10123 Session 2022
A1 - Rotating Electrical Machines
PS 2 / Asset Management of Electrical Machines

Data Science and AI for On-line Diagnosis of rotating Machines from Pre-existing Sensors, with applications in Hydro Generators and Wind Generators

Marcos E. G. ALVES*[1] - Brazil – marcos.alves@radicetech.com
Gabriel S. P. GOMES[4] – Brazil – gabriel.spgomes@usp.br
Murilo M. PINTO[1] – Brazil – murilo.marques@radicetech.com
Guilherme TOYOSHIMA[2] - Brazil - guilherme.toyoshima@ibituenergia.com
Rafael P. FEHLBERG[1] – Brazil – rafael.fehlberg@radicetech.com
Catia P. URAS[1] – Brazil - catia.pedrosa@radicetech.com
Daniel C. P. ARAUJO[1] – Brazil – daniel.carrijo@radicetech.com
Bruno F. SARDINHA[1] – Brazil – bruno.sardinha@radicetech.com
Gilberto A. MOURA[1] – Brazil - gilberto.amorim@radicetech.com
Iony P. SIQUEIRA[3] – Brazil - iony@tecnix.com
Camila BARBOSA[4] – Brazil – camilagomss@gmail.com
Rogério A. FLAUZINO[4] – Brazil - raflauzino@usp.br

RADICE TECHNOLOGY[1], IBITU ENERGIA[2], TECNIX[3], USP[4]

SUMMARY

Accidents with rotating machines, such as wind turbines and hydro generators, caused by factors such as weathering, internal defects, or others, can lead to great losses not only of a financial nature, but also of a human and environmental nature. Because of this, the concern about the condition of the machines and their operating context has risen and have become increasingly relevant in recent decades, intensifying research in several areas related to the problem, such as their structures and compounds. However, as much as the compounds, the design and the manufacturing process of rotating machines can improve due to these researches, all components still suffer from several failure modes, in addition to possible random failures, related to defects caused by the aging of the machine or due to operating conditions not foreseen in the project.

Thus, ensuring the proper functioning of this equipment is an extremely essential service for the electricity sector. Therefore, it is increasingly necessary the development of new techniques and tools to support maintenance and asset management teams in identifying defects in rotating machines that are still in an incipient stage, enabling these teams to better support decision making, in order to meet the utility's organizational objectives more efficiently.

In the specific case of wind generators, a very specific and commonly found business model is also added to this context, in which the manufacturer of the wind turbines fully takes care of the park and machine maintenance during their warranty period. This model is largely motivated by the fact that large-scale adoption of this generation technology is relatively recent, limiting the availability of

professionals and specialized knowledge. As the parks' original warranty periods are coming to an end, utilities often have a need to take on not only the maintenance of equipment, but also the full consequences of failures.

In order to address these challenges and needs, this work is developing a methodology to monitor and diagnose online the main failure modes of wind turbines and hydro generators through the use of maintenance history and data from the pre-existing sensors in the machines, often already connected and having data collected in the plant's SCADA system, associated with advanced statistical techniques and artificial intelligence. This methodology is being coded in open Python language on an asset management platform, an environment that allows the native integration of intelligent algorithms with maintenance order data, malfunction registration, registration, and sensor data in real time.

The methodology aims to support the maintenance engineering and asset management of the generation utility with the objectives of increasing safety and reliability, optimizing maintenance resources and reducing the loss of revenue due to unavailability of the machines.

This work will present the operation method of the developed methodology, going through the steps listed below:

1. Acquisition of sensor data - such as temperatures, load currents, power consumption of auxiliary motors, speed, vibration, voltage, angle, and power of different parts of the machine
2. Pre-processing, sanitization, and extraction of characteristics from the sensor data,
3. Inference process by machine learning models (Machine Learning) and advanced statistical algorithms and, finally,
4. Analysis of the result and generation of the diagnosis.

