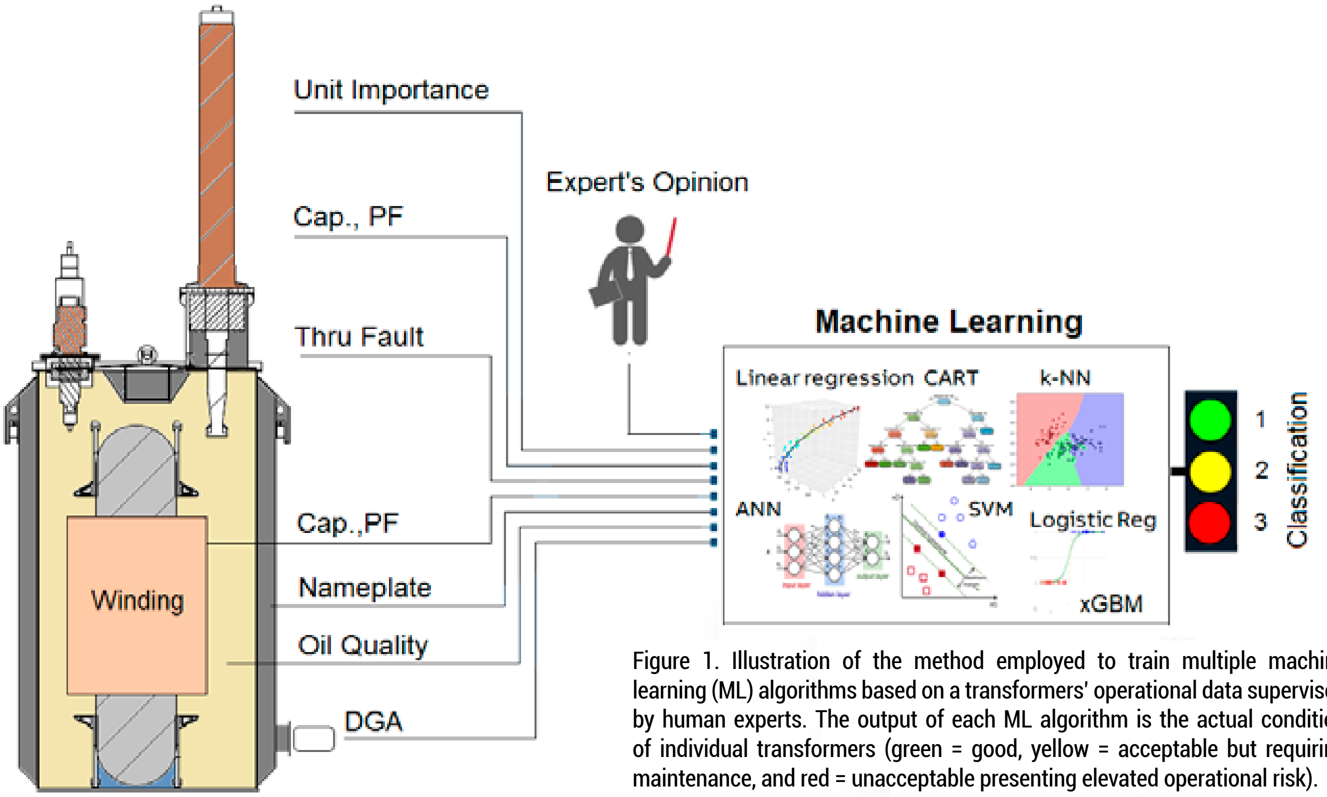

Figure 1. Illustration of the method employed to train multiple machine learning (ML) algorithms based on a transformers' operational data supervised by human experts. The output of each ML algorithm is the actual condition of individual transformers (green = good, yellow = acceptable but requiring maintenance, and red = unacceptable presenting elevated operational risk).

