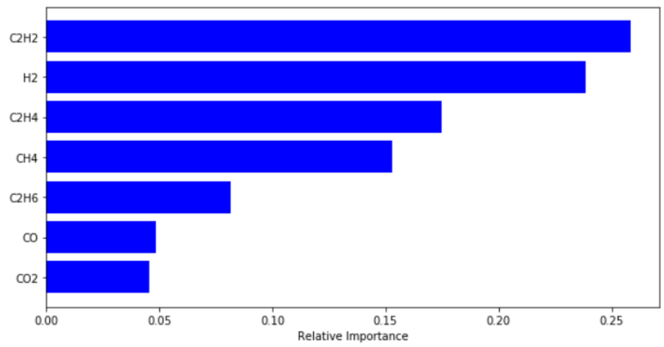

Figure 13. Relative Importance of Each Gas for Faulty or Not

As can be seen, acetylene (C_2H_2), hydrogen (H_2) ethylene (C_2H_4) and methane (CH_4) have together around 80% of the importance, roughly indicating that these three gases alone can be used to identify the existence of a fault in a reliable way. This result is physically consistent with Cigre - JWG D1/A2.47 (2019). Further studies will be conducted on feature extraction and selection to confirm these facts.

For the second level of analysis, the results can be seen in Table 3. The accuracy obtained in the training set was greater than the obtained in both validation and test. This shows that the model is over-fitted. To solve this problem, more data is needed, once hyper-parameter tuning did not solve this problem. Nevertheless, the model is also consistent, as can be seen considering the small standard deviation presented by the cross-validation.

Table 3. Results for the second level of analysis

	Train	Validation (kfold)	Test
Accuracy	100 %	91 $\pm$0.15%	94 %

The third level of analysis performed almost like the second one as can be seen in Table 4. It also suffered over-fitting, as can be seen in the difference between train, validation and test. It is a big challenge to solve the over-fitting problem of this level, because it is hard to find reliable data for training the model. More post-mortem equipment analysis is needed.

Table 4. Results for the third level of analysis

	Train	Validation (kfold)	Test
Accuracy	100 %	90 $\pm$0.15%	92 %

6. CONCLUSIONS

As technology develops, the existence of true specialists is becoming more vital to the life of any asset in the power system, but a lot of them are getting retired and stopping to work. With this in mind, true specialists work hours become scarce and more expensive.

The idea of this paper is not to replace specialist analysis, but to help them to obtain more quickly and reliable results. Data analysis and neural network can be powerful tools in the hands of maintenance personnel and engineers if its limitations are well known, considered throughout the modeling and, when possible, dealt with.

A stairway model for DGA analysis was proposed on this paper in order to help decision-making process, specially reducing the need of equipment specialists for simple cases of DGA tests. In the first level, it provides information about the main state of the equipment, concerning its condition (faulty or not). The second step provides information about the general type of the fault, dividing it in three possible major types: partial discharges, energy discharges and thermal fault. The third step provides a more detailed analysis of the fault type, dividing it in five types: partial discharges, low energy discharges, high energy discharges, low and medium temperature thermal fault and high temperature thermal fault. Moreover, it provides a different way of gathering data, where data for the first level are easier to obtain and more reliable than for the next steps respectively. The models developed to perform the analysis of each step reached a test accuracy of 100 % (first step), 94 % (second step) and 92 % with a very low standard deviation for the validation set, showing that the analysis is consistent. Even so, the limitations of the models were shown and explicitly considered as points where a specialist analysis might still be necessary.

As known limitations, due to the small amount of data, the methodology proposed by this paper serves as a starting point for the development of a robust model where more data will be used to train, test and validate the models. The research next step will be getting more data for training the first and the second step, considering it is easier to find data for them when comparing with the third step.

7. ACKNOWLEDGMENTS

The authors would like to thank **CEB-D**, **Radice** and **SEL-EESC-USP** for all support in PD-5160-1804/2018 ANEEL whose main objective is to create an optimum maintenance policy based on power assets condition and maintenance data. We also thank to **Treetech** for collaborating by providing experts who supported the development team during this work.

8. AUTHOR'S ORCID IDS

Gabriel de Souza Pereira Gomes ; Daniel Carrijo Polonio Araujo ; Mateus Batista de Morais ; Rafael Prux Fehlberg ; Murilo Marques Pinto ; Arthur Franklim Marques de Campos ; Rogério Andrade Flauzino .

REFERENCES

Abu-Siada, A., Hmood, S., and Islam, S. (2013). A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis. *IEEE Transactions on Dielectrics and Electrical Insulation*, 20(6), 2343–2349.

Bakar, N.A., Abu-Siada, A., and Islam, S. (2014). A review of dissolved gas analysis measurement and interpretation techniques. *IEEE Electrical Insulation Magazine*, 30(3), 39–49.

Carrijo, D. (2009). *Estudo de Metodologia e Técnicas para Execução de Ensaios de Resposta em Freqüência em Transformadores de Potência*. Master's thesis, Universidade Federal de Minas Gerais - UFMG, Brasil. doi: 10.13140/RG.2.2.34496.84486.

Chatterjee, K., Dawn, S., Jadoun, V.K., and Jarial, R. (2019). Novel prediction-reliability based graphical dga technique using multi-layer perceptron network & gas ratio combination algorithm. *IET Science, Measurement & Technology*.

Cigre - JWG D1/A2.47 (2019). TB-771 - Advances in DGA Interpretation. *D1/A2 Technical Brochure*, 1.

Cruz, V.G., Costa, A.L., and Paredes, M.L. (2015). Development and evaluation of a new dga diagnostic method based on thermodynamics fundamentals. *IEEE Transactions on Dielectrics and Electrical Insulation*, 22(2), 888–894.

Duval, M. (2008). The duval triangle for load tap changers, non-mineral oils and low temperature faults in transformers. *IEEE Electrical Insulation Magazine*, 24(6), 22–29.

Duval, M. and DePabla, A. (2001). Interpretation of gas-in-oil analysis using new iec publication 60599 and iec tc 10 databases. *IEEE Electrical Insulation Magazine*, 17(2), 31–41.

Duval, M. and Lamarre, L. (2014). The duval pentagon-a new complementary tool for the interpretation of dissolved gas analysis in transformers. *IEEE Electrical Insulation Magazine*, 30(6), 9–12.

Khan, S.A., Equbal, M.D., and Islam, T. (2015). A comprehensive comparative study of dga based transformer fault diagnosis using fuzzy logic and anfis models. *IEEE Transactions on Dielectrics and Electrical Insulation*, 22(1), 590–596.

Kim, S.w., Kim, S.j., Seo, H.d., Jung, J.r., Yang, H.j., and Duval, M. (2013). New methods of dga diagnosis using iec tc 10 and related databases part 1: application of gas-ratio combinations. *IEEE Transactions on Dielectrics and Electrical Insulation*, 20(2), 685–690.

Noori, M., Effatnejad, R., and Hajihosseini, P. (2017). Using dissolved gas analysis results to detect and isolate the internal faults of power transformers by applying a fuzzy logic method. *IET Generation, Transmission & Distribution*, 11(10), 2721–2729.

Rogers, R. (1978). Ieee and iec codes to interpret incipient faults in transformers, using gas in oil analysis. *IEEE Transactions on Electrical Insulation*, EI-13, 349–354.