The basis on which the methodology was developed is an exploratory analysis carried out on the data to find cause and effect correlations between the main failure modes of the machines and the data coming from the sensors, subsidizing the extraction of characteristics and the efficient creation of models.

To evaluate the effectiveness of the developed system, a pilot application of the methodology and intelligent software is being carried out on 210 wind turbines from two different manufacturers and 3 hydro generators in Ibitu Energia.

The work is being developed within the scope of the ANEEL R&D Project 0119, titled Intelligent System for Optimized Management of Wind Turbines and Hydro Generators.

KEYWORDS

Machine learning, wind turbines, fault detection, time series, pitch system failures

1. INTRODUCTION

Accidents with rotating machines, such as wind turbines and hydro generators, caused by factors such as weathering, internal defects, or others, can lead to great losses not only of a financial nature, but also of a human and environmental nature. Because of this, the concern about the condition of the machines and their operating context has risen and have become increasingly relevant in recent decades, intensifying research in several areas related to the problem, such as their structures and compounds. However, as much as the compounds, the design and the manufacturing process of rotating machines can improve due to these researches, all components still suffer from several failure modes, in addition to possible random failures, related to defects caused by the aging of the machine or due to operating conditions not foreseen in the project.

Thus, ensuring the proper functioning of this equipment is an extremely essential service for the electricity sector. Therefore, it is increasingly necessary the development of new techniques and tools to support maintenance and asset management teams in identifying defects in rotating machines that are still in an incipient stage, enabling these teams to better support decision making, in order to meet the utility’s organizational objectives more efficiently.

In the specific case of wind generators, a very specific and commonly found business model is also added to this context, in which the manufacturer of the wind turbines fully takes care of the park and machine maintenance during their warranty period. This model is largely motivated by the fact that large-scale adoption of this generation technology is relatively recent, limiting the availability of professionals and specialized knowledge. As the parks' original warranty periods are coming to an end, utilities often have a need to take on not only the maintenance of equipment, but also the full consequences of failures.

In order to address these challenges and needs, this work is developing a methodology to monitor and diagnose online the main failure modes of wind turbines and hydro generators through the use of maintenance history and data from the pre-existing sensors in the machines, often already connected and having data collected in the plant's SCADA system, associated with advanced statistical techniques and artificial intelligence. This methodology is being coded in open Python language on an asset management platform, an environment that allows the native integration of intelligent algorithms with maintenance order data, malfunction registration, registration, and sensor data in real time.

The methodology aims to support the maintenance engineering and asset management of the generation utility with the objectives of increasing safety and reliability, optimizing maintenance resources and reducing the loss of revenue due to unavailability of the machines.

This work will present the operation method of the developed methodology, going through the steps listed below:

1. Acquisition of sensor data - such as temperatures, load currents, power consumption of auxiliary motors, speed, vibration, voltage, angle, and power of different parts of the machine
2. Pre-processing, sanitization, and extraction of characteristics from the sensor data,
3. Inference process by machine learning models (Machine Learning) and advanced statistical algorithms and, finally,
4. Analysis of the result and generation of the diagnosis.

This article is divided as follows: in the first chapter, an introduction is presented. In the second chapter, the methodology applied to wind turbines is described, and the results of this methodology are presented in real cases. In the third chapter, the methodology applied to hydrogenerators is presented, and a case study is presented using data from the literature. In the last chapter, the conclusions of this work are presented.

2. WIND TURBINE METHODOLOGY

2.1 DATA DESCRIPTION

A three-year period of data from sensors extracted from the wind turbine SCADA system was used to carry out this work. The data was organized by entity model, where each entity corresponded to a wind turbine, and the data of this entity corresponded to the values acquired by the machine sensors. In total, 23 wind turbines were used, each of which had data from 609 attributes related to the sensors. Each sensor transmits machine information to the SCADA system, which records the database every 10 minutes. For each magnitude measured by the sensors, 4 different calculations are sent: maximum, minimum, mean, and standard deviation of the last 10 minutes, in order to represent a probability distribution curve for the corresponding magnitude.