AI - machine learning algorithms applied to transformer diagnostics

ABSTRACT

With the arrival of the age of big data, e-commerce, and smartphones, there has been a growing interest in the application of fast and sophisticated tools, namely machine learning algorithms, to handle massive amounts of data and extract meaningful information that can boost and speed up regression and classification problems, as for example in short term load forecasting and asset condition assessment. This paper describes the use of machine learning (ML) algorithms as supporting tools for the automatic classification of power transformers operating conditions. The work [1] consists of training of multiple ML algorithms with real-life data from 1,000 (one thousand) transformers that were individually analyzed by human experts. Each transformer in the database was scored with a 'green,' 'yellow' or 'red' card depending on the data and the interpretation of human experts, thus serving as the target variable in ML supervised training mode. The paper describes the main steps towards the training of the multiple ML algorithms and the stunning output produced by those algorithms when requested to analyze 200 unseen transformer cases (new cases).

KEYWORDS

automated tool, condition assessment, machine learning algorithms, transformer diagnostics

Machine learning algorithms can be interpreted as a universal non-linear approximator which can be used for fitting very complex multidimensional data with an arbitrary number of inputs and outputs

Table 1. Structure of training dataset, showing 10 random samples

Age	IMP	HV	MVA	TF	IFT	DS	PF25	H_2O	H_2	...	CO_2	O_2	N_2	H1PF	H1Cap	Bsh-PF	Bsh-Cap	CO-$_2$CO	O_2N_2	Class
43	70	345,0	201,6	1,00	33,1	44,1	0,005	16,0	10		913	12479	65221	0,51	35,9	0,51	359	14,27	0,19	2
20	57	141,0	93,0	14,00	33,0	35,0	0,030	3,9	2		210	1900	110000	0,39	2562,8	0,52	190	42,00	0,02	3
44	70	345,0	33,3	0,10	33,4	34,4	0,036	20,0	8		333	6910	32940	0,40	36,5	0,40	365	8,12	0,21	2
44	100	765,0	100,0	0,10	33,9	42,0	0,078	6,6	50		3484	377	26202	0,25	44,0	0,25	440	11,28	0,01	1
34	60	20,9	39,2	0,20	31,0	35,0	0,020	33,0	8		2012	260	21440	0,42	1051,2	0,35	151	15,24	0,02	3
25	85	345,0	660,8	0,66	26,0	35,0	0,051	9,0	3		8818	5715	70864	0,19	8000,0	0,41	1838	58,79	0,08	3
22	30	230,0	53,3	0,66	42,0	39,0	0,013	12,0	11		540	2135	79702	0,36	1542,8	0,32	179	22,50	0,03	3
23	100	765,0	500,0	2,00	34,0	25,0	0,042	7,2	48		2710	28215	79492	0,24	38,9	0,24	389	3,07	0,35	1
51	47	161,0	230,0	0,20	33,0	42,0	0,180	13,0	25		5472	1103	68585	0,61	3396,3	0,60	308	195,43	0,02	2
10	100	765,0	112,4	0,10	25,0	35,0	0,005	2,7	24		5608	8661	24715	0,37	58,6	0,37	586	4,38	0,35	3

Table 2. Statistical description of all features of the transformer dataset

	Age	IMP	HV	MVA	TF	IFT	DS	PF25	H_2O	H_2	...	CO_2	O_2	N_2	H1PF	H1Cap	Bsh-PF	Bsh-Cap	CO-$_2$CO	O_2N_2	Class
Count	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000		1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Mean	28,8	54,0	290,3	191,4	7,6	34,7	36,6	0,121	8,9	63,4		2156,6	6221,2	57002,1	0,40	1149,2	0,38	394,8	38,20	0,11	
Stdev	16,2	30,8	250,3	202,9	14,6	5,5	8,0	1,534	9,2	568,0		2594,6	8256,0	33658,7	0,26	1308,2	0,21	615,7	69,47	0,11	
Min	1,0	0,6	4,2	0,2	0,0	0,0	7,0	0,001	1,0	1,0		15,0	62,0	53,0	0,04	0,0	0,12	0,0	0,07	0,00	1
0,25	16,8	29,0	138,0	47,0	0,2	33,0	35,0	0,005	3,9	8,0		517,8	960,0	28962,3	0,26	40,2	0,27	156,8	7,46	0,02	1
0,5	30,0	50,0	161,0	100,0	0,7	33,0	35,0	0,023	5,0	11,2		1290,5	3113,5	56716,0	0,35	964,8	0,36	217,2	13,35	0,06	3
0,75	39,0	80,0	345,0	260,8	5,0	37,8	37,0	0,060	11,0	25,0		2643,0	8577,8	77402,8	0,45	1764,0	0,42	408,5	34,86	0,18	3
Max	79,0	100,0	765,0	1000,0	79,0	56,8	75,0	35,300	117,0	15092,0		22200,0	74556,0	300210,0	4,63	9195,6	2,90	7062,2	700,00	0,50	3

