Fault Detection in Three Phase Transformer Using Ensemble Model

Tejasvi Roop Singh, Viplove Verma, Shivam, Shivam Chawla, Sargam Ghorela, Sukirat Singh Mann and Prashant Singh Rana
Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala-147004, India

Abstract—Three phase transformers are an important part of electrical power distribution systems. This type of transformer is composed of primary and secondary windings and can be connected in either star or delta formations. Detection of faults in such high power transformers is crucial and can help in reducing such faults up to a great extent. The ensemble model is used for the detection of five faults in three phase transformer. Seven feature selection techniques and seven machine learning models have been used, out of which five best possible combinations were selected for ensemble model. Five different operating conditions are discussed namely normal condition, inrush, over excitation, internal fault and external fault. These faults are classified by 30 features i.e. 10 samples of three phase differential current have been used. 420 simulated samples have been generated using sim power systems of MATLAB/SIMULINK under different conditions of Y-Y transformer. Cross Validation is done to show the robustness and consistency of the best predictive models on which ensemble model is applied. The results show that on applying ensemble techniques, the accuracy improves significantly to about 85.652% in fault prediction based on the dataset.

Index Terms—Fault Detection, Three Phase Transformer, ARC, Random Forest, Gini, Internal Fault Simulations.

1 INTRODUCTION

An electric power system is a network of different electrical components installed at various locations to supply, transfer, store, and use electric power. The transformer is a major equipment of this system. Three phase transformers supply large loads and have large power distribution and hence they prove to be more economical. Three phase transformer has various advantages compared to the single phase transformer. It has better operating characteristics and is easier to construct for the same power handling capacity as offered by the single phase transformer. Therefore the protection of these transformers becomes a necessity to maintain the safety of power system. Differential protection is one such method. It uses the principle of comparing the primary and secondary currents of the three phase transformer using differential relay.[1] If there is any imbalance between primary and secondary currents the relay will initiate and inter trip both the primary and secondary circuit breaker of the transformer. Differential protection is one of the best methods for fault discrimination. Differentiating internal fault with other operating conditions is one of the major problems faced by the differential protection method. Artificial neural network trained using the PSO(particle swarm optimisation) is used for implementing differential protection.[5] It differentiates among the different types of faults in the transformer namely magnetizing inrush[9], normal condition, over excitation, internal and external fault condition in the transformer. The second harmonic component in transformer blocks the differential relay in three phase transformer. Magnetizing inrush current can be found by calculating the ratio of second harmonic power spectrum to the power spectrum found by the autoregressive process where current values depend upon the preceding values. Several methods have been proposed using artificial neural network and signal processing. Wavelet packet are also used for the protection of transformer. It differentiates between the different transformer fault conditions and magnetizing inrush condition by describing different types of features of differential current. Minimum length description data criteria is used for selecting the optimal wavelet.[16] Methods based on correlation are used for differentiating between short circuit condition and magnetizing inrush condition. Signals can be described in terms of time and frequency component by employing discrete wavelets. Correlation coefficient describes relationship between different wavelet energies at different scales of the signal resolution.[14] Three phase power transformer can also be protected using a multi-region adaptive differential relay. It is used to differentiate internal fault conditions from the disturbance based on the

• *Tejasvi Roop Singh, Viplove Verma, Shivam, Shivam Chawla, Sargam Ghorela and Sukirat Singh Mann are students of B.E.,Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala-147004, Punjab India.E-mail: tejasvichandi@gmail.com*

TABLE 1: Description of the transformer faults

Fault Class	Type	Description
1	Normal	No Fault
2	Magnetising Inrush	Circuit breaker closing at different voltage angles
3	Over-Excitation	Occurs when the angles are different
4	Internal Fault	Transformer secondary is shorted
5	External Fault	Fault at transmission lines