Along with data from the sensors, machine unavailability data over the years 2016 to 2018 was used. This data contained the machine that suffered the outage, the time when it occurred, the duration of the outage, and the affected system. To simplify the model, the pipeline presented in this article handles unavailability caused only by the pitch control system, the one that suffered from more outages.

In this work, only the stops that occurred due to a defect in the pitch with outages of more than one day were used. This filter was applied because downtimes of less than one day can be resolved by resetting the machine, not requiring the intervention of the maintenance team. In the data, there were 1,326 outages caused by a defect in the pitch system.

2.2 FAULT DETECTION PIPELINE

The fault detection pipeline is described in Figure 1 and consists of the data processing stage, followed by feature engineering, model selection, hyperparameter optimization, and evaluation.

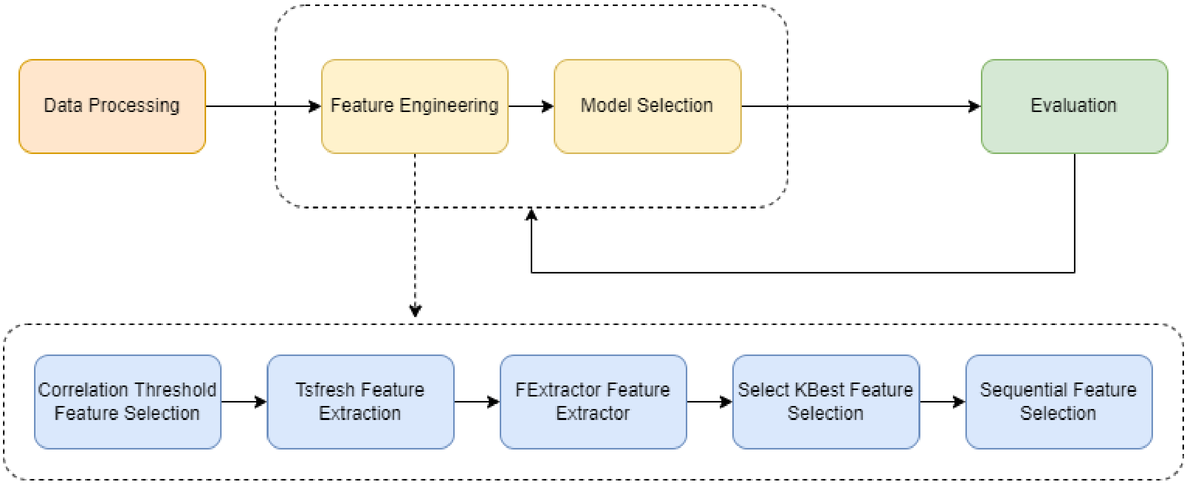
Data Processing
Feature Engineering
Model Selection
Evaluation
Correlation Threshold Feature Selection
Tsfresh Feature Extraction
FExtractor Feature Extractor
Select KBest Feature Selection
Sequential Feature Selection

Figure 1 – Fault detection pipeline diagram

Initially, it is performed a feature selection based on correlation. This process aims to eliminate features highly correlated, keeping only uncorrelated features. This intends to avoid overfitting and training problems for some models that do not behave well with highly collinear features. After that, it is performed the extraction of temporal features such as lag and trend using the tsfresh library [1]. In this work, 3 windows were used for the lag and trend. To move a team to carry out an intervention on the wind turbine requires at least two days, thus, the values of the selected windows were one week, three weeks, and six weeks. These windows were chosen because they corresponded to three different time intervals, the first window being considered a small-time interval, the second a median, and the third a large one.

After the feature extraction, a selection of the best features is performed to predict the remaining time. In this work, a value of k equal to 150 was used. Finally, with the remaining 150 features, a Sequential Feature Selection (SFS) process is performed. Those features that do not significantly change the result will be removed. In this work, after the SFS process, only 81 features remained.