Fitting of input data to desired output data using ML algorithms is called learning or training, and once the ML model is trained, it can be used for predicting output value for arbitrary inputs

Machine learning algorithms and techniques are used for the assessment of the transformer's conditions

1. Introduction

1.1 Dataset

The dataset employed to train the machine learning algorithms contained 24 typical transformer parameters such as nameplate data, DGA, oil quality, insulation power factor, etc. As illustrated in Table 1 and Table 2, it provides a general statistical description of each parameter for the whole dataset.

1.2 Machine learning training with 10-fold cross-validation

The training was achieved by first random partitioning the original dataset with 1,000 transformers into two subsets, in which one dataset contained data for 800 transformers (training dataset), and the remaining 200 transformers were used as validation or test dataset. The training process was a supervised learning based on a 10-fold cross-validation procedure with 3 repeats, yielding 30 output accuracies for each machine learning algorithm [2-5], with each accuracy corresponding to each fold in a given repeat process. The supervised learning was applied with the support of human experts who have analyzed the same 1,000 cases provided to the machine learning algorithms.

Machine learning algorithms

The following 12 ML algorithms were trained and compared in the present work:

Linear algorithms

1. General linear regression (logistic regression) - GLM
2. Linear discriminant analysis – LDA

Non-linear algorithms

3. Classification and regression trees (CART)
4. C5.0 (a type of CART algorithm)
5. Naïve Bayes algorithm (NB)
6. K-nearest neighbor (KNN)
7. Support vector machine (SVM)

Ensemble algorithms

8. Random forest (stochastic assembly of a large number of CART algorithms)
9. Tree bagging (tree bagging)
10. Extreme gradient boosting machine (XGB tree)
11. Extreme gradient boosting machine (XGB linear)

Machine learning algorithms have the possibility to estimate a part of the missing data, which is extremely important in the transformer diagnostics application

