Defence Science Journal, Vol. 69, No. 3, May 2019, pp. 249-253, DOI : 10.14429/dsj.69.14413

Fault Detection and Isolation in Electrical Machines using Deep Neural Networks

M. Sai*, Parth Upadhyay, and Babji Srinivasan

Department of Electrical Engineering, Indian Institute of Technology, Gandhinagar - 382 355, India
**E-mail: reachmsai@gmail.com*

ABSTRACT

Condition and health monitoring of electrical machines during dynamic loading is a common, yet challenging problem in main battle tanks. Existing methods address this issue by extracting various features which are subsequently used in a classifier to isolate faults. However, this approach relies on the feature set being extracted and therefore most of the time does not provide expected accuracy in identification of faults. In this work, we have used convolution neural network that utilises the original time domain measurements for fault detection and isolation (FDI). Results from experimental studies indicate that the proposed approach can perform FDI with more than 95\% accuracy using commonly available current measurements.

Keywords: Electric machine; Non stationary; Faults; Convolution neural network

1. INTRODUCTION

Prognosis and diagnosis is a key aspect in maintaining the life of the vehicle, and it becomes more important when it comes to combat vehicles. Modern vehicles are laden with sensors for tracking various mechanical and electrical parameters to ensure performance and safety of the vehicle. Owing to an emphasis on electric vehicles, monitoring of electric motors has become more significant task compared to the past. Incipient faults in the electrical machines result in degradation of performance, ultimately leading to compromise of the mission in combat vehicles. Several studies have focused on condition monitoring of electrical machines. However, there are very few methods which can perform prognostics of machines under continuously varying load conditions (non-stationary operation), which is more realistic than constant loadings. This is due to the fact that the fault signatures of the electrical machines are dependent on the loading conditions and therefore does not provide clear fault signatures under non-stationary operation. In combat vehicles, depending on the mode and mission, the load on the electrical machines vary continuously. This has to be viewed with the introduction of hybrid combat vehicles wherein the power requirement from the machines vary continuously due to load sharing with batteries. Therefore, it is important to develop prognostics and diagnosis tools that can monitor the machines even under non-stationary operation of vehicles. Also, it is desirable to perform this operation with data from a minimum number of sensors, preferably using current measurements, as the addition of sensors is difficult due to space constraints in combat vehicles. Further, current is a more intrinsic parameter for the motor and commonly measured for various purposes. Our current work is aimed to address the challenges pertaining to the prognosis of electrical machines in combat vehicles.

A methodology for prognosis and diagnosis of electrical machines operating under non-stationary conditions using simple current measurements is proposed. In this work, we considered bearing, inter turn, and misalignment faults with bearing fault being the most common among all[1]. Literature in other domains reports that deep neural networks perform better than traditional methods for pattern recognition. One of the advantage with these methods is that it acts on the raw signal rather than on some transform[2]. So, this makes it independent of the transform used and also more information can be extracted from the original signal. Motivated by this, we have used convolution neural network (CNN) where feature extraction and classification is done by the same network.

2. EXPERIMENTAL METHODOLOGY

For demonstration purpose, a 2.2 hp, 10 A separately excited DC motor (coupled with a generator) with hall effect current sensors for current measurements was setup at IIT Gandhinagar as shown in Fig. 1. Current is measured using Hall effect current sensor (LA-25p), and it is logged in the pc through dspace with a sampling frequency of 20 KHz. In order to introduce non-stationary load conditions, the load on the motor is continuously varied using the coupled generator through electronically controlled switches. The algorithm for loading the motor is implemented using a software so that resistive loads can be turn on and turn off without manual intervention.

The electrical motor is custom made for introduction of the following faults.

a. *Bearing Fault*: It is the most common of all and is very difficult to diagnose in early stages. However, once diagnosed it is very easy to replace and prevent the system from the failure. Bearing faults can be a localised defect

Received : 30 August 2018, Revised : 28 February 2019
Accepted : 13 March 2019, Online published : 30 April 2019

