Article

Two-Layer Ensemble-Based Soft Voting Classifier for Transformer Oil Interfacial Tension Prediction

Ahmad Nayyar Hassan and Ayman El-Hag *

Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada; anhassan@uwaterloo.ca

* Correspondence: ahalhaj@uwaterloo.ca; Tel.: +1-519-277-2984

Received: 16 March 2020; Accepted: 3 April 2020; Published: 5 April 2020

Abstract: This paper uses a two-layered soft voting-based ensemble model to predict the interfacial tension (IFT), as one of the transformer oil test parameters. The input feature vector is composed of acidity, water content, dissipation factor, color and breakdown voltage. To test the generalization of the model, the training data was obtained from one utility company and the testing data was obtained from another utility. The model results in an optimal accuracy of 0.87 and a F1-score of 0.89. Detailed studies were also carried out to find the conditions under which the model renders optimal results.

Keywords: Interfacial tension; machine learning; transformer oil parameters

1. Introduction

Power and distribution transformers are one of the most significant and expensive assets in any power system grid. Internal faults in the transformer such as partial discharge (PD) or overloading may lead to insulation deterioration and eventually to complete failure of the transformer. This causes catastrophic transformer outages, which lead to both direct and indirect costs. Hence, assessing the transformer's health condition and continuous monitoring of the insulation system ensures its satisfactory performance, maintains efficiency, and prolongs its lifetime.

Together, the oil and insulation paper constitute the transformer's insulation system and have two important functionalities [1]: to act as an insulation to insulate high voltage from the ground and as a coolant to dissipate the generated heat efficiently. The overall health condition of a transformer depends largely on the state of its oil and paper insulation system [2]. Ageing of the transformer oil, which is a natural process in any insulation system, results in the formation of sludge particles, which in turn damages the properties of other insulation components like cellulose paper in the transformer winding. Therefore, it becomes very critical to monitor the transformer oil quality by regularly inspecting samples using different electrical, physical and chemical methods.

There are several elements that can be measured to quantify the transformer oil ageing condition. They can be classified into three categories: dissolved gas analysis (DGA), furan content and oil tests. DGA analysis is conducted mainly to detect the emergence of different faults inside the transformer winding, like arcing or PD activities. Furan, on the other hand, is measured to estimate the health condition of the transformer paper insulation. Finally, oil tests reveal information about several aspects of the electrical, physical and chemical condition of the transformer oil. For example, oil tests include water content, breakdown voltage (BDV), interfacial tension (IFT), dissipation factor (DF), color and acidity [3]. Conducting such tests routinely adds to the overall maintenance cost of the transformer. The cost of the oil sample varies from one country to another, for example, testing one oil sample (BDV, acidity, water content and IFT) in Dubai would cost around USD 1500 [4]. Thus, instead of testing these samples, it is more economical to predict their values. This is particularly so, given the recent advancement in machine learning (ML) algorithms as they have proven efficacy in many applications.

Among all oil tests, the IFT conducted as per the ASTM D971 standard has the highest cost, requires specific expertise and specialized instruments [2].

The IFT of mineral oil is related to the aging of the oil sample. Mineral oil is essentially a non-polar saturated hydrocarbon fluid and when it undergoes oxidative degradation, oxygenated species are formed such as carboxylic acids, which are hydrophilic in nature. The presence of these hydrophilic components in the transformer oil can influence the chemical (acidity), electrical (BDV), and physical (IFT) properties of the oil sample. Measuring the IFT is basically conducted by measuring the surface tension of an oil sample against that of water, which is highly polar. The more the two liquids (oil and water) are similar in their polarity, the lower the value of the surface tension between them. Thus, the higher the concentration of hydrophilic materials in the oil sample, the lower will be the interfacial tension of the oil measured against water. So, the magnitude of the IFT is inversely related to the concentration of the hydrophilic degradation products that result from the aging of the oil. Since hydrophilic materials are usually highly polar and thus not very soluble in non-polar oil, the presence of these species can result in sludge formation that in turns contributes to the further degradation of the transformer insulation system [5].