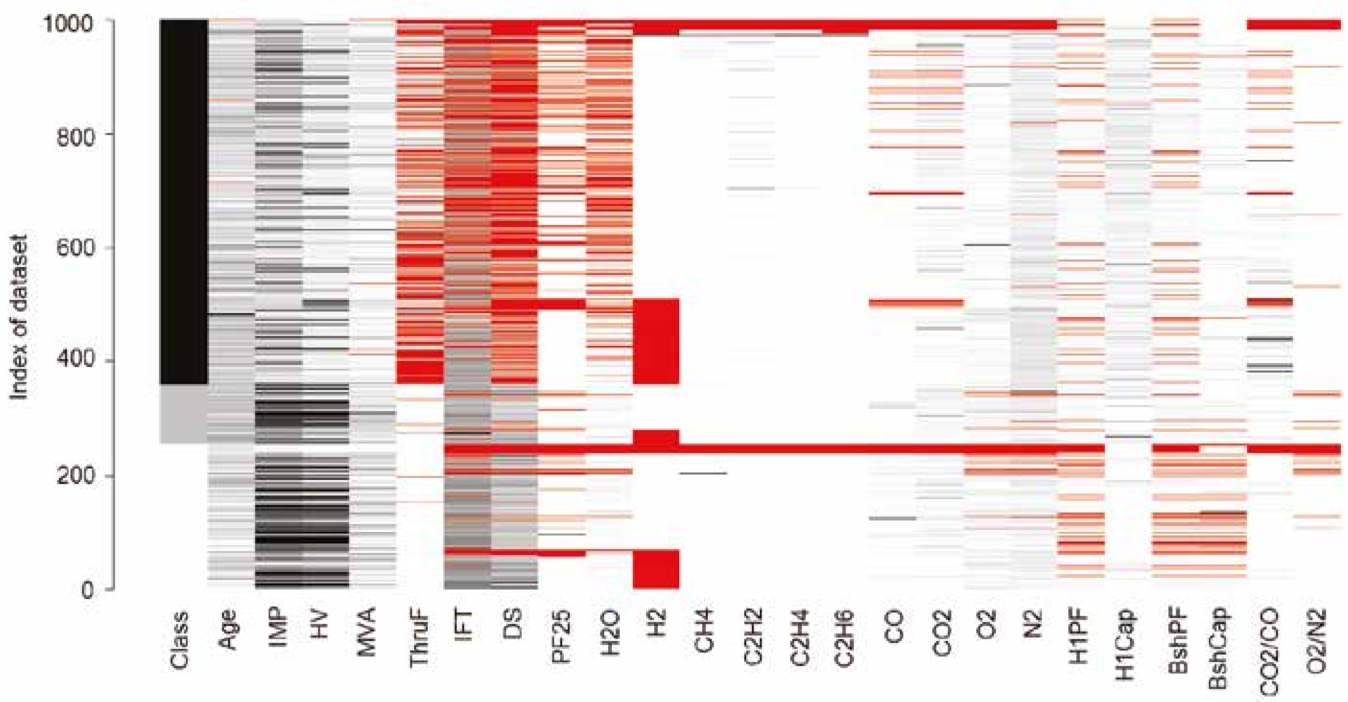

Figure 2. Map of missing values in the 1,000 cases used in the current paper to train and test the machine learning algorithms. Red lines show missing values for each column of data. Greyscale color shows available data, varying from low numbers (white) to high numbers (black).

12. Artificial neural networks (ANN – not deep learning yet)

The following section describes the meaning of statistical learning for each algorithm.

2. Statistical learning process

Statistical learning has a different interpretation for each of the above-indicated algorithms. In linear regression, for example, the learning process is associated with the optimal search of the linear model coefficients that best correlate inputs to outputs in a given problem. In the classification and regression trees, the training is related to a statistical method that optimizes the breakdown of the feature space (transformer parameters) into a decision tree, capable of classifying transformers based on the class distribution inside the tree. The support vector machine (SVM) is the so-called "widest band classifier" that optimizes the separation between different classes in a given dataset. In the neural networks method, the learning process is related to finding the optimal distribution of weights interconnecting multiple nodes in different layers until the classification error reaches a maximum acceptable threshold. The fact of the matter is that, although there are different algorithms and learning methodologies, the so-called "learning" is only possible due to the robust statistical procedures applied to each individual algorithm, through repetition of thousands of examples of different class types, until each algorithm is capable of outputting an acceptable level of accuracy.

Each random case illustrated in Fig. 1, out of the 800 training cases, contains the features and classification (expert judgment) necessary in the learning phase of each ML algorithm. In the end, each algorithm maps input data to output class in a statistical process that is characteristic of that specific algorithm. The supervised learning takes place in a comparison between the output of each individual machine learning algorithm and the one posted by the human expert for each individual example. An error or cost function is defined, and a proper statistical process is employed to minimize such a cost so that each algorithm will provide the best possible accuracy based on each model's training strategy. After the learning, algorithms are tested against the 200 unseen cases, and another accuracy is calculated. An interesting method of showing the accuracy of such a test is through the so-called "confusion matrix" to be described in the sections below.