Though other models were tested, this work presents the results obtained with an optimized Ridge Regression [2]. The ratio between train and test data was 90/10%. The evaluation consisted in analysing the learn curve analysis (LCA), R-squared, mean absolute error and the predictions plots.

2.3 RESULTS, DISCUSSION & CONCLUSION

The pipeline presented in this work focused on time series feature extraction to build a strong model that determines the number of days left before a failure in the pitch control system of a wind turbine occurs. In other words, a regression problem.

The tests made for the purpose of this work used data from the wind turbine ICD13. It followed the pipeline previously described.

For the training metrics, R-squared was 0.98, which means that only 2% of the data cannot be explained by the model. In Figure 2, it is shown two periods of time from ICD13, with the real data (blue) and the predicted time to fault (red). The behaviour seen in the graphs reassures us that the metrics presented are real and that the predicted values follow the pattern of the real ones.

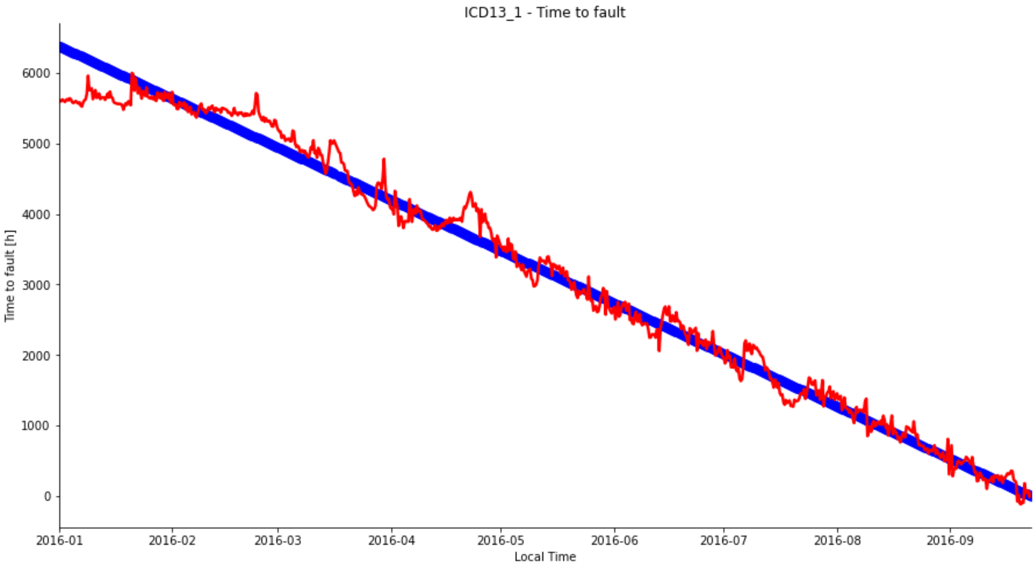

Figure 2 – Prediction Graph of 2016 from ICD13 (Training Data)

The learning curve, presented in Figure 3, helps to understand how well the model is adjusted to the data. It is shown that the performance of the model in both the train and cross-validation data have a mean absolute error close to zero. The minimum difference between the metrics indicate that the features selected to train the model represent the data.