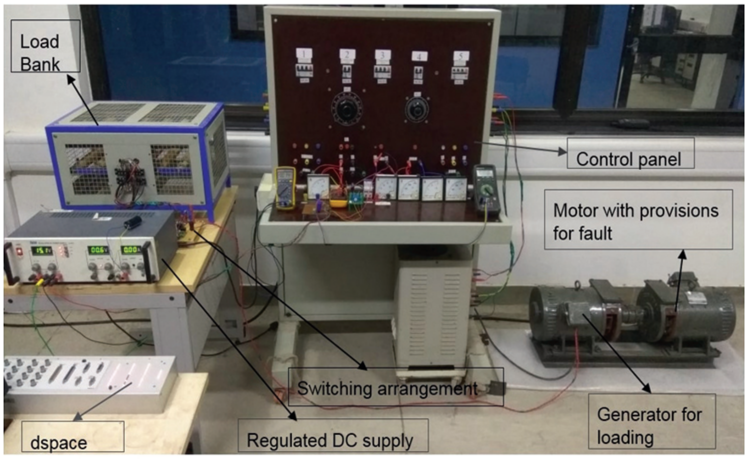

Figure 1. Experimental setup.

at certain places. This is also known as single point defect and is induced in the current setup by drilling a hole of 1.5 mm dia in the outer race-way of the bearing. Increase in generalised roughness occurs when the surface of the bearing degrades over a larger area, more common than single point defect and is relatively difficult to detect them[3]. This can be induced by passing the a.c. current through the bearing for certain time and this period will depend upon the bearing specification[5].

b. *Inter turn Fault*: It occurs when the turns of the winding get shorted. We have made the provision in the motor such that one to five turns can be shorted. For demonstration of the proposed FDI framework, we have collected data with Three turns (shorted) faults.

c. *Misalignment*: This fault arises when the centrelines of load shaft and motor shaft are not coinciding with each other. If centrelines intersect at some angle but remains in the same plane, it is called angular misalignment, and if the centrelines are parallel but offset by some distance, then it is called parallel misalignment.

In this work, we have introduced inter turn fault with three turns shorted and bearing outer race fault. We have two working condition. Thus we have three classes including the healthy state in each working condition. Figure 2 shows the plots of the current for 0.3 s under constant loading condition, and Fig. 3 shows the plot for 300 s under non-stationary condition.

The input data (current) needs to be preprocessed before applying any algorithm so that analysis can be done effortlessly. Following are the steps taken for the CNN framework.

- The obtained current data is normalised to have zero mean and unit standard deviation. It eases the process of optimisation, speeds up the learning process and also take cares of the condition when features (input) are of different scale[5].
- This data is subsequently scaled linearly to the range of [-1 1].
- The length of each sample is 2000 time stamps, and it is deliberately taken so that it includes at least one period of the signal[6].
- Total training, validating and testing samples for the stationary condition are 12000, 3000 and 1500 respectively. Samples are uniformly divided among three classes. Similarly, 14400, 3600 and 1500 are training, validating and testing samples respectively for non-stationary condition.

3. CONVOLUTION NEURAL NETWORK

The idea of convolution neural network (CNN) is inspired from the animal visual cortex. Cortex has a complex arrangement of cells which respond to the stimuli only in a restricted region of the visual field known as a receptive field. These small regions overlap to form the complete visual field[7]. The filters in the network act as cells which search the spatial correlation present in the input. CNN has relatively very less parameter compared to the fully connected network due to following two properties:

- *Sparse Connection*: The output of convolution operation is known as feature map. Each entry in feature map is only dependent on a subset of input.
- *Shared Parameters*: filters are used to extract particular features from the input data. If there is a certain feature in some part of the input, then it is also probable that it will be present in other part of the data. So one filter can work for whole input in extracting a particular feature.

Convolution neural network consists of convolution and pooling layer followed by fully connected layers at the end. The operation in each type of layer is briefly described as follow

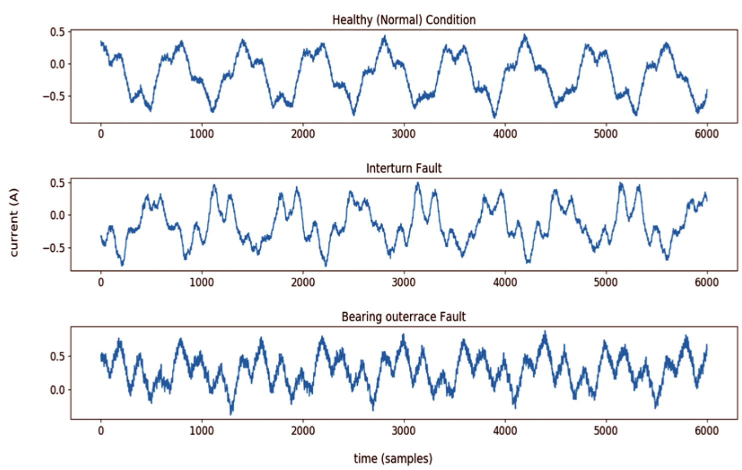

Figure 2. Current under different faults during constant load.