Recently, the application of machine learning in transformer assessment has become more widespread. Most of the reported studies have concentrated on predicting the transformer health index (HI). The transformer HI is a calculated number that estimates the health condition of oil-filled transformers [6]. In [7], a fuzzy logic-based approach was used to predict the HI value using the oil quality, dissolved gas and furan content parameters as inputs. The reported classification success rate was 97% based on a three-class classification system. Moreover, in [8], an artificial neural network (ANN) approach was proposed to classify the condition of the transformer based on the predicted HI value. The input features used in this model are oil test parameters, DGA and furan content. Based on the testing outcomes, 97% of the testing samples were correctly classified into a three-class condition problem. To further enhance the HI calculation, a reduced model was implemented [9]. It has been found that a HI with relatively high accuracy can be achieved with few tests.

Few studies have been conducted to estimate transformer oil characteristics such as water content and breakdown voltages [10–12]. A cascaded ANN was used to predict transformer oil parameters using the Megger test [10]. Also, ANN with stepwise regression was implemented to predict the transformer furan content [11]. These studies were only conducted on a moderate number of transformers, which makes it hard to generalize the conclusions. A polynomial regression model has been developed to predict the breakdown voltage as a function of the transformer service period and other oil testing parameters like total acidity and water content. Except for a few cases, the percentage error between the actual and predicted values of transformer breakdown voltage was less than 10% [12]. However, the model needs the water content and total acidity as an input to predict the breakdown voltage. Hence, while this model saves the cost of conducting the breakdown voltage test, there is still a need to conduct two other oil tests. Moreover, the values of the water content and total acidity need to be collected at different time intervals to formulate the mathematical model and predict the value of the transformer oil breakdown voltage, which adds to the overall transformer oil maintenance cost.

In this paper, the authors investigated the ability of ensemble methods to predict the class of IFT. An ensemble method is a learning technique that uses several base models in order to produce one optimal predictive model [13]. The key idea behind any learning-based problem is to find a single model that best predicts the output. Instead of depending on only one model and hoping that it might be the most accurate we can come up with, ensemble methods take a myriad of models into account and leverage these to produce one final model. In our problem, we use two layers of these ensemble models using soft voting. The concept behind a voting classifier is to combine different machine learning classifiers and use a voting criterion of some sort to predict the class label [13]. A classifier of this sort can balance out the individual weakness of the classifiers involved. There are two types of voting classifiers: (i) majority/hard voting and (ii) soft voting. The former uses the mode of the class labels predicted by the individual classifiers while the later returns the class label as argmax of the sum

of predicted probabilities. In other words, each classifier is assigned a weight and the class label that has the maximum weighted average is selected as the output class label.

2. Materials and Methods

2.1. Dataset

Two different datasets were used in this study, i.e., a training dataset and a testing dataset. The training set consists of the oil tests of 730 transformers with a high voltage rating of 66 kV and power rating ranges from 12.5 to 40 MVA. The testing dataset consists of 36 transformers with a high voltage rating of 13.8 kV and power rating ranges from 0.5 to 1.5 MVA. It is apparent that these two datasets have no overlap in terms of the transformer rating or in terms of their geographical location as they were obtained from two different countries in the Gulf region. While the aging mechanism may be different in these two different categories of oil-filled transformers due to the different loading conditions, the impact of aging on the oil chemical, electrical and physical properties will be similar. The input features included in the dataset are water content, acidity, breakdown voltage, dissipation factor (DF) and color and the output variable is the interfacial tension (IFT). The output feature vector was divided into two categories (good and bad) based on their values. Figure 1 depicts the distribution of data between the two classes for both the training and testing sets. Oil samples with IFT ≥ 30 dyne/cm are considered as “Good” oil, otherwise the sample is considered as “Bad”.