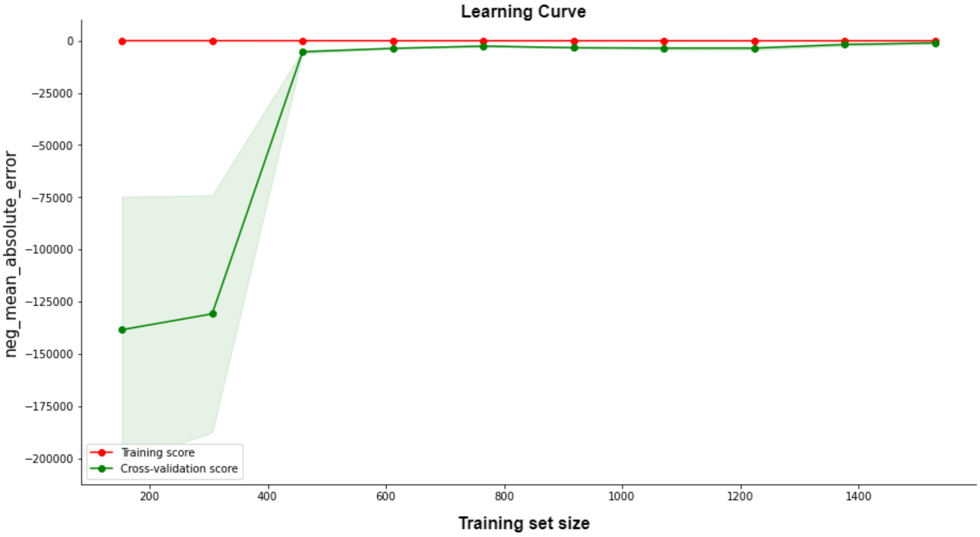

Figure 3 – Learning Curve of the model

As for the testing metrics, the R-squared value above 1 represents an abnormal case, has no logical meaning, and may result from the small sample size. On the other hand, Figure 4 shows that the predicted values are able to follow the real ones. The prediction only differs slightly from the time to failure close to the outage.

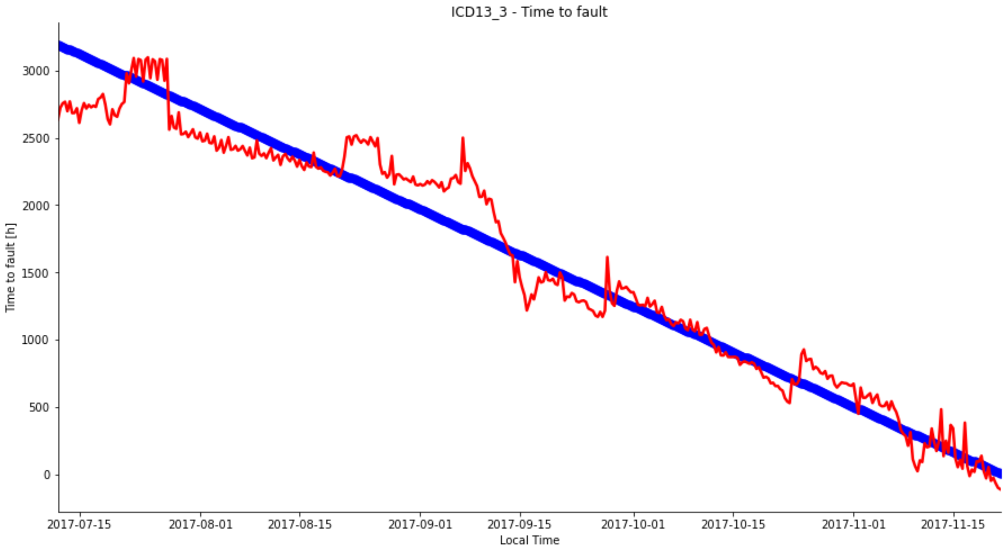

Figure 4 - Prediction Graph of 2016 from ICD13 (Testing Data)

This being said, it is possible to conclude that the pipeline proposed in this paper is capable of predicting the time to failure of a wind turbine. For future work, it is interesting to experiment with models with more generalization capability, making it possible to have only a couple of models for an entire wind plant. Some points should be noted about the approach used in this work: when the model is getting closer to the failure point, its error increases. That happens due to same reasons. The first one is the window used for training: because the window used for training are equal or bigger than one week, the model doesn't prioritize the knowledge of data closer to the failure date. To address this problem, smaller windows should be used, or another model must be created to predict end point failures. The second one is that features that matter for long term prediction can be different for features considering short term prediction. Also, the data resolution matters in this case. New studies are being conducted in order to clarify these points and will be presented in a future article.