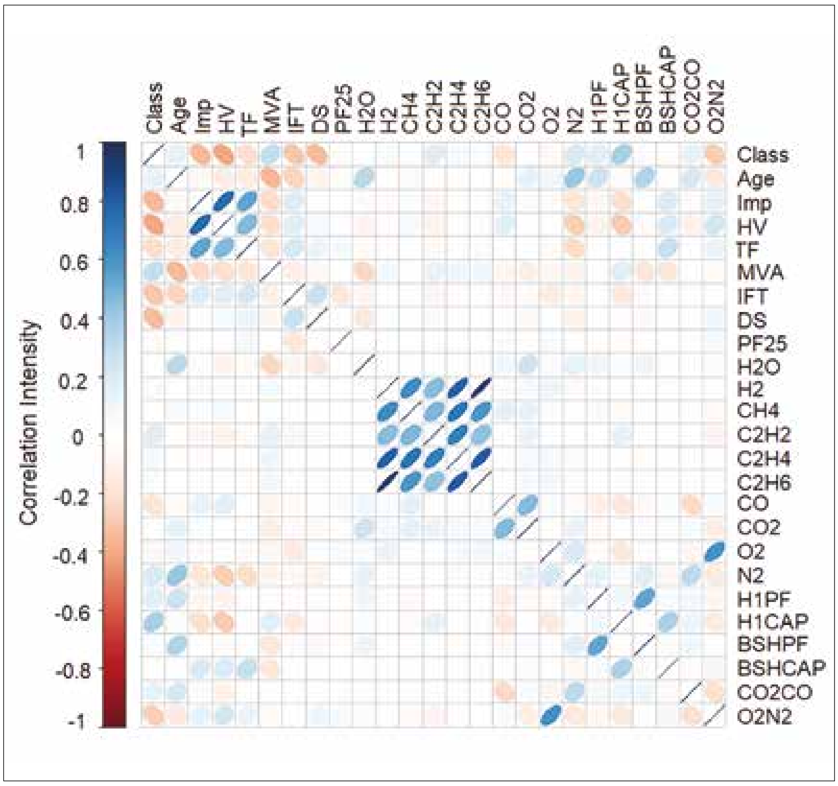

Figure 3. Correlation matrix of all 24 variables used in the study. Blue ellipses indicate a positive correlation; red ellipses show a negative correlation. Color intensity is proportional to correlation. Blank squares denote that there is no correlation.

The condition assessment data generated by the transformer experts have been used to train the various types of ML algorithms, and their accuracy was compared

3. Handling missing data

Missing data is perhaps the single most important aspect of any machine learning technique, and it is also extremely important in any transformer diagnostic process since human experts are typically forced to make decisions based on incomplete data. Fig. 2 shows a concise map of the actual missing data in the current dataset of 1,000 transformers.

There are several possible approaches to the problem of missing data, but one can say for sure that missing data is like a medical issue: it will not go away just because it has been overlooked. A human expert will intuitively handle missing data by, for example, assuming a missing parameter (say bushing power factor) is *normal* and, as such, will not influence the decision about the condition of the transformer. This is called "*single value imputation*" or "*educated guess*" since it replaces missing data with "normal" values. The most common imputation procedures are:

a) Single imputation (educated guess, mean or even median value of a distribution),
b) Feature correlation (make the column of missing data a function of all other parameters),
c) Multiple imputations (find the probability distribution function that best adheres to the data),
d) Use of probabilistic belief propagation algorithm (such as in Bayesian networks).

The algorithms that showed the best performance were those based on aggregation or ensemble of classification and regression trees, with an accuracy close to 97 %

Statistically speaking, a single value imputation (like in the educated guess or in the replacement of missing data by the mean or median, for example) may work well in certain applications but suffers from a possibly significant change in the original distribution of data. Another very powerful technique is the so-called statistical multiple imputations, although it is not of trivial application. The idea is to replace each missing datum with a randomly selected value from the actual probability density function that best fits the remaining data for that parameter. Feature correlation may also work, but it depends on a complex analysis regarding the level of correlation between features and the target variable (class, in the current example), as illustrated in Fig. 3. Several methods have been tested, but the simple imputation by the median showed good enough results in the present work.

4. Best performing machine learning algorithms

After replacing the missing data for each transformer, and training and testing all 12 machine learning algorithms with the available dataset, duly analyzed by transformer experts, the algorithms that showed the best performance were those based on aggregation or ensemble of classification and regression trees (CART). One should mention that no optimization procedure was applied to any of the tested algorithms and that the so-called "deep learning" was not employed with the artificial neural networks.

4.1 Principle behind CART

A full explanation of entropy and CART is beyond the scope of the present paper but interested readers may find a wealth of information in the references below and further. Finally, it is important to mention that tree bag, random forest, and the extreme gradient boosting machines (xGBM1, xGBM2 in the present work) are different forms of association of multiple CARTs, so that statistical combinations of weaker algorithms may lead to much stronger outputs.