Figure 1. The data distribution of the (a) training dataset and (b) testing dataset.

2.2. Data Pre-Processing

Data pre-processing was divided into two main steps. The first step involves outlier removal while the second step includes normalization of the data. An outlier is an observation that lies outside the overall pattern of a distribution. Outliers severely skew the performance of a classifier, therefore, there is a strong requirement to remove them if any exist. In order to remove them, the mean and standard deviation of each column is computed and any observation whose absolute difference from the mean exceeds three times the standard deviation is detected as an outlier and is removed. In order to make sure that all the features have the same significance, at the start it is very important that the scales of all the features remain the same. In order to make sure that all the feature values are in the same scale, each individual feature is converted into a number between zero and one. This is done by min-max scaling, that is, for each reading, the minimum value is subtracted and the result is divided by the maximum value of that feature.

2.3. Data Visualization

The heatmap of correlation was calculated to see whether there is multi-collinearity among different features or not. The heatmap is shown in Figure 2 and it depicts the correlation score between different input features.

Even though there is some relatively high correlation between some of the features like acidity and color (correlation = 0.74) and DF and acidity (correlation = 0.75), no correlation between any of the features is greater than 0.80, which indicates that no multi-collinearity exists. It is worth mentioning here that the calculated correlation is for the whole 730 transformers. The correlation will be more evident for severely aged transformers. For example, the correlation between IFT and acidity drops from -0.72 to -0.77 if the correlation was calculated for transformers with an IFT value less than 20.

	Water	acidity	BDV	DF	Color	IFT
Water	1	0.36	-0.25	0.33	0.31	-0.36
acidity	0.36	1	-0.088	0.75	0.74	-0.72
BDV	-0.25	-0.088	1	-0.056	-0.11	0.092
DF	0.33	0.75	-0.056	1	0.71	-0.71
Color	0.31	0.74	-0.11	0.71	1	-0.88
IFT	-0.36	-0.72	0.092	-0.71	-0.88	1

Figure 2. Correlation matrix of the input features in the dataset.

2.4. Machine Learning Model Architecture

The machine learning model proposed in this paper is a two-layer ensemble-based soft voting classifier, which uses a total of eight different classifiers. The first layer consists of two main blocks with four classifiers in each block. The first block consists of four classical machine learning algorithms, which are non-ensemble-based followed by a soft voting classifier module. These four learning algorithms are naïve Bayes, support vector machine with radial basis function as the kernel function, logistic regression and k-nearest neighbors. The output of each classifier is then passed to the voting classifier, which does soft voting based on the argmax of the sum of predicted probabilities of the class labels. The second layer consists of four ensemble-based classifiers and each one's output is again fed into a separate voting classifier that performs soft voting. The four ensemble classifiers used in this block are random forest, decision tree-based bagging model, Ada-boost and gradient boosting classifier. The output of each of the two blocks in layer one is finally fed to another voting module which performs soft voting and generates an output label. The block diagram shown in Figure 3 demonstrates the structure of the model.

The key idea behind using the consensus of multiple classical machine learning algorithms in the first block is to overcome the limitations of some algorithms, as the shortcoming of one algorithm might be a strength of another. For example, naïve Bayes assumes that the presence of one feature is unrelated to another feature [14]. This assumption might not be valid in many physical systems due to the inherent correlation among the predictor variables as shown in Figure 2. However, it is extremely fast and easy to compute for generating the output labels for the test set. K-nearest neighbors, on the other hand, assumes that similar values that are close to each other perhaps belong to the same class [15]. In order to assign label to a test data point it loads all the labelled data points in the memory and computes the distances between all the label input data and the test data points. Furthermore,

it decides the first K neighbors on the criterion of the smallest distance, and then finally computes the mode of the class label of these K neighbors as the output label. As this algorithm keeps everything in the memory during testing and computes all the distances while making a prediction, it is extremely slow. However, it generates a decision boundary between classes that is highly nonlinear, and therefore accommodates linear classifiers like logistic regression.

