A Novel Approach for Multi Class Fault Diagnosis in Induction Machine Based on Statistical Time Features and Random Forest Classifier

Deepak M. Sonje
Electrical Engineering Department
RHSCOEMS& R, Nasik, India
1deepaksonele23@gmail.com

P.Kundu2, A.Chowdhury3
Electrical Engineering Department, SVNIT, Surat, India
2pk@eed.svnit.ac.in, 3ac@eed.svnit.ac.in

Abstract: Fault diagnosis and detection is the important area in health monitoring of electrical machines. This paper proposes the recently developed machine learning classifier for multi class fault diagnosis in induction machine. The classification is based on random forest (RF) algorithm. Initially, stator currents are acquired from the induction machine under various conditions. After preprocessing the currents, fourteen statistical time features are estimated for each phase of the current. These parameters are considered as inputs to the classifier. The main scope of the paper is to evaluate effectiveness of RF classifier for individual and mixed fault diagnosis in induction machine. The stator, rotor and mixed faults (stator and rotor faults) are classified using the proposed classifier. The obtained performance measures are compared with the multilayer perceptron neural network (MLPNN) classifier. The results show the much better performance measures and more accurate than MLPNN classifier. For demonstration of planned fault diagnosis algorithm, experimentally obtained results are considered to build the classifier more practical.

Keywords: Fault Detection, Multilayer Perceptron Neural Network (MLPNN), Random Forest Classifier (RF), Statistical Time Features.

1. INTRODUCTION
Induction machines are playing most important part of the process by providing the uninterrupted continuation and production in many industries. They are mainly subjected to mechanical, electrical, and thermal stresses during running condition. If any of these stresses become severe enough then various faults may initiate in the induction machine. The faults in the machine can be segregated into mainly stator, rotor, bearing and eccentricity related faults. If the faults are not sensed at an initial stage, results in premature damage of the machine and costly downtime of the plant. Numerous methods and scheme are designed for fault identification.

Many existing methods are applicable for bigger size induction machines, and very few are applicable in small size machines due to restrictions related to sensor size and cost of data.
acquisition. The major techniques for monitoring the fault are based on vibration, current, temperature, air gap torque, magnetic flux, and partial discharge measurement [1-4]. The methods which are non-invasive and non-intrusive are mostly considered for fault diagnosis of induction machines which monitors the motor’s condition using only electrical parameters. Motor current signature analysis (MCSA) is traditional non-invasive technique which utilizes the spectral analysis of stator current for fault analysis [5]. But the limitation of MCSA is that the magnitudes of characteristic frequency components are relying on the load variation which becomes difficult for fault analysis in the motor. The sensitivity of MCSA is enhanced by combining conventional MCSA method with wavelet transform, short time Fourier transform and expert system [6].

The condition based monitoring and fault analysis of induction machines have stimulated in recent years from conventional methods to artificial intelligence (AI) methods. These methods don’t require knowledge of machine parameters or any modeling of the system required. It should be noted that plenty of neural based methods are available in literature [7]–[10]. However, in the neural network (NN) based methods, architecture of a NN are not known in advance; and are obtained after a trial-and-error procedure. Some of AI methods use expert systems [11] and support vector machine (SVM) [12]. These classifiers for fault analysis are affected when the class is more in number. It is also found that the response of the classifiers becomes slow. Many detection schemes are very costly and also applicable to large size machines.

The main intention of the paper is to build up a novel method to detect and classify mixed and individual fault in induction machine. In this scheme three phase stator currents are captured, preprocessed and simple statistical parameters are estimated which are considered as the inputs to the classifier to detect and classify four conditions of motor. In classifier based fault-detection schemes, along with accuracy other performance measures are also important parameter to judge classifier performance. Accordingly, accuracy and other performance parameters such as mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE), true positive rate (TP), false positive rate (FP), precision, recall and F-measure are evaluated. For demonstration of planned fault analysis techniques, experimental results are generated to build the classifier more realistic. The rest of the paper is arranged as follows. The section 2 covers the system description and data acquisition. Section three discusses about the statistical feature extraction. The information of proposed classifier is included in section 4. The results and discussions are provided in section 5. Paper is concluded in section 6 with brief remarks and inferences.