trajectory of differential current and weighted factors depending on the locus of differential current in the relay characteristic. Differential relay has a dual slope and is divided into three operating regions with each region having an associated weight factor.[3] PSO (particle swarm optimization) is used to find the ideal smoothing factor of probabilistic neural network. It determines the operating condition of transformer by calculating the ratio of voltage-to-frequency and amplitude of differential current. To distinguish between external fault and internal fault of a three phase power transformer, combination of probabilistic neural network (PNN) and discrete wavelet transformer (DWT) is used. To show a proper differentiation between internal faults and other disturbances such as over-excitation condition and inrush current, an advanced technique is developed which is based on the concept of support vector machine (SVM). To find the fault in radial distribution system, we use the support vector machine and artificial neural network models. To predict the faults which can occur in the three phase transformer, we propose a predictive model using various machine learning algorithms. Seven feature selection techniques and seven machine learning models are applied on the dataset generated using sim power systems of MATLAB/SIMULINK. 420 samples of current are generated using the simulation system under different types of fault conditions. The feature selection techniques used are Gini, Relief, OneR, Chi- Square, Pearsons Correlation, Consistency and BCBCSF whereas the machine learning models used are ARC, Decision Tree, Linear Classification, Neural Network, Node Harvest, Random Forest and Support Vector Machine. Our proposed model is then obtained by applying the ensemble technique. The predictive model will give robust and consistent results.

TABLE 2: Sample Dataset

Fault Class	F1	F2	F3	....	F28	F29	F30
2	-0.11455	-0.16691	-0.20828	....	0.308804	0.30001	0.2725
5	-926.131	-1335.84	.1615	....	0.2141	0.1993	0.165088
1	-0.001	-0.00102	-0.00094	....	0.000611	0.000328	0.00001
4	-0.00168	-0.00139	-0.00097	....	0.00003	-0.00053	-0.00097
3	0.00055	0.00105	0.00145	....	-0.00153	-0.0015	-0.00126
1	-0.00077	-0.00052	-0.00022	....	-0.0003	-0.00059	-0.00082
4	0.1141	0.052571	-0.01418	....	0.1192	0.1685	0.2013

TABLE 3: Description of the features

Feature	Information
F1-F10	Current in Phase A
F11-F20	Current in Phase B
F21-F30	Current in Phase C

2 Materials and Methods

2.1 Modelling and simulation of the power system to prepare the patterns

A 100 km transmission line has been incorporated into a three-phase 220/6.3 KV power system, as shown in Fig 2. They have been used to develop the required tests and training patterns. Sim power systems of MATLAB/SIMULINK were used for the simulation.Table 1 shows the different types of faults that can occur in a three phase transformer. Different voltage angles and different loads are the two different conditions for inrush current and over excitation state. Waveforms that are generated by different operating conditions are considered as an inputs

2.2 Data Set

420 simulated samples of currents have been generated using sim power systems of MATLAB/SIMULINK under different fault conditions. Table 2 shows a sample data set. These faults are classified by 30 features i.e. 10 samples of three phase differential current have been used and is shown in Table 3. Distribution of fault counts in data set is shown in Figure 1. Different types of current patterns for various fault conditions are shown in Figure 3.

3 Methods Applied

3.1 Methodology

The methodology is shown in Fig 4. The first step involves generating 420 simulations of currents using sim power simulation under different faults conditions. In the second step, data cleansing takes place which removes missing entries and duplicate values in the data set. In the third step we use seven different feature selection techniques mentioned below to select the required features. In the fourth step we use seven different types of classification based machine learning models which are trained on 70% data set with their default parameters and tested on remaining 30% of the data set. After the results are obtained, in the fifth step we choose the five best possible combinations and apply the ensemble model on them to obtain the accuracy. Finally, K-fold cross validation is performed to check the robustness and consistency of the model.