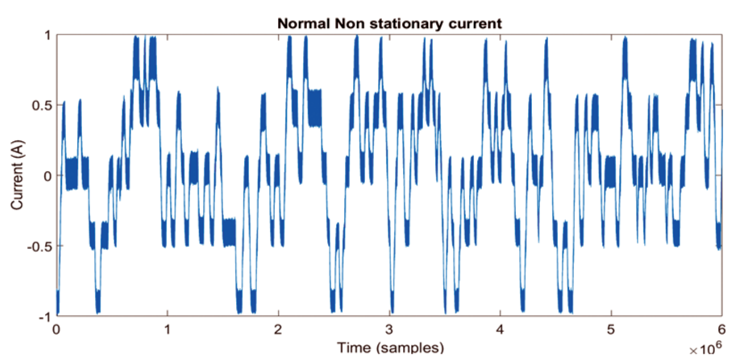

Figure 3. Current under non-stationary normal condition.

The convolution operation is generally represented by the following equation[5].

$$z(t) = I(t) * K(t) \tag{1}$$

where I is the input and k is filter or kernel, and z is the feature map. For two-dimension input and Kernel, we have the following expression

$$Z(x,y) = I(x,y) * K(x,y) = \sum_m \sum_n I(m,n)K(x-m, y-n) \tag{2}$$

In multidimensional cases, data is represented as a matrix where convolution between a patch of input and filter can be viewed as the weighted sum of the input. Then filter is moved by some defined steps known as stride, and the above mentioned operation is repeated to cover the whole input. The dimensionality of the output shrinks due to this operation and so padding is required. In padding, additional strip of elements is embedded along the boundary of the input matrix so that input and output size remains same. The dimension of the output after convolution is given as

$$((N + 2P - F) / S + 1)X((N + 2P - F) / S + 1) \tag{3}$$

where N is the height or length of the input, F is the Filter size, P is the Padding size, and S is the stride length.

Pooling is like sub-sampling operation which prevents the size of the convolution output from growing too much and at the same time preserve the features learned by the network. This operation gives some statistics of the elements under the selected region[5]. For instance, max pooling with pooling size of 3 will give the maximum value of the elements under 3X3 square region in the two-dimension input.

Each neuron in a layer performs two operations. First, it calculates a weighted sum of all inputs and then applies some nonlinear function on it. To quantify how good our model is performing for the current set of parameters, we need some function which is known as loss function. Structure of general loss function is given as

$$J(\theta) = \frac{1}{m}\sum_i^M [L(Y^i, \hat{Y}^i)] + \lambda\phi(\theta) \tag{4}$$

where θ is the parameters of the model to be learned, $\hat{Y}$ is the output of the model, M is the number of samples and Y is the corresponding target (desired) output.

The loss function consists of two terms. The first term will be some function of Y and $\hat{Y}$ and second term is regularisation or penalty term. Depending upon the problem, there are various functions like mean square error, cross entropy, negative logarithmic likelihood, etc. Categorical cross entropy is frequently used for multiclass classification problems. Minimising norm-2 of the parameters is one common practice for regularisation.

Any machine learning algorithms including deep learning ultimately results in optimisation of some loss function so that model can be learned by updating its parameters. In a neural network, due to nonlinearity the objective function is non convex and may have many local minima and saddle points[5]. Here, there is no convergence criterion like in the case of other machine learning algorithms, and it also depends upon the initial values of parameters. So in minimisation problem, we are satisfied with the values that are very low but not minimal[5]. We always optimise the function using gradient descent algorithm and there are various other methods which are just the modification of the gradient descent. In deep learning, there is no fixed rule for training the model, so we have to train a lot of models to find the best one. Learning can be speed up by selecting a good optimisation algorithm. Instead of changing weights after whole training set, we can change weights after certain training samples, known as mini-batches. The size of mini batch lies between 1 and M such that it balances the oscillation and speed of the learning. There are various other modifications like gradient descent with momentum, rmsprop, Adam, etc. The update rule with basic gradient descent is given as follows

$$dW^{[L]} = dJ / dW^{[L]}$$
$$W^L(t+1) = W^L(t) - \alpha dW^L$$
$$b^L(t+1) = b^L(t) - \alpha dB^L \tag{5}$$

where α is the learning rate determining the step size, t is the time instant, W and b are the weights and bias in the layer L.