5. ML algorithms output

Fig. 4 below shows the boxplots with comparative results of training accuracy for the 12 described machine learning models. Notice that the top 5 best performing models are all variations and ensembles of CART, and their major

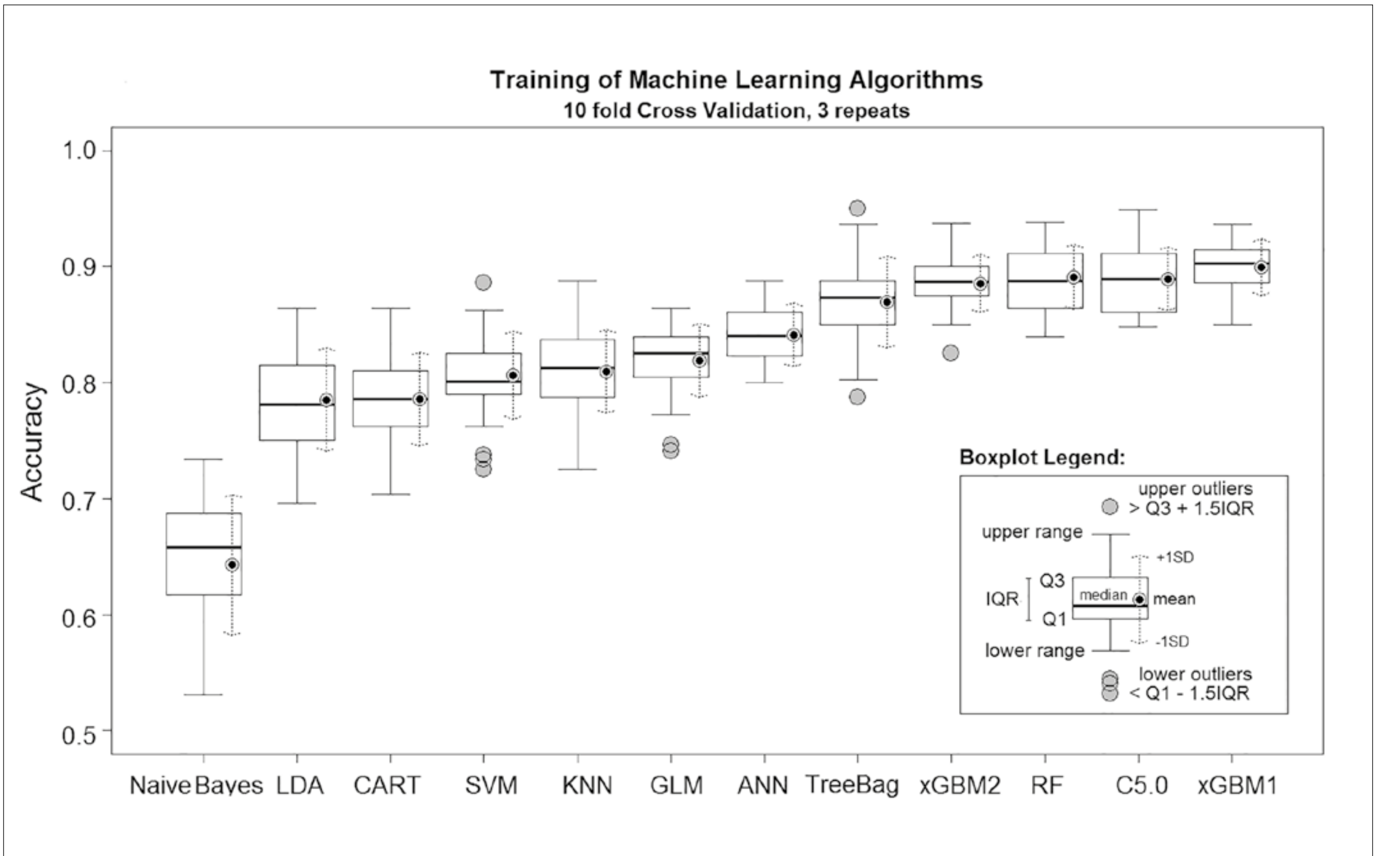

Figure 4. Comparative accuracy of machine learning algorithms after training 12 models with 80 % of the available data, by 10-fold cross-validation (CV) and 3 repeats. The ML algorithms were Naïve Bayes, linear discriminant analysis (LDA), classification and regression trees (CART), general linear model (GLM), support vector machine (SVM), K-nearest neighbor (KNN), artificial neural networks (ANN), tree bagging, extreme gradient boosting machine (xGBM1 and xGBM2), random forest (RF) and C5.0.