2. DATA ACQUISITION AND PREPROCESSING

In this paper three incipient faults which are normally occur in induction motor are investigated. The experimentation and data acquisition is performed on custom designed machine whose details are furnished in Table I. The machine is coupled with mechanical arrangement for loading i.e. adjusting weight on spring and belt. The schematic diagram of the scheme is given in Fig. 1 and actual experimental set-up in shown in Fig. 2. Three phase autotransformer is used to provide rated three phase balanced power supply to the motor under various conditions. The currents flowing through the lines are sensed by the three current transformers. The diagnostic instrumentation system used is National Instruments data acquisition card model NI-6212. The data is sampled at frequency of 16.896 kHz. The outputs of current transformers are fed to the data acquisition board. The instrumentation system is
supported with Lab VIEW 2015 and MATLAB R2014a is used for the processing and analysis. In the stator current, harmonics may present due to the load and supply conditions corresponding to normal running operation of motor. In order to eliminate those harmonic components that do not provide useful failure information, selective low pass filter is designed. The statistical tool box of MATLAB is used to estimate the statistical time features.

2.1 Stator Faults

To simulate inter turn fault, the star connected stator winding has been customized by addition of a number of tappings to the coils. The winding of the machine is having 180 turns per path per phase and has two parallel paths. The tappings are brought outside and connected to an external panel from which turns for creating inter turn faults can be accessed. Minor inter turn fault involving 1.39% (5 turns) of total turns in each phase of the winding is possible with the setup. Machine is tested under different conditions of minor inter-turn faults are furnished in Table II.

2.2 Rotor Faults

For demonstrating the rotor fault, three identical rotors having 28 bars are considered. Out of these rotors, one is considered as healthy and in the other two rotors damaged artificially by drilling the holes on the rotor bars. Three bar broken and five bar broken are considered for the rotor fault analysis. The actual pictures of faulty as well as healthy rotor are given in Fig. 3.

2.3 Mixed Faults

For the analysis of mixed faults, rotor and stator faults are combined together. The various mixed combinations are considered for the analysis.

The motor is run at no load and various loading conditions with the healthy faulty cases. The four classes have been considered as healthy (H), stator fault (S), rotor fault (R) and mixed fault (M) for classification problem. The recorded observations have been arranged systematically for the classification purpose. The data size for each class is 200 x 10000 samples chosen.

Table I. Specifications of three phase induction motor

| Rated Power (P) | 2.2 kW | Connection (Star/Delta) | Star |
|----------------|--------|-------------------------|------|
| Rated Speed (N) | 1440 rpm | pole pairs (p) | 2 |
| Rated Voltage (V) | 400 V | Total stator slots | 32 |
| Rated Current (I) | 4.5 Amp | Total rotor bars | 28 |
| Frequency (f) | 50 Hz | | |

Table II. Stator inter-turn fault cases

| Sr. No. | Inter-turn fault type | No. of turns shorted | % of stator winding turns per phase shorted |
|---------|-----------------------|---------------------|------------------------------------------|
| 1       | Healthy               | 0                   | 0                                        |
| 2       | Fault1                | 5                   | 1.39                                     |
| 3       | Fault2                | 10                  | 2.77                                     |
| 4       | Fault3                | 15                  | 4.16                                     |
| 5       | Fault4                | 20                  | 5.55                                     |
| 6       | Fault5                | 25                  | 6.94                                     |
3. Feature Extraction

In conventional fault diagnostic methods, electrical parameters such as voltage, currents and mechanical parameter such as vibration data are considered as features. During experimentation, no noticeable changes are observed in current waveforms of stator under healthy and various faulty condition as given in Fig. 4. There is necessity to extract some unique features from these currents. Accordingly, set of common and simple statistical-time features are estimated from the three phase currents. Fourteen features are suggested to each condition which is estimated from (1)-(14). The least number of time features to be analyzed covers the
maximum, minimum, mean, median values, sum and standard deviation. The mean and median are estimated on individual dimension basis. The mean or variance, are utilized to describe the probability density function of a time-varying signal. The symmetry of distribution is measured by skewness. If the distribution of data is close to symmetrical in the region of its mean, the value of skewness is near to zero. If the distributed data has a longer tail to the right of the mean represents positive value of skewness and vice versa. The higher order statistical time features such as kurtosis, which gives warning of the proportion of samples which diverge from the mean by a minute value compared with those which diverge by a big value. These features have the important property that they are not sensitive to Gaussian distributed noise.