Figure 3. Block diagram of the proposed two-layer ensemble-based soft voting classifier.

The concept remains the same in the second block as well. The unanimity of numerous different algorithms will reduce the weaknesses and enhance the strength of the combined model in general. However, in order to deal with the issue of limited data, this block uses the idea of bootstrap plus aggregation and ensemble models. Bootstrap plus aggregation means that from the main pool of data, a large number of samples are collected with replacement to make multiple datasets. Then, many models are trained using the boosted number of data points across multiple datasets generated from the original data, and their outputs are combined. Bootstrap plus aggregation is commonly known as bagging. Not only does bagging help in dealing with the issue of a lesser volume of data, it also reduces variance in highly variant models like decision trees, which severely overfit the training data. Decision trees (mainly CART) are greedy in nature—the splitting variable is chosen on local and not global minimization of the error [16]. This in turn results in decision trees having similar structure and predictions that are highly correlated to each other. This problem is solved by using random forests. Random forests adjust the splitting criterion such that the resulting predictions from different trees have less correlation among each other. In order to make sure that the split is not greedy, random forests ensure that the learning algorithm looks through all variables and their values in order to select the most optimal split point. Random forests run multiple trees in parallel, therefore it remains unbiased towards miss-classifications.

With all these learning algorithms that combine both classical and ensemble via the final voting classification layer, the model learns the data in the best possible manner and renders highly accurate results. All the hyperparameter tuning is done using grid search cross-validation, which uses 5-fold cross-validation. Grid search cross-validation does an exhaustive search over specified parameter values for an estimator.

2.5. Evaluation Metrics

In the testing phase, the generated classification model along with the chosen features was evaluated using the testing dataset. A confusion matrix can be constructed to show the actual and predicted classifications. Table 1 shows a confusion matrix for binary classification problems, in which the class is either Yes or No. The size of the testing data is determined in relation with the over-all size of available data. Since the testing data have known classes, the classifier accuracy rate can be calculated. The F-measure is another reliable statistical evaluation measure that is widely used to evaluate and compare the performance of classifiers on binary classification problems. The F-measure is the harmonic mean of the precision (P) and recall (R), as shown in Equation (1).

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F_1 = 2 \frac{P \times R}{P + R} \quad (1)$$

where P is the ratio of the correctly predicted positive instances over all predicted positive instances, i.e., true positive (TP) and false positives (FP); R is the ratio of correctly predicted positive instances over all actual (real) positive instances, i.e., true positive (TP) and false negatives (FN); and $F1$ is the harmonic mean of precision and recall. P , R , and $F1$ are useful measures for binary classification problems and imbalanced classification problems.

Table 1. Confusion matrix of binary classification problems.

Really Is	Classified As	
	Yes	No
	Yes No	TF FP FN TN

As a generalization for multi-class classification problems, the overall classification accuracy measure is given in Equation (2).

$$Accuracy\ Rate = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

3. Results and Discussion

The aforementioned evaluation metrics of the different combined and individual classifiers are presented and discussed in the following subsections.

3.1. Classification Accuracy of Both Individual and Combined Classifiers

As previously addressed, the first dataset (730 transformers) was used for training purposes and the second dataset (36 transformers) was used for testing. In a previous publication, the authors used the dataset for the 730 transformers for both training and testing [17]. An overall accuracy of 95.5% was achieved when 10-fold cross-validation was used for the training and testing of the data. The accuracy, F1-score, precision and recall for both the classical and ensemble machine learning algorithms are shown in Tables 2 and 3, respectively. It is evident from Tables 2 and 3 that there is no single classifier that gives the best results for all metrics. Also, the maximum overall accuracy achieved was 86.1% using AdaBoost as a classifier. Moreover, the only classifier that did not show any FN is the naïve Bayes classifier. However, it shows the highest FP among all individual classifiers. It is interesting to note that except for naïve Bayes, the precision is higher than the recall for all other classifiers.