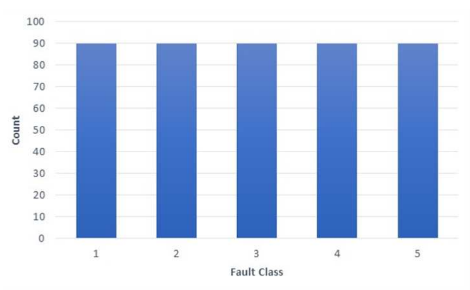


Fig. 1: Fault Counts in the Dataset

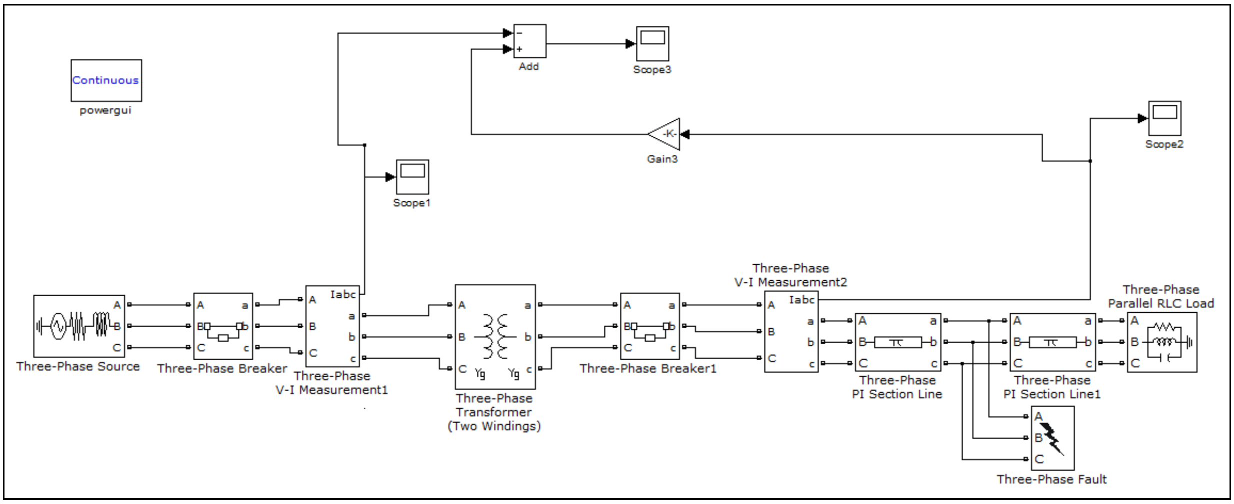
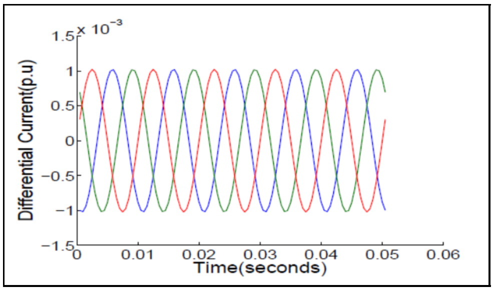
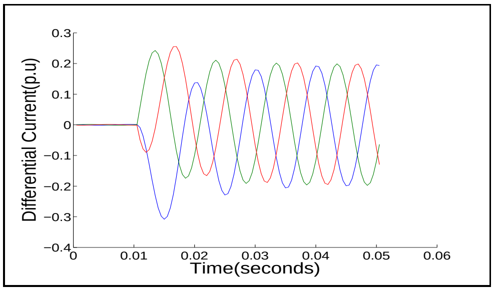


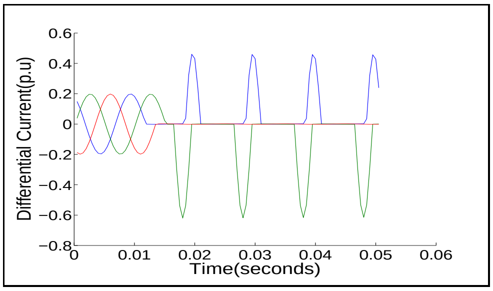
Fig. 2: Simulated power system model for three phase transformer.