Minimum Value \[ X_{\text{min}} = \min(x_i) \]  
Maximum Value \[ X_{\text{max}} = \max(x_i) \]  
Mean \[ \mu = \frac{1}{N} \sum_{i=1}^{N} x_i \] 
N= the total number of samples
Median \[ \text{median} = \left( \frac{(N+1)}{2} \right)^{\text{th}} \text{ value} \] 
Standard Deviation \[ S \left( \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N - 1} \right) \] 
Variance \[ S^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N - 1} \] 
Sum \[ \text{Sum} = \sum_{i=1}^{N} x_i \] 
Skewness \[ \text{Skewness} = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \mu)^3}{\sigma^3} \]

Fig. 4. Current waveforms acquired at constant load condition: Healthy, stator fault, rotor fault and mixed fault.
4. Classifier

4.1 Random Forest Classifier [13]

RF classifier is based on the grouping of trees for regression and classification. It is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The distinctive training set is completed by using bagging. The procedure bagging means extracts a fixed quantity from a training set randomly improve classification, and regression models according to stability and classification accuracy. This process decreases variance and avoids over-fitting. The absolute results makes by adding the scores of component predictors on every class and then choosing the successor class in terms of the number of scores to it. The error for forests converges to an optimized value as the number of trees in the forest becomes large. The error of classifiers also depends on the strength of the individual trees and the correlation between them.

4.2 Multi Layer Perceptron Neural Network Classifier (MLPNN) [13]

Artificial neural networks model is motivated from human brain and biological learning process and they have been developed in the form of parallel distributed network. MLPNN classifier is a feed forward model that maps groups of input data into a set of fitting outputs. During the training neural networks are fitted to the data by learning algorithms. In this paper, supervised learning has been used with input of forty two features and four outputs. It consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron with a nonlinear activation function. It utilizes a back propagation for training the network. Two Sigmoid activation are used in the classifier which are described by (15),

\[ y(v_j) = \tanh(v_j), \quad y(v_j) = \left(1 + e^{-v_j}\right)^{-1} \]  (15)
first is a hyperbolic tangent which ranges from -1 to 1, and the other one is logistic function but ranges from 0 to 1. Here $y_i$ is the output of the $i^{th}$ node and $v_i$ is the weighted sum of the input synapses. The error is minimized by using back propagation algorithm. We represent the error in output node $j$ in the $n^{th}$ data point by $e_j(n) = y_j(n) - d_j(n)$, where $d$ is the target value and $y$ is the value produced by the perceptron. The error is optimized in the entire output given by (16),

$$
e(n) = \frac{1}{2} \sum e_j^2(n)$$

(16)

5. RESULT AND DISCUSSION

As mentioned in earlier, fourteen statistical features are calculated for each phase of filtered current. These features are given as an input to random forest classifier. For testing and training the algorithm, Weka software (Ver. 3.8.0) is used as a tool [14]. The use of this data mining and machine learning software is very common in both academic researches and industrial studies [13-14]. The designed parameters used for algorithm are the same as default values determined by the software. The algorithm is tested on data sets of 200, 400, 600 and 800 numbers of observations to check the performance classifier. It is observed that as the data size is increased, classifier performance gradually improves. Finally, data sets of 800 observations are considered for training and testing.

After selecting the data set, initially small data is presented for training and large data is kept for testing. Gradually, training data size is increased and classifier performance is observed. In the next stage, the classifier is tested on cross validated (CV) data of 10 fold. The results of the classifier are compared with the known MLPNN classifier as given in Table III-V. It is observed from Table III that RF classifier performance is independent on testing and training ratio and performance measures are consistently optimum for all the combinations in comparison with overall accuracy and various errors. The various errors are also reduced to significantly as given in Table III. 100 % accuracy is obtained for all the combinations of data sets. The class wise performance measures such as TP, FP, precision, recall and F-measure are listed in Table IV. TP rate, precision, recall and F-measure also achieved desired value as 1 and FP rate is zero. The confusion matrix gives the information about the false and correct observations represented in Table V. Out of 800 observations false prediction rate is reduced to zero in case of RF classifier and thirty five in case of MLPNN classifier given in Table V. The results obtained from the proposed classifier gives more accurate results as compared to MLPNN classifier. Hence, the proposed classifier has been found more suitable for the fault detection in induction machine.
Table III. Comparison of RF Classifier and MLPNN classifier