4. RESULTS AND ANALYSIS

We have used six layer convolution neural network with the first four for extracting features and last two for classification. Number of filters in first convolution layer is 16, each having dimension of 50. Pooling is done with 1X3 sub region. In Second convolution layer, there are 8 filters, each having dimension of 50. In last two layers, there are 50 and 3 neurons respectively. ReLu function is used throughout the network except for last layer where softmax is used. Dropout fraction in the fifth layer is 0.3, i.e. randomly 30 per cent of the neurons will not be connected to next layer during training time. It is a type of regularisation which prevents overfitting[5]. For multiclass classification, categorical cross entropy is used as loss function which is minimised by rmsprop optimizer as its speed up the gradient descent. Mini-batch size of 64 is used which means parameters are updated once after 64 inputs are processed. This model is trained in the python-3.6 using Tensorflow and Keras library.

There are various measures to quantify the quality of predictions. Some of them are confusion matrix, recall score,

Table 1. Confusion matrix for stationary condition

Predicted → True ↓	Normal	Interturn	Bearing outerrace	Accuracy (per cent)
Normal	500	0	0	100
Interturn	0	500	0	100
Bearing outerrace	0	0	500	100

Table 2. Confusion matrix for non-stationary condition

Predicted → True ↓	Normal	Interturn	Bearing outerrace	Accuracy (per cent)
Normal	499	0	0	99.80
Interturn	0	500	0	100
Bearing outerrace	0	0	498	99.60

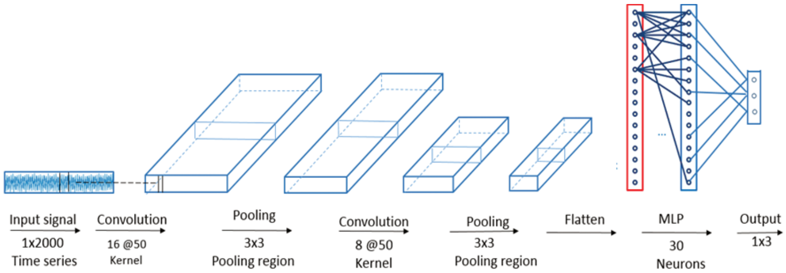

Figure 4. CNN structure.

precision score, F-score, Cohen's Kappa, Receiver operating characteristics (ROC), etc. Confusion matrix is a table that is used to describe the performance of a classification model on a set of test data for which the true labels are known. The horizontal axis gives predicted label, and vertical axis gives a true label. The precision is the ability of the classifier not to label as positive a sample that is negative (exactness) and recall is the ability of the classifier to find all the positive samples[8]. A low recall indicates many false negatives and low precision indicate a large number of false positives.

TP=True positive FP=False positive

FN=False negative TN=True negative

Classification accuracy = $(TP+TN)/(TP+TN+FP+FN)$

Precision = $TP/(TP+FP)$

Recall = $TP/(TP+FN)$

Figure 5. Confusion matrix table for binary classification.

The F1 score is the weighted average of the precision and recall. It conveys the balance between the precision and the recall. F1 score reaches its best value at one and worst score at zero[8].

F1 score = $2*(precision*recall)/(precision+recall)$

For multiclass problem, binary metric calculations are averaged across the different classes.

For stationary condition, total classification accuracy is 100 per cent, and all other performance metrics which are mentioned above are 1.0 as shown in Table 1.

In non-stationary condition, Overall Accuracy of the classifier is 99.80 per cent. Precision is 0.9980, Recall is 0.9979 and F1 score is 0.9979 as shown in Table 2.

We have also implemented one of the traditional methods for the comparison purpose. Dual tree complex wavelet packet transform is used to transform the signal from time domain to time-frequency plane[9,10]. Then Shannon entropy of selected frequency band is used as a feature along with SVM to classify the faults. The overall accuracy comes out to be 97 per cent and 89 per cent for stationary and non-stationary condition respectively.