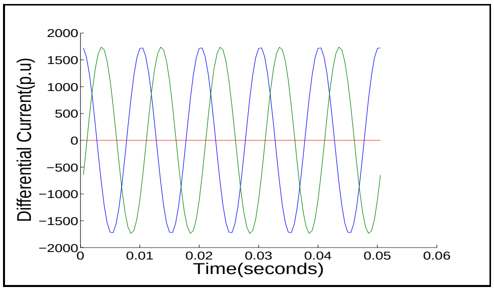
(a) Normal Condition



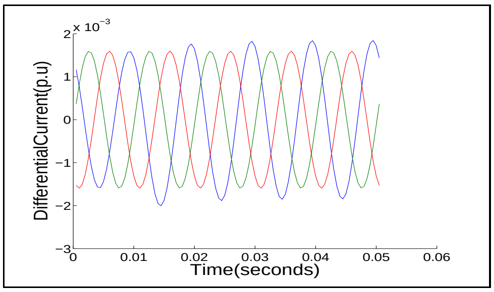
(b) Inrush



(c) Over Exictation



(d) Internal Fault



(e) External Fault

Fig. 3: Different Types of Currents Patterns for fault conditions.

3.2 Feature Selection Models

3.2.1 Gini

Gini feature importance is a random forest classifier and it handles the task of eliminating the feature explicitly. Features selected are more accurate and robust compared to others. Random forest feature selection uses two methods for feature selection namely mean decrease accuracy and mean decrease impurity.[12]

3.2.2 Relief

Relief feature selection is based on the concept of attribute weighing as the relief based algorithm gives a weight between -1 and 1 as the output for each attribute. More positive weights indicate that attribute is predictive. The weight of the attribute is updated after each iteration. This is a simple, fast and effective approach towards feature selection.[15]

3.2.3 One Rule(OneR)

As the name suggests, in this machine learning classification algorithm for each predictor in the data, one rule is generated and then the rule with the smallest error is selected as the one rule. It handles numeric data and missing values together with extensive diagnostic functions.[19]

3.2.4 Chi-Square

Chi Square technique is used in statistics to study the independence of two events. It uses the concept of frequency distribution. If a dataset is given about two events, then we can compute the observed count and expected count. Chi-Square score estimates how much the observed count and expected count deviate from each other. The events which can occur in feature selection can be occurrence of class and occurrence of features. Therefore, using Chi-Square we find whether the specific feature or the specific class occurring are dependent upon each other or not.[10]

3.2.5 Pearson's Correlation

Pearson correlation coefficient is used as a measure for quantifying linear dependence between two continuous variables X and Y. The resulting value lies between -1 and 1. Result of -1 would mean that there is a perfect negative correlation i.e. if one variable increases, the other variable decreases. If the result is +1 ,it means there is a perfect positive correlation. If the result is 0 , it means that there is no linear correlation between the two variables under consideration.[6]

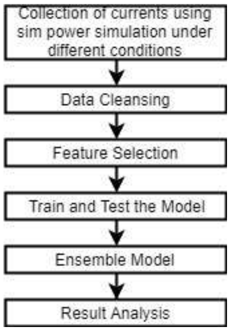


Fig. 4: Methodology used

$$Correlation = \frac{n(\sum_{x=1,y=1}^{n} x*y)-(\sum_{x=1}^{n} x)(\sum_{y=1}^{n} y)}{\sqrt{[n(\sum_{x=1}^{n} x^2)-(\sum_{x=1}^{n} x)^2][n(\sum_{y=1}^{n} y^2)-(\sum_{y=1}^{n} y)^2]}} \tag{1}$$

3.2.6 Consistency

We aim to find optimal subset of features and goodness of features is evaluated using consistency measure. A measure of consistency is intuitively defined as a metric which can calculate the distance of subset of features from the consistent state. When a feature subset is found to be consistent, the inconsistency value is 0 and as it moves towards a more consistent state, the rate of approaching to 0 decreases.[18]

3.2.7 Bias-Corrected Bayesian Classification with Selected Features(BCBCSF)

BCBCSF method uses the incomplete information obtained from the feature selection method, along with some retained features. On obtaining this information of features, it forms a correct unbiased or posterior distribution of some hyper parameters in the Bayesian model which is of hierarchal nature and it regulates the signal-to-noise ratio of the dataset on which the model is applied. BCBCSF gives a better prediction than the two most commonly used high dimension classification methods namely, diagonal linear discriminant analysis and prediction analysis for microarrays.[13]