3. HYDROGENERATORS

In contrast to wind turbines, fault/defect annotation is uncommon in hydrogenerators. It signifies that only a limited amount of data exists to link sensor measurements to fault/defect events. This fact turns the problem into a hard-to-solve machine learning problem due to the small label quantity. In this case, non-supervised machine learning can be used to detect uncommon patterns in signals. But the link between the sensors and the faults are still missing. To address this problem, this initial study was started with a specialized Bayesian network which will be in the future connect to Anomaly detection models, in order to detect changes in signals and predict what these changes probably mean in terms of machine components. In this way, specialist systems and Anomaly Detection Systems will create a set up for the detection of incipient faults in hydrogenerators, overcoming Anomaly Detection and Specialists Systems problems that make their application or condition monitoring hard. The probabilities were set up by specialists in the area and evaluated against fault data found in the literature.

3.1 PROBABILISTIC SPECIALIST SYSTEMS

Stevens, in [3], defines expert systems as machines that think and reason as an expert would in a given domain. For example, an expert medical diagnostic system would request as input the patient's symptoms, test results, and other relevant facts and, using them as indicators, it would search its database for information that could lead to identification of the disease. Furthermore, an Expert System not only performs traditional computer functions to handle large amounts of data, but also manipulates that data in such a way that the output is a meaningful answer to an unspecified question.

An expert system is basically composed of two parts: a knowledge base and an inference mechanism. The knowledge base is where all the specific knowledge about the domain of a given problem is stored. For example, in a medical application, the knowledge base would contain all the information about the relationship of symptoms and diseases. The inference mechanism is the set of algorithms to process the knowledge stored in the knowledge base, along with any other specific information available about a given application.

3.2 KNOWLEDGE BASE

There are many knowledge base types that go beyond the scope of this work. In this work, we will use the Relevant Symptoms Model Independent, which is a simplification of the Independent Relevant Symptoms Model [4]. In this simplification, it is assumed that the failures are independent of each other and the symptoms are independent of each other but dependent on the diseases. Furthermore, sets of symptoms relevant to each failure mode are defined; that is, not all symptoms are relevant to all types of failure. Figure 5 depicts this model, in which only some sets of symptoms are connected in each type of failure, for example, only symptoms S1 and S2 are relevant for failure f1.

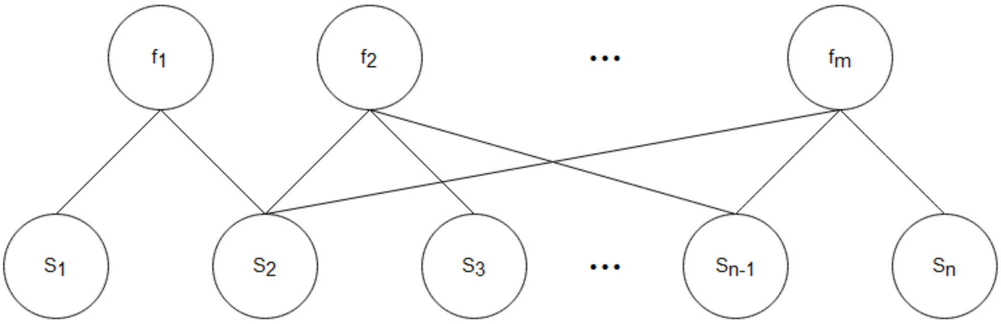

Figure 5 - Independent Relevant Symptoms Model

Considering that S_1, ... , S_{ri} are relevant symptoms for the fault f_i and that the remaining symptoms S_{ri} + 1, ... , Sn are irrelevant, the conditional probability distribution is calculated by Equation 1.