5. CONCLUSIONS

In the present work, we attempted to provide a solution for FDI of machines under continuous loading conditions using CNN. The proposed CNN uses alternate convolution and pooling layers to extract features. Further, we have only used the current for the diagnosis which removes the requirement of installing additional sensor like accelerometer, acoustic sensor, etc. This method can outperform conventional methods as its accuracy is more than 99 per cent in performing FDI. The only constraint with this method is a requirement of a large amount of data which is not a difficult task for this industrial era where digitalisation and IoT are commonly available. One dimension convolution is used in the network which is computationally cheap compared to other complex operation, so this diagnosis can be done in real time and can be implemented in FPGA board[11]. In the future, we will introduce more faults and would like to train the model with the noisy data so that the classifier becomes more robust.

REFERENCES

1. SubCommittee, P. S. E.; MRW, Group & PSE, Committee. Report of large motor reliability survey of industrial and commercial installations: Part 3. *IEEE Trans. Industry Appl.*, 1987, **23**(1), 153–158 .
2. Ince, T.; Kiranyaz, S.; Eren L.; Askar, M. & Gabbouj, M. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Industrial Electron.,* 2016, **63**(11),7067-7075
   doi: 10.1109/TIE.2016.2582729
3. Stack, J.R.; Habetler, T.G. & Harley, R.G. Fault classification and fault signature production for rolling element bearings in electric machines. *IEEE Trans. Industry Appl.*, 2004, **40**(3), 735-739.
   doi: 10.1109/TIA.2004.827454
4. Stack, J.R.; Habetler, T.G. & Harley, R.G. Experimentally generating faults in rolling element bearings via shaft current. *IEEE Trans. Industry Appl.*, 2005, **41**(1), 25-29.
   doi: 10.1109/TIA.2004.840966
5. Goodfellow, I.; Bengio, Y.; Courville, A. & Bengio, Y. Deep learning. Cambridge, MIT press; 2016.
   doi: 10.4258/hir.2016.22.4.351
6. Liu, R.; Meng, G.; Yang, B.; Sun, C. & Chen, X. Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. *IEEE Trans. Industrial Info.*, 2017, **13**(3), 1310-20.
   doi: 10.1109/TII.2016.2645238
7. Hubel, D.H. & Wiesel, T.N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*, 1959, 148(3), 574-591.
   doi: 10.1113/jphysiol.1959.sp006308
8. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. & Vanderplas, J. Scikit-learn: Machine learning in Python. *J. Mach. Learning Res.,* 2011, 2825-30.
9. Selesnick, I.W. On the dual-tree complex wavelet packet

and m-band transforms. *IEEE Trans. Signal Proces.*, 2008, **56**(6), 2298-2310.
doi: 10.1109/TSP.2007.916129

10. Seshadrinath, J.; Singh, B. & Panigrahi, B.K. Incipient turn fault detection and condition monitoring of induction machine using analytical wavelet transform. *IEEE Trans. Industry Appl*., 2014, **50**(3), 2235-2242.
doi: 10.1109/TIA.2013.2283212
11. Farabet, C.; Poulet, C; Han, J.Y. & LeCun, Y. CNP: An FPGA-based processor for convolutional networks. *In* International Conference on Field Programmable Logic and Applications. 31 Aug. 2009. IEEE - FPL 2009. pp. 32-37.
doi: 10.1109/FPL.2009.5272559

CONTRIBUTORS

Mr M. Sai, has completed BTech (Electrical Engineering) from BIT Sindri and MTech (Electrical Engineering) from Indian Institute of Technology Gandhinagar.
In the current study, he developed machine learning models and handle other data mining work.

Mr Parth Upadhyay, received his BTech from L.D. College of Engineering, and MTech from PDPU, Gujarat. He is currently pursuing PhD in Department of Electrical Engineering, Indian Institute of Technology, Gandhinagar.
In the current study, he helped in testbed setup for the collection of fault data.

Dr Babji Srinivasan received his BTech (Instrumentation and Control Engineering) from the Madras Institute of Technology, India, in 2005, and PhD (Chemical Engineering) from Texas Tech University, in 2011. Presently working as Assistant Professor position at Indian Institute of Technology Gandhinagar. His research interests include: design, monitoring and control of cyber physical systems.
The current study is conceptualised and executed by him.