3.3 Machine Learning Models

3.3.1 Association Rule Classification(ARC)

ARC method applies the Classification based on Association Rules(CBA). There are various convenience methods available in this package that allow to automatically set some CBA parameters like minimum support and minimum confidence etc. This method also manages attributes of numeric type by incorporating a pre-discretization step. There is an 'arules' package which can handle the association rules generation phase. Post processing can be performed by the 'qCBA' package as it reduces the size of the CBA models which are produced by the 'arc' package.[20]

3.3.2 Decision Tree

Decision Trees are a type of supervised machine learning where the data is continuously split according to a certain parameter. In supervised machine learning algorithm the input data is explained and the corresponding data is the trained data. Decision tree comprises of two objects, i.e. leaves and decision nodes. The final outcomes are the leaves of the

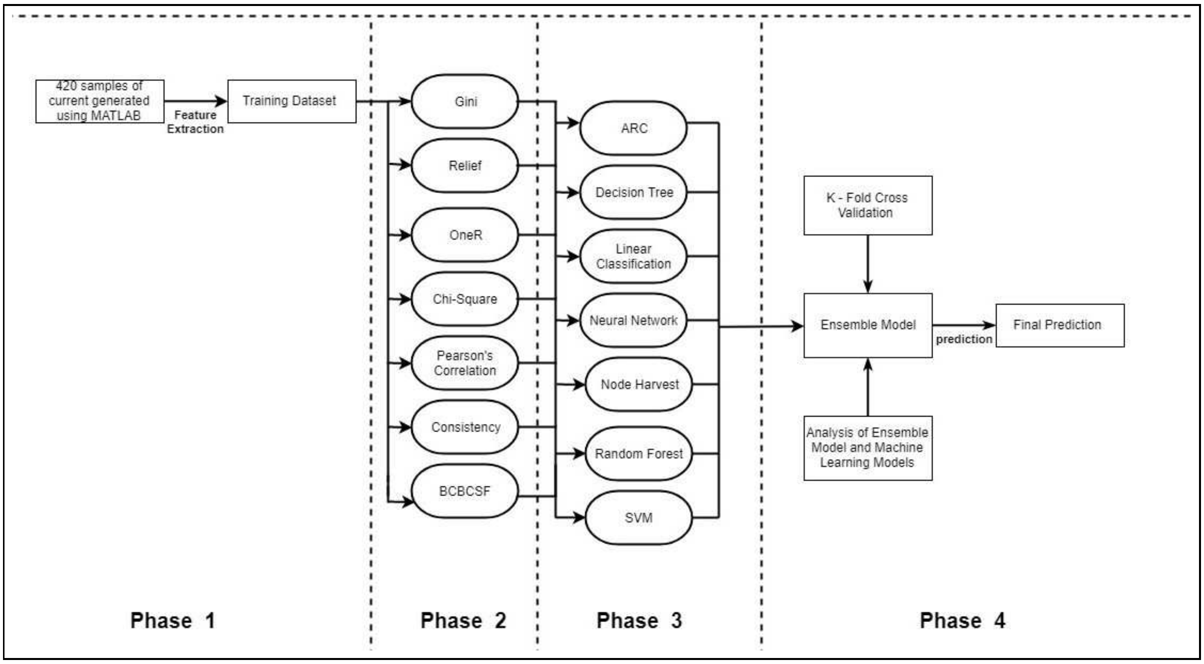


Fig. 5: Working of proposed model

decision tree and the decision nodes are where the data is branched. The most widely used algorithm for constructing a decision tree is Iterative Dichotomiser 3(ID3) algorithm. This algorithm measures the entropy i.e. the measure of randomness in the data and information gain which gives us the effective change is entropy after selecting a particular attribute.[7]

$$H = \sum_{x=1,y=1}^{n} p(x) * log_2(1/p(x)) \tag{2}$$

$$IG(S, A) = H(S) - H(S, A) \tag{3}$$

3.3.3 Linear Classification

Classification basically means using an object characteristics to identify which category/class the object belongs to. Linear classification algorithm also known as the classifier makes its classification based on a linear predictor function combining a set of weights with the feature vector. A linear classification model is mainly used when the speed of classification plays a major role because this model is quite fast and responsive. Also it is quite widely used when the number if input feature dimensions are quite large. It follows two approaches- Discriminant approach and probabilistic approach.[4]