Equation 1 – Desease Probability

$$p(f_i|\mathrm{s_i}, \dots, s_n) = \frac{p(f_i)\prod_{j=1}^{r_i} p(s_j|f_i)\ \prod_{j=r_i+1}^{n} p_j}{p(s_1, \dots, s_n)}$$

From Equation 1, we have that for the calculation of the conditional probability p (fi | S1, ..., Sn) is used only as marginal probabilities p (fi) for all possible values of F, as conditional probabilities p (Sj | fi) for each combination of F-value and its relevant symptoms and the probability pj for each f-value that has at least one irrelevant symptom. Thus, in a scenario with m possible failures in binary symptoms, the required number of parameters is m - 1 + n - ai = 1mri, where ri is the number of parameters relevant to a failure and the number of factors relevant to all failures.

3.3 MODEL

In order to model a Probabilistic Specialist Hydrogenerator Failure Diagnosis System, a critical task is to analyze the types of failures, their characteristics and associated hydrogenerator systems.

Thus, an extensive survey of failure data for hydrogenerators present in the literature was carried out in this work, where the main failure modes of the equipment, the main failure characteristics, their associated alarms and the main systems of the hydrogenerator affected by the defect were identified. The data collected were validated by a group of specialists in hydrogenerator maintenance with years of experience in the area.

The data collected defined the relationship between 10 (ten) symptoms of the hydrogenerator (which are: Undervoltage, Overcurrent, Heating, Vibration, Noise, Harmonics, Short-circuit, Overspeed, Underspeed and Instability) with 6 (six) different failure modes (Stator Insulation Failure, Fatigue, Speed Regulator Failure, Cavitation, Bearing Corrosion and Cooling System Sub-cooling). The evaluated Failure Modes cover some of the main machine systems: Generator, Turbine, Bearings and Cooling System.

Thus, the conditional probabilities p(f_i|S_1, S_2, ..., S_{10}) were then calculated and the Independent Relevant Symptoms network structured. A part of the network, for two Failure Modes and four Symptoms, is shown in Figure 6.

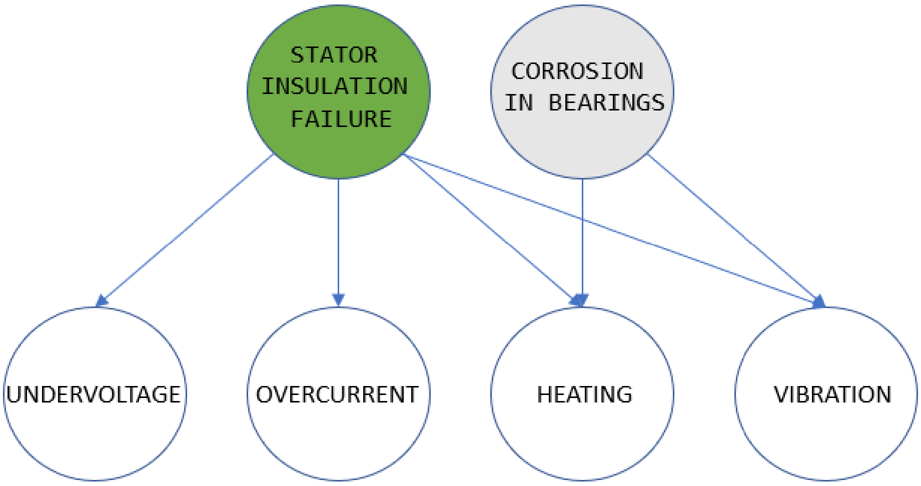

Figure 6 – Case study

3.4 CASE STUDY

For the evaluation and validation of the model, a case study of the application of the probabilistic expert system for the diagnosis of hydrogenerators developed was carried out. Four failure cases found in the literature were evaluated by the model, as follows:

In [5] a case of failure in the Cooling System of a hydrogenerator is presented. In (14), a failure of the type Corrosion in the machine bearings is presented. In both cases, the failures were preceded by an increase above the normal temperature of the hydrogenerator, that is, by a symptom of heating in the equipment. Presenting this symptom to the Expert System, we have a 58% probability of the machine presenting the failure mode Sub-cooling in the Cooling System and 39% of presenting Corrosion in the Bearings. The other failure modes evaluated had less than 1% probability. With that, we can conclude that for this case the model presented a satisfactory result, indicating, in the presence of the Warming symptom, much higher probabilities for the failures reported in the literature than for the other evaluated failures.