Table 2. Evaluation metrics of each of the individual classical machine learning classifiers.

Classical Machine Learning Algorithm	Accuracy	F1-Score (F1)	Precision (P)	Recall (R)
Naïve Bayes	0.833	0.892	0.806	1.0
K-Nearest Neighbors	0.777	0.826	0.904	0.760
Logistic Regression	0.833	0.874	0.913	0.84
Support Vector Machine	0.75	0.809	0.863	0.76

Table 3. Evaluation metrics of each of the individual ensemble classifiers.

Ensembling Method	Accuracy	F1-Score (F1)	Precision (P)	Recall (R)
Bagging with Decision trees	0.75	0.791	0.944	0.68
Random Forests	0.833	0.870	0.952	0.80
AdaBoost	0.861	0.875	0.952	0.81
Gradient Boosting	0.857	0.873	0.934	0.82

After combining the classifiers, a marginal improvement in the classification accuracy is evident only for the two-layer ensemble-based technique, as depicted in Table 4. Nevertheless, the two-layer ensemble-based technique does not show superior performance in other metrics such as *F1* and recall. Thus, it can be stated that no ML algorithm can guarantee the best performance for all evaluation metrics. Moreover, and similar to the individual classifiers, the precision was higher than the recall for all ensembling blocks.

Table 4. Evaluation metrics of each of the individual and combined ensembling blocks.

Ensembling Method Blocks	Accuracy	F1-Score (F1)	Precision (P)	Recall (R)
Ensemble with classical ML models	0.861	0.898	0.917	0.88
Ensemble with tree based models	0.833	0.870	0.952	0.80
Two layer ensemble-based	0.871	0.894	0.955	0.84

3.2. Classification Under Reduced Number of Features

Reducing the number of tests required to predict the IFT will further reduce the cost of transformer oil assessment. Different techniques have been implemented to change the size of the input feature vector. Both principal component analysis (PCA) and linear discriminant analysis (LDA) are used to vary the feature space. PCA is a feature extraction technique. It projects the data onto a lower dimensional feature space by using an orthogonal transformation based on the maximization of variance. The resulting dimensions are reduced in number with respect to the total number of features and are also orthogonal (have no overlap) to each other. PCA is performed on the dataset to find the number of transformed dimensions that capture the maximum variance. Figure 4 shows the variance shared by each component. The first three components are responsible for most of the data variance (93%), therefore we tested the proposed model using only the first three components.

Testing the PCA proposed model with reduced number of features reduces the accuracy to about 60.4%, as shown in Table 5. A drastic drop in all other measuring metrics is also evident. PCA is an unsupervised dimensionality reduction and therefore it does not consider the information of class labels. This results in the generation of transformed dimensions that do maximize variance of data but also make the separation between classes difficult.

Figure 4. Relative variance of transformed components shown through a bar graph and explicit shared variance values.

Table 5. Evaluation metrics of reduced input feature vector.

Feature Extraction/Selection Technique	Accuracy	F1-Score (F1)	Precision (P)	Recall (R)
PCA	0.604	0.683	0.722	0.587
LDA	0.75	0.809	0.863	0.76
Extra tree classifier	0.833	0.875	0.913	0.84

In order to solve this problem, we use LDA which is a supervised feature extraction technique. The key concept behind LDA is that in order to find the new axis, the optimization problem should be such that it minimizes the intraclass (within class) variance and maximizes the distance between projected class means so it is easier to do the classification once the dimensions have been reduced. In order to have a fair comparison with PCA, the number of dimensions were kept constant at three. Testing the proposed model on reduced dimensions with LDA improves the accuracy to 75%, but this is still less than the accuracy achieved on the original dataset. This means that even though the last two components have less share in the variance, they are important to accomplish the previously achieved classification accuracy.