3.3.4 Neural Network

A Neural Network is an information processing model. It works on the principle of biological nervous systems, such as the brain, and how they process information. A large number of highly interconnected processing elements or neurones make up the neural network system. Neural Networks use a through learning process and works with applications like pattern recognition and data classification. Neural networks are widely used to obtain patterns, derive meaning from complicated data and identify trends which are complex for humans to understand.[17]

3.3.5 Node Harvest

Node harvest is a self-explainable tree-like predictive model for regression and classification problems of high dimension. From a large group of nodes a few nodes are selected, each related with a weight of positive value. New observations can belong to one or several nodes and predictions made by the model are the weighted average values amongst all the groups. Predicted values can be a new observation along with the average values of training observations in each node.[2]

3.3.6 Random Forest

Random forest is basically an extension of decision trees. Multiple decision trees are merged together to build a random forest. Doing so allows random forest algorithm to give more accurate and efficient prediction compared to decision trees and is used both for classification and regression. Random forest uses the concept of searching the best feature among a random subset of features rather than searching for best features while splitting a node. This approach makes random forest more diverse and better than decision trees. Random forest is known for its good accuracy, robustness and ease of use.[11]

3.3.7 Support Vector Machine(SVM)

A support vector machine comes under the category of supervised machine learning algorithm and is

TABLE 4: Features Selected

Feature Selection Models	Number of Features Selected
Gini	7
Relief	10
OneR	15
Chi-Square	1
Pearson's Correlation	5
Consistency	13
BCBCSF	3

mainly used for classification purposes. In SVMs our focus is to find a hyperplane that divides our datasets in to classes, each class consisting of support vectors. Support vectors are the data points which lie nearest to the hyperplane. They are a crucial part of the data set because if they are removed then the position of hyperplane could be altered. The main purpose is to map the data into higher and higher dimensions until a hyperplane is generated to separate it. Since SVM uses a subset of training points, it is more efficient and accurate. It works well only for smaller data sets because larger data sets have high training time with SVMs[8]

3.4 Ensemble Model

As we have seen above, after applying the feature selection algorithms, we get different set of features for each method. Each of the seven machine learning models discussed above are trained and tested on each of these subsets of features 10 times and the accuracy in each attempt is calculated. The average accuracies calculated of each possible combination are given in Table 5. The highest accuracy of 82.667% is achieved by the combination of Gini and ARC models. Now we apply the ensemble model in which five best viable combinations are selected. In ensemble model, voting technique is used I.e. predicted output is chosen which matches the actual value the most number of times. In case of a tie, that model is selected which had a higher accuracy. The ensemble model is run 10 times and model evaluation parameters are calculated as discussed in the next section. The working of the proposed model is shown in figure 5.

4 MODEL EVALUATION

We have calculated three evaluation parameters to predict the performance of our ensemble model. A confusion matrix is obtained from the results which shows the accuracy, average class error and overall error. The confusion matrix shows the information about actual and predicted values which are determined using the ensemble model. The diagonal elements of matrix represent the number of objects for which the predicted label is equal to the true label, while non-diagonal elements are those that are mislabelled by the ensemble model. To get a better accuracy, the value of diagonal elements should be higher. If there are n number of classes then the value Cij of the confusion matrix of size nn represents the number of patterns of class i predicted in class j. Accuracy from the confusion matrix can be calculated as follows:

$$Accuracy = \frac{\sum_{i=1}^{n} C_{ii}}{\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}} \tag{4}$$

The confusion matrix will also help us to find two types of errors, average class error and overall error. The average class error of a confusion matrix will be the mean of errors calculated for each class whereas the overall error of a confusion matrix would find the total error which will occur in the prediction after running the ensemble model. Average class error from the confusion matrix can be calculated as follows:

$$AverageClassError = \frac{\sum_{i=1}^{n} \frac{\sum_{j=1}^{n}(C_{ij}-C_{ii})}{\sum_{j=1}^{n} C_{ij}}}{n} \tag{5}$$

Overall error can be calculated from the confusion matrix as follows:

$$OverallError = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(C_{ij}-C_{ii})}{\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}} \tag{6}$$

K-fold cross validation is performed to measure the robustness and consistency of the ensemble model. The training dataset(70%) is divided into k equal partitions or subsets where each partition is called a fold. From these k number of folds, one of them is used for validation or testing the model and the rest k-1 folds are used for training the model. Accuracies are estimated by averaging the accuracy calculated in all k cases of cross validation.