| Parameter                  | RF Classifier | MLPNN classifier |
|----------------------------|---------------|------------------|
|                            | C.V. | Ratio 0.5 | Ratio 0.6 | Ratio 0.7 | Ratio 0.8 | Ratio 0.9 | C.V. | Ratio 0.5 | Ratio 0.6 | Ratio 0.7 | Ratio 0.8 | Ratio 0.9 |
| Time taken to build        | 0.31 | 0.26      | 0.28      | 0.27      | 0.27      | 0.27      | 18.7 | 18.8      | 18.8      | 18.8      | 18.8      | 18.8      |
| % Overall Accuracy         | 100  | 100      | 100      | 100      | 100      | 95.6      | 97.8 | 98.1      | 96.3      | 98.8      | 98.8      |
| Mean Absolute Error        | 0    | 0.01     | 0        | 0        | 0        | 0.03      | 0.02 | 0.02      | 0.02      | 0.01      | 0.02      |
| Root Mean Squared Error    | 0.01 | 0.03     | 0.01     | 0.01     | 0.01     | 0.12      | 0.11 | 0.07      | 0.09      | 0.05      | 0.06      |
| % Relative Absolute Error  | 0.43 | 2.44     | 1.13     | 0.63     | 0.36     | 0.36      | 6.88 | 6.34      | 5        | 5.58      | 3.43      | 4.21      |
| % Root relative squared Error | 1.55 | 6.71     | 3.2      | 2.28     | 1.33     | 1.44      | 28.4 | 24.8      | 15.8      | 20.7      | 11.5      | 13.9      |

Table IV. Detailed Accuracy by class

| Parameter   | RF Classifier | MLPNN classifier |
|-------------|---------------|------------------|
|             | H  | S  | R  | M  | H  | S  | R  | M  |
| TP Rate     | 1  | 1  | 1  | 1  | 1  | 0.93 | 0.98 |
| FP Rate     | 0  | 0  | 0  | 0  | 0.02 | 0    | 0.07 |
| Precision   | 1  | 1  | 1  | 1  | 0.93 | 1    | 0.98 |
| Recall      | 1  | 1  | 1  | 1  | 1    | 0.93 | 0.98 |
| F-Measure   | 1  | 1  | 1  | 1  | 0.97 | 1    | 0.95 |

Table V. Confusion matrix

| Classified Output | RF Classifier | MLPNN classifier |
|-------------------|---------------|------------------|
|                   | H  | S  | R  | M  | H  | S  | R  | M  |
| H                 | 200 | 0  | 0  | 0  | 200 | 0  | 0  | 0  |
| S                 | 0   | 200 | 0  | 0  | 0   | 200 | 0  | 0  |
| R                 | 0   | 0  | 200 | 0  | 21  | 6   | 173 | 0  |
| M                 | 0   | 0  | 0   | 200 | 1   | 1   | 6   | 192 |

6. Conclusion

This paper has presented novel and accurate approach for individual and mixed fault in induction machine. Fourteen statistical parameters for each phase current have been calculated for various faulty conditions and feed input to the RF and MLPNN classifier. From obtained results, it has been observed that RF classifier performance is better than MLPNN classifier. The accuracy is reach to 100% and various errors are drastically reduced to almost 0% using proposed classifier. The false prediction of the classifier is also reach to zero in the 800 observations of data set. The other performance measures such as TP rate, FP rate, Precision, recall and F-measure also improved significantly. The classifier accuracy is almost constant for cross validated data and for various testing and training ratio of data sets. The RF classifier found to be superior in accuracy in comparison with MLPNN algorithms. The results obtained using proposed classifier method is promising therefore; proposed method can be better choice for individual and mixed fault analysis and classification in induction machine.
ACKNOWLEDGMENT

The authors thankfully acknowledge the support provided by S.V.N.I.T., Surat for conducting the research work. Deepak M. Sonje also thanks KKWIIER, Nashik and RHSCOEMS&R, Nashik for their help in assisting while conducting the experiments.

REFERENCES

[1] P. Zhang, Y. Du, T. G. Habetler, B. Lu, “A survey of condition monitoring and protection methods for medium-voltage induction motors,” IEEE Trans. on Industry Applications, vol.47, no.1, pp. 34-46, 2011.