As an alternative to PCA and LDA, the features are selected directly based on their importance and not by transforming them into a different domain and then reducing the dimension. In order to select the top three features, we rank the features by assigning them relative importance. This is done using the extra tree classifier, which is a variant of a decision tree. However, when looking for the best split to separate the samples of a node into two groups, random splits are drawn for each of the selected features and the best split among these is chosen. The results for feature importance using the extra tree classifier are shown in Figure 5, which gives color, dissipation factor and acidity the highest scores. Therefore, we checked our model with only this subset of features. This agrees with the correlation matrix that shows that these three features have the highest correlation with the IFT. Using only these features results in an accuracy of 83.3%, which is close to the accuracy using the original dataset.

Figure 5. The feature importance bar graph and relative importance score using extra tree classifier.

3.3. Classification Under Balanced Number of Input Features

One of the problems in the training dataset is the imbalance between the two classes as depicted in Figure 1. To overcome this problem, two different approaches were investigated, namely, up sampling and down sampling. In a previous study, up sampling improved the classification accuracy of furan content in transformer oil [18]. Both the up sampler and down sampler were implemented using a random sample from the original data. In the case of up sampling, random sampling without replacement is done from the minority class to increase the number of data points in the minority equal to the majority class. For down sampling, random sampling without replacement is done from the majority class to down size the number of data points to the number of samples in the minority class. The results of using these techniques is shown in Table 6. While both techniques did not contribute to the improvement of the classification accuracy, up sampling resulted in relatively better results than down sampling. This could be attributed to the low number of testing samples and hence improving the training data may not result in improvement in the testing samples. As samples are selected at random for down sampling, there is a high probability that data points that are essential in causing the separation between the two classes are not selected and hence the output is not accurate.

Table 6. Evaluation metrics of two different data balancing techniques.

Input Data Balancing Technique	Accuracy	F1-Score (F1)	Precision (P)	Recall (R)
Up Sampling	0.805	0.844	0.95	0.76
Down Sampling	0.72	0.773	0.894	0.68

4. Conclusions

Transformer oil IFT is a very important parameter that needs to be evaluated to assess the condition of transformer oil. Compared to other oil tests, IFT is relatively harder and more expensive to conduct. In this paper, the viability of using multiple machine learning algorithms (as individuals and combined) to predict the transformer oil IFT was investigated. In this investigation, two different datasets for transformers from two different geographical locations were used with one used for training and the other used for testing. This was implemented to ensure the robustness of the proposed method. No single technique showed superior performance on all measured metrics. However, combining different ML algorithms and applying voting technique generally resulted in better measured metrics

than individual ML algorithms. Moreover, it was found that reducing the input feature vector using PCA resulted in a significant reduction in all measured metrics. However, when feature selection was based on the feature correlation with the IFT, much better results were achieved. One of the drawbacks of the proposed technique is the shortage of testing samples. As a future work, the authors will try to collect more samples to further validate the proposed algorithm.