differences are in the process of building multiple trees and their combinations that will best separate the data after learning from the training dataset. The test accuracy is obtained by comparing the output of the system when classifying data that were not used during training (200 new cases not used during training) against the human experts' opinion for those new cases. This is typically given in the format of the so-called *confusion matrix* illustrated in Table 3 for the best performing method *Extreme Gradient Boosting Machine* 1 (xGBM1).

The machine learning algorithms have shown an impressive accuracy when analyzing complex power transformer data, however, human expert judgment is crucial in their training process

6. Discussion and conclusions

The machine learning algorithms have shown an impressive accuracy when analyzing complex power transformer data without using any engineering model whatsoever. In other words, the algorithms were not provided with reference levels or flags to indicate that a given parameter was within the acceptable range or outside the "normal" range. The 12 ML models were only provided with the final classification between *green*, *yellow*, and *red* previously established by transformer human experts. The best performing algorithm (xGBM1) presented near 97 % accuracy when analyzing the 200 new test cases unseen during training. It missed one *green* case that was "wrongly" but conservatively classified as *red*, 3 *yellow* cases that were wrongly classified as *green*, and 3 *yellow* cases that were wrongly classified as *red*. No red case was wrongly classified. The significant number of misses in practical terms is 3 *yellows* classified as *green* cases out of 200, leading to 3/200 = 1.5 % real miss, since the other misses were conservative and would not lead to any unfavorable situation like a possible failure. The paper has also demonstrated the importance of human expert judgment in the training and learning process of the ML algorithms, particularly with respect to power transformer diagnostics.

Bibliography

[1] L. Cheim, *"Machine Learning Tools in Support of Transformer Diagnostics"*, Cigre Paris Session, August 2018, paper A2-206 (*Best Paper Award*)

[2] J. H. Friedman, R. Tibshirani, T. Hastie, *"The Elements of Statistical Learning"*, 2nd. Edition, Springer 2009

[3] M. Kuhn, K. Johnson, *"Applied Predictive Modelling"*, Springer 2013

[4] J. Berger, *"Statistical Decision Theory and Bayesian Analysis"*, 2nd Edition 1993

[5] I. Witten, E. Frank, M. Hall, *"Data Mining: Practical Machine Learning Tools and Techniques"*, 3rd edition, Elsevier 2011

Dr. Luiz Cheim

Dr. Luiz Cheim is a Senior Principal R&D Engineer at Hitachi ABB Power Grids, with over 30 years' experience in the power transformers industry. His major activities and interests are in the development of transformer condition assessment, performance models and algorithms, as well as new sensors and state-of-the-art monitoring technologies. Dr. Cheim developed the transformer algorithms in Ellipse APM solution and is the proponent of the new Transformer Inspection Robot (TXploreTM). In August 2018, he was granted the Best Paper Award by Cigre organization in Paris, Study Committee A2-206/PS2. He has over 20 patents granted or filed, over the last 10 years alone, and has been recently selected to represent ABB at the Public Utilities Magazine (PUC, November 2019 issue) as a top innovator. He has been a long-standing member of both Cigre and the IEEE and has taken several active roles in both organizations.

Table 3. Confusion matrix and statistics (200 new test cases, ML = xGBM1)

	Human expert classification ➔		
ML Prediction 🡣	**Green**	**Yellow**	**Red**
Green	61	3	0
Yellow	0	14	0
Red	1	3	118
Totals	62	20	118

Algorithm accuracy = (61 + 14 + 118) / 200 = 96.5% Eq. (1)