In [6] a case in which the hydrogenerator presented Overspeed is presented. This overspeed indicated a failure in the hydrogenerator speed regulator. As in the previous cases, simulating the Overspeed symptom in the developed model, we have an indication of the system of 89% probability of the hydrogenerator presenting Speed Regulator Failure. The second most likely failure was Cavitation, at 6%. The result obtained for this test was also satisfactory, indicating a much higher probability for the failure presented in the reference in relation to the other failure modes.

Finally, in [7] we have a case of failure in the stator insulation that had as symptoms the presence of harmonics and a short circuit between the turns of the hydrogenerator. For these symptoms, the expert system had a 55% probability of the machine having the failure mode reported in the reference. For this case, two other failures stand out, being Speed Regulator Failure, with 21%, and Fatigue, with 17%. However, even though the probability of these failures is reasonable, the probability of the machine presenting a Stator Insulation Failure was much higher. We can conclude then, that for this case, the performance of the model was also satisfactory.

This is shown in the table below:

Table 1 – Study case results

Symptoms	Real defect	Predicted defect
Overheating	Cooling system failure	Undercooling (58%) and Corrosion in bearings (39%)
Overheating	Corrosion in Bearings	Undercooling (58%) and Corrosion in bearings (39%)
Overspeed	Speed regulator failure	Speed regulator failure (89%) / Cavitation (6%)
Harmonics and short-circuit	Stator Insulation Failure	Stator Insulation Failure (55%), Speed Regulator Failure (21%) and Fatigue (17%)

4. CONCLUSION

This work presented two methodologies for condition diagnosis in wind turbines and hydrogenerators. The two methodologies differ in their approach, the first using machine learning and the second adopting an expert probabilistic system. For wind turbines, the results showed that the methodology was able to learn to identify the turbine failure period. The test results proved to be promising, demonstrating that the model understood the wind turbine's failure mechanism. However, when the moment of occurrence of the defect approached, the error of the predicted model increased. This is an indication that a short-term model could help to solve this problem. In hydrogenerators, the result was also promising, correctly identifying the defects of the 4 analyzed cases extracted from the literature. Both models developed for aero and hydrogenerators can be extended for any type of defect or fault. Furthermore, this work demonstrated the possibility of creating failure predictive systems in the presence of labels and in their absence, obtaining satisfactory results in both cases.

The work is being developed within the scope of the ANEEL R&D Project 0119, titled Intelligent System for Optimized Management of Wind Turbines and Hydro Generators.

BIBLIOGRAPHY

[1] Christ, Maximilian, et al. "Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package)." Neurocomputing 307 (2018): 72-77.

[2] McDonald, Gary C. “Ridge regression.” Wiley Interdisciplinary Reviews: Computational Statistics 1.1 (2009): 93-100.

[3] STEVENS, Lawrence. Artificial Intelligence, the Search for the Perfect Machine. Prentice Hall, 1985.

[4] CASTILLO, Enrique. Expert systems; uncertainty and learning. 1991.

[5] MILIĆ, Saša D.; ŽIGIĆ, Aleksandar D.; PONJAVIĆ, Milan M. Online temperature monitoring, fault detection, and a novel heat run test of a water-cooled rotor of a hydrogenerator. IEEE Transactions on Energy Conversion, v. 28, n. 3, p. 698-706, 2013.

[6] YUCESAN, Melih; KAHRAMAN, Gökhan. Risk evaluation and prevention in hydropower plant operations: A model based on Pythagorean fuzzy AHP. Energy policy, v. 126, p. 343-351, 2019.

[7] SHUTING, Wan et al. The analysis of generator excitation current harmonics on rotor winding inter-turn short circuit Fault. Automation of Electric Power Systems, v. 27, n. 22, p. 64-67, 2003.