Author Contributions: Conceptualization, A.N.H. and A.E.-H.; methodology, A.N.H.; software, A.N.H.; validation, A.N.H.; formal analysis, A.N.H.; investigation, A.N.H. and A.E.-H.; resources, A.N.H. and A.E.-H.; data curation, A.E.-H.; writing—original draft preparation, A.N.H.; writing—review and editing, A.E.-H.; visualization, A.N.H. and A.E.-H.; supervision, A.E.-H.; project administration, A.E.-H.; funding acquisition, N/A". All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bernard, N.; Cucek, B. Methods for monitoring age-related changes in transformer oils. In Proceedings of the 2014 IEEE International Conference on Liquid Dielectrics, Bled, Slovenia, 30 June–3 July 2014.
2. Abu-Siada, A.; Abu Bakar, N. A novel method of measuring transformer oil interfacial tension using UV-Vis Spectroscopy. *IEEE Electr. Insul. Mag.* **2016**, *32*, 7–13.
3. Singh, J.; Sood, R.Y.; Verma, P. The influence of service aging on transformer insulating oil parameters. *IEEE Trans. Dielectr. Electr. Insul.* **2012**, *19*, 421–426. [\[CrossRef\]](#)
4. Dubai Electricity and Water Authority. *Services Guides*, 5th ed.; Dubai Electricity and Water Authority: Dubai, United Arab Emirates, 2016.
5. Gray, I.A.R. Interfacial Tension. Available online: https://www.satcs.co.za/Interfacial_Tension-IARGray2007.pdf (accessed on 14 February 2020).
6. Jahromi, A.; Piercy, R.; Cress, S.; Service, J.; Fan, W. An Approach to Power Transformer Asset Management Using Health Index. *IEEE Electr. Insul. Mag.* **2009**, *25*, 20–34. [\[CrossRef\]](#)
7. Abu-Elanien, A.E.B.; Salama, M.M.A.; Ibrahim, M. Calculation of a Health Index for Oil-Immersed Transformers Rated Under 69 kV Using Fuzzy Logic. *IEEE Trans. Power Deliv.* **2012**, *27*, 2029–2036. [\[CrossRef\]](#)
8. Abu-Elanien, A.E.B.; Salama, M.M.A.; Ibrahim, M. Determination of Transformer Health Condition Using Artificial Neural Networks. In Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications (INISTA), Istanbul, Turkey, 15–18 June 2011.
9. Alqudsi, A.; El-Hag, A. Application of Machine Learning in Transformer Health Index Prediction. *Energies* **2019**, *12*, 2694. [\[CrossRef\]](#)
10. Shaban, K.; El-Hag, A.; Matveev, A. A Cascade of Artificial Neural Networks to Predict Transformers Oil Parameters. *IEEE Trans. Dielectr. Electr. Insul.* **2009**, *16*, 516–523. [\[CrossRef\]](#)
11. Ghunem, R.A.; Assaleh, K.; El-Hag, A.H. Artificial Neural Networks with Stepwise Regression for predicting transformer oil Furan Content. *IEEE Trans. Dielectr. Electr. Insul.* **2012**, *19*, 414–420. [\[CrossRef\]](#)
12. Wahab, M.A.; Hamada, M.M.; Zeitoun, A.G.; Ismail, G. Novel modeling for the prediction of aged transformer oil characteristics. *Electr. Power Syst. Res.* **1999**, *51*, 61–70. [\[CrossRef\]](#)
13. Saqlain, M.; Jargalsaikhan, B.; Lee, J.Y. A Voting Ensemble Classifier for Wafer Map Defect Patterns Identification in Semiconductor Manufacturing. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 171–182. [\[CrossRef\]](#)
14. Mitchell, T.M. Generative and Discriminative Classifiers: Naive Bayes And Logistic Regression. In *Machine Learning*; McGraw Hill: New York, NY, USA, 2015; Available online: <http://www.cs.cmu.edu/~jguestin/Class/10701-S05/slides/NBayesLogReg-2-05.pdf> (accessed on 14 February 2020).
15. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2002; ISBN 978-0-12-374856-0.
16. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)

17. Bhatia, N.; El-Hag, A.H.; Shabaan, K. Machine Learning-based Regression and Classification Models for Oil Assessment of Power Transformers. In Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT), Doha, Qatar, 2–5 February 2020.
18. Benhmed, K.; Shaban, K.B.; El-Hag, A. Cost Effective Assessment of Transformers Using Machine Learning Approach. In Proceedings of the 2014 IEEE Innovative Smart Grid Technologies Conference—Asia (ISGT ASIA), Kuala Lumpur, Malaysia, 20–23 May 2014.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).