[2] A. Siddique, G. S. Yadava, B. Singh., “A review of stator fault monitoring techniques of induction motors ”, IEEE Trans. on Energy Conversion, vol. 20, pp. 106-114, 2005.

[3] Deepak M. Sonje, A. Chowdhury, P. Kundu, “Park’s Vector Approach for Fault Diagnosis of Induction Motor,” Proc. of IEEE Conference on Advances in Electrical Engineering, ICAEE, pp.1-4, 2014

[4] A. Bellini, F. Filippetti, C. Tassoni, and G. A. Capolino, “Advances in diagnostic techniques for induction machines,” IEEE Trans. on Industrial Electronics, vol. 55, pp. 4109–4126, Dec. 2008.

[5] W. T. Thomson and M. Fenger, “Current signature analysis to detect induction motor faults,” IEEE Industrial Application Magazine, vol. 7, pp. 26–34, Aug. 2001.

[6] M. E. H. Benbouzid, “A review of induction motors signature analysis as a medium for faults detection,” IEEE Trans. on Industrial Electronics, vol.47, 5, pp. 984–993, 2000.

[7] V. N. Ghate, S.V. Dadul, “Cascade neural-network-based fault classifier for three-phase induction motor”, IEEE Trans. on Industrial Electronics, vol. 58, pp. 1555–1563, Dec. 2011.

[8] C. T. Kowalski and T. Orlowska-Kowalska, “Neural networks application for induction motor faults diagnosis,” Math. Comput. Simul. vol. 63, no. 3–5, pp. 435–448, Nov. 2003.

[9] M. B. K. Bouzid, G. Champenois, N. M. Bellaaj, L. Signac, and K. Jelasli, “An effective neural approach for the automatic location of stator inter turn faults in induction motor,” IEEE Trans. on Industrial Electronics, vol. 55, no. 12, pp. 4277–4289, Dec. 2008.

[10] F. Filippetti, G. Franceschini, C. Tassoni, and P. Vas, “Recent developments of induction motor drives fault diagnosis using AI techniques,” IEEE Trans. on Industrial Electronics, vol. 47, no. 5, pp. 994–1004, Oct. 2000.

[11] B. Ayhan, M. Y. Chow, and M. H. Song, “Multiple discriminant analysis and neural-network-based monolith and partition fault-detection schemes for broken rotor bar in induction motors,” IEEE Trans. on Industrial Electronics, vol. 53, no. 4, pp. 1298–1308, Jun. 2006.

[12] V. N. Ghate, S.V. Dadul, “Induction machine fault detection using support vector machine based classifier”, WSEAS Trans. on Systems, vol. 8, pp. 591-603, May 2009.

[13] Raj Kumar Patel, V.K. Giri, “Feature selection and classification of mechanical fault of an induction motor using random forest classifier”, Elsevier Engineering and Material Sciences, vol.8, 334-337, 2016.

[14] http://www.cs.waikato.ac.nz/me/weka
Deepak M. Sonje received the B. E. (Electrical) degree from Shivaji University in 1999, M.E.(Electrical) degree from SPPU, Pune, in 2009, and pursuing Ph. D degree in electrical engineering from SVNIT, Surat. Since 2010 he has been with Electrical Engineering Department, GES’s RHSCOEMS&R, Nashik. His research areas are fault diagnosis of electrical machines, machine learning tools, modeling of electrical machines and drives. Four papers have been published in proceedings of international conferences and journals. He is a member of the Institution of Engineering and Technology (IET).

Prasanta Kundu obtained the B. E. (Electrical) degree in 1992, M. E. (Electrical) degree in high voltage from IISC Bangalore, in 1994 and Ph. D degree in electrical engineering from IIT, Kharagpur, in 2009. He is currently working as an Assistant Professor with SVNIT, Surat, India. He is the recipient of 2007 IEEE DEIS graduate student fellowship. His area of research is condition monitoring, DSP techniques and applications, electromagnetic field computation. He is an associate member of the Institution of Engineers (IE).

Anandita Chowdhury received her B. E. and M. E. from university of Calcutta and Ph. D degree in electrical engineering from IIT, Kharagpur, India. Presently, she is working as an Associate Professor in the electrical engineering department, SVNIT, Surat. She has more than twenty two years of teaching experience. Her area of research includes electrical machines, drives and power system stability.