5 DISCUSSION AND RESULTS

After applying the feature selection algorithms and running the machine learning models on each of the selected features, we chose the five best possible combinations to run the ensemble model 10 times. The combinations chosen for the ensemble model are shown in Table 6. Combination of Gini and ARC is chosen because it has the highest accuracy of 82.667% amongst all possible cases. In each attempt, a confusion matrix is created which tells the accuracy, average class error and the overall error in each matrix. The confusion matrix with the best accuracy is shown in Table 7. The confusion matrix shows that Fault 1 is predicted correctly 27 times and similarly for all the fault conditions. The three model evaluation parameters and their values calculated after each attempt of running the ensemble model are shown in Table 8. After executing the model 10 times and averaging the results, we find out that average accuracy of predicting the fault is 85.652%, the overall error is 14.349% and the average class error is 14.175%. K-fold cross validation is done which shows that our predictive model is robust and consistent. The value

TABLE 5: Performance Comparison of Machine Learning Models for Average Accuracies

	ARC	Decision Tree	Linear Model	Neural Network	Node Harvest	Random Forest	SVM	Proposed Model
Gini	82.667	76.299	47.165	35.432	44.802	82.282	33.699	82.667
Relief	76.641	77.428	32.109	26.333	40.069	69.117	36.570	77.428
OneR	79.263	80.926	63.169	28.870	43.832	79.964	25.572	80.926
Chi-Square	66.317	67.629	19.860	42.781	N/A	50.569	19.248	67.629
Pearson's Correlation	56.17	61.244	33.419	34.909	18.024	49.432	39.982	61.244
Consistency	81.189	81.365	59.668	46.431	43.220	81.627	45.669	81.627
BCBCSF	81.451	81.015	55.030	37.183	37.095	81.540	48.906	81.451

TABLE 6: Combinations selected for Ensemble Model

Feature Selection Model	Machine Learning Model	Accuracy
Gini	Random Forest	82.282
Gini	ARC	82.677
Consistency	Decision Tree	81.365
Consistency	Random Forest	81.627
BCBCSF	ARC	81.451

TABLE 7: Confusion Matrix with Best Accuracy

Actual/Predicted	Fault 1	Fault 2	Fault 3	Fault3	Fault 5
Fault 1	27	0	0	0	0
Fault 2	0	20	4	1	0
Fault 3	0	1	19	2	0
Fault 4	0	3	1	24	0
Fault 5	0	0	0	0	25

of average accuracy doesnt deviate much and lies between 80-90% in all the attempts. The slope of the graph doesnt change much at each point. Hence our predictive model is robust and consistent. Results of the k-fold cross validation for accuracy are shown in Figure 6.

6 Conclusion

We have explored various machine learning models with various current samples as the dataset to predict the faults in three phase transformers and differentiate it with other fault conditions such as over-excitation and magnetizing inrush. Seven feature selection techniques and seven machine learning models are then applied on the data set to predict the faults. Ensemble model is then used where five best possible combinations are chosen. The result indicates that ensemble model is robust with accuracy varying between 80-90%. The work can be extended for more samples, full cycle data window and other computational methods to enhance the performance of machine learning methods.

TABLE 8: Model Evaluation Parameter

	Accuracy	Overall Error	Average Class Error
Attempt 1	85.826	14.173	12.527
Attempt 2	83.464	16.535	15.003
Attempt 3	82.677	17.322	17.612
Attempt 4	83.464	16.535	17.841
Attempt 5	90.551	9.448	9.584
Attempt 6	88.188	11.811	12.862
Attempt 7	81.102	18.897	17.542
Attempt 8	88.976	11.023	11.772
Attempt 9	86.614	13.385	12.830
Average Accuracy	85.652	14.348	14.175

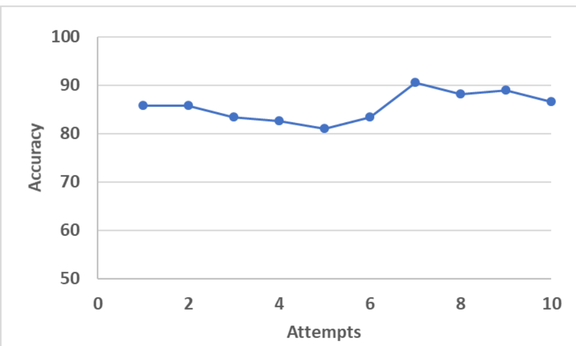


Fig. 6: K-Fold Cross Validation

References

[1] Harish Balaga, DN Vishwakarma, and Amrita Sinha. Numerical differential protection of power transformer using ann as a pattern classifier. In *IEEE Int. Conf. on Power, Control and Embedded Systems*, pages 1–6, 2010.

[2] Antonio Criminisi, Jamie Shotton, Ender Konukoglu, et al. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7(2–3):81–227, 2012.

[3] Hamideh Dashti and Majid Sanaye-Pasand. Power transformer protection using a multiregion adaptive differential relay. *IEEE Transactions on Power Delivery*, 29(2):777–785, 2014.

[4] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *Journal of machine learning research*, 9(Aug):1871–1874, 2008.

[5] M Geethanjali, S Mary Raja Slochanal, and R Bhavani. PSO trained ANN-based differential protection scheme for power transformers. *Neurocomputing*, 71(4):904–918, 2008.

[6] Mark A Hall. Correlation-based feature selection of discrete and numeric class machine learning. 2000.

[7] Chen Jin, Luo De-Lin, and Mu Fen-Xiang. An improved id3 decision tree algorithm. In *Computer Science & Education, 2009. ICCSE'09. 4th International Conference on*, pages 127–130. IEEE, 2009.

[8] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.

[9] H Kazemi Kargar, M Jabbari, and S Golmohammad-zadeh. Inrush current identification based on wavelet transform and correlation factors. In *IEEE Int. Conf. on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, volume 1, pages 50–53, 2009.

[10] Henry Oliver Lancaster and Eugene Seneta. *Chi-square distribution*. Wiley Online Library, 1969.

[11] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[12] Bjoern H Menze, B Michael Kelm, Ralf Masuch, Uwe Himmelreich, Peter Bachert, Wolfgang Petrich, and Fred A Hamprecht. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10(1):213, 2009.

[13] Jong Hee Park, Andrew D Martin, and Kevin M Quinn. Cran task view: Bayesian inference. 2005.

[14] M Rasoulpoor and M Banejad. A correlation based method for discrimination between inrush and short circuit currents in differential protection of power transformer using discrete wavelet transform. *Int. Journal of Electrical Power & Energy Systems*, 51:168–177, 2013.

[15] S Francisca Rosario and K Thangadurai. Relief: Feature selection approach. *International Journal of Innovative Research and Development*, 4(11), 2015.

[16] A Saleh and MA Rahman. Modeling and protection of a three-phase power transformer using wavelet packet transform. *IEEE Transactions on Power Delivery*, 20(2):1273–1282, 2005.

[17] Vidushi Sharma, Sachin Rai, and Anurag Dev. A comprehensive study of artificial neural networks. *International Journal of Advanced research in computer science and software engineering*, 2(10), 2012.

[18] Kilho Shin, Danny Fernandes, and Seiya Miyazaki. Consistency measures for feature selection: A formal definition, relative sensitivity comparison, and a fast algorithm. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 1491, 2011.

[19] Haleh Vafaie and Ibrahim F Imam. Feature selection methods: genetic algorithms vs. greedy-like search. In *Proceedings of the International Conference on Fuzzy and Intelligent Control Systems*, volume 51, page 28, 1994.

[20] Chengqi Zhang and Shichao Zhang. *Association rule mining: models and algorithms*. Springer-Verlag, 2002.