Application of Machine Learning Techniques for fault classification in Induction Machines

Open Elective Presentation

Soham Chaudhuri, Sneha Ray, Uttam Kanti Dutta, Srayan Bhattacharyya, Mainak Seal, Rohan Pandey, Sugam Das, Priyansu Karmakar, Prabir Paul

Department of Electrical Engineering, Jadavpur University

December 8, 2024

Objective and Methodology

- **Develop a Machine Learning Classifier:**
  - Create a robust ML model that can classify motor conditions as healthy or with pre-defined fault types. This allows us to overcome the drawbacks of conventional hardware loop based monitoring.
- **Enhance Fault Detection Capabilities**:
  - Implement a solution that accurately detects the considered pre-defined faults regardless of varying load conditions.
  - Provide a non-invasive diagnostic approach that improves on conventional methods' capabilities.
- **Support Predictive Maintenance**:
  - Enable early fault detection to minimize machine downtime and maintenance expenses.
  - Help plan maintenance schedules more effectively to prevent costly breakdowns.

Types of Faults in an IM

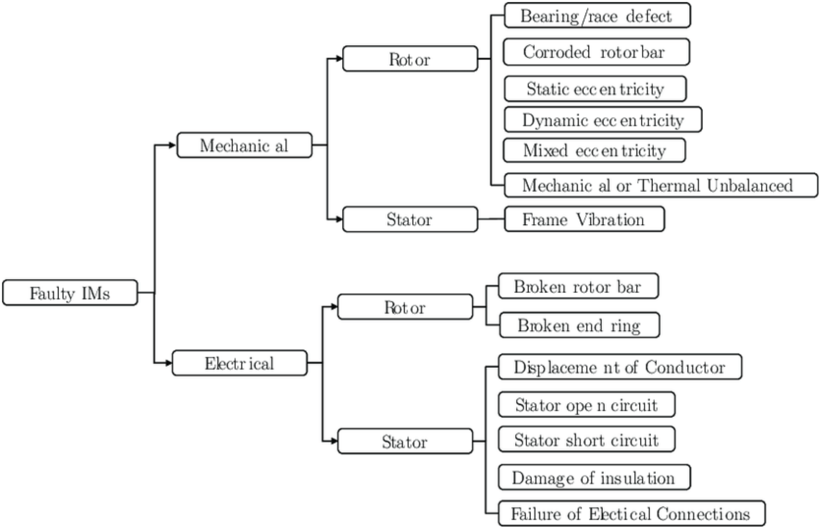


Figure 1: Summary of types of Electrical and Mechanical Faults in an IM

Dataset Details

Experimental Bench Specifications

Specification	Value	Unit
Motor	1/4	CV DC
Frequency range	700-3600	rpm
System weight	22	kg
Axis diameter	16	mm
Axis length	520	mm
Rotor	15.24	cm
Bearings distance	390	mm

Specification	Value	Unit
Number of balls	8	
Balls diameter	0.7145	cm
Cage diameter	2.8519	cm
FTF	0.3750	CPM/rpm
BPFO	2.9980	CPM/rpm
BPFI	5.0020	CPM/rpm
BSF	1.8710	CPM/rpm

Figure 2: Experimental Bench Specifications

Dataset Details

Data Acquisition System

- **Three Industrial IMI Sensors, Model 601A01** accelerometers on the radial, axial, and tangential directions:
  - **Sensitivity:** ($\pm$20%) 100 mV per g (10.2 mV per m/s^2)
  - **Frequency range:** ($\pm$3 dB) 16–600000 CPM (0.27–10,000 Hz)
  - **Measurement range:** $\pm$50 g ($\pm$490 m/s^2)
- **One IMI Sensors triaxial accelerometer, Model 604B31**, returning data over the radial, axial, and tangential directions:
  - **Sensitivity:** ($\pm$20%) 100 mV per g (10.2 mV per m/s^2)
  - **Frequency range:** ($\pm$3 dB) 30–300000 CPM (0.5–5,000 Hz)
  - **Measurement range:** $\pm$50 g ($\pm$490 m/s^2)
- **Monarch Instrument MT-190** analog tachometer
- **Shure SM81** microphone with frequency range of 20–20,000 Hz
- **Two National Instruments NI 9234** 4-channel analog acquisition modules, with a sample rate of 51.2 kHz

Dataset Details

Sequences

Each sequence was generated at a 50 kHz sampling rate during 5 s, totaling 250.000 samples. Below is a summary of each type of sequence:

- Normal Sequence: There are 49 sequences without any fault, each with a fixed rotation speed within the range from 737 rpm to 3686 rpm with steps of approximately 60 rpm.
- Imbalance Faults: There are 33 sequences. Simulated with load values within the range from 6 g to 35 g. For each load value below 30 g, the rotation frequency assumed in the same 49 values employed in the normal operation case.
- Horizontal Parallel Misalignment: This type of fault was induced into the MFS by shifting the motor shaft horizontally of 0.5 mm, 1.0 mm, 1.5 mm, and 2.0 mm.
- Vertical Parallel Misalignment: This type of fault was induced into the MFS by shifting the motor shaft vertically of 0.51 mm, 0.63 mm, 1.27 mm, 1.40 mm, 17.8 mm and 1.90 mm

Model Preparation and Training

- **Label Encoding**: Converted target labels (e.g., "healthy", "Imbalance", "Vertical Alignment", "Horizontal Misalignment") to numeric format.
- **Data Split**: Divided data into training and testing sets.
- **Classifiers Used**:
  - Multiple classifiers, including Logistic Regression, Random Forest, SVM, KNN, Decision Tree, AdaBoost, and Gradient Boosting, and Stacking Classifier were trained and evaluated to identify best performing models.
- **Training Process**: Trained models on `(x_train, y_train)` and made predictions on both training and testing data.
- **Evaluation Metrics**:
  - Accuracy, classification report, and confusion matrix generated for each model.

Imbalance - Without Pre-Processing

For this model, a total of 45 sequences were used for each weight class (6, 10, 15, 20, 25, 30, and 35 g).

- *Classifiers Used*: Logistic Regression, Random Forest, Support Vector Classifier, Decision Tree, K-Nearest Neighbors, Naive Bayes, Gradient Boosting, AdaBoost, Ridge Classifier, XGBoost, Stacking Classifer
- We found XGBoost to be the best performing classifier with an accuracy of 0.88. Below is the classification report for the same.

```
Model: XGBoost
Accuracy: 0.88
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.87      0.88      3000
           1       0.87      0.89      0.88      3000

    accuracy                           0.88      6000
   macro avg       0.88      0.88      0.88      6000
weighted avg       0.88      0.88      0.88      6000
```

Imbalance - with Pre-Processing

- *Classifiers Used*: Same classifiers as for Imbalance faults without pre-processing.
- Pre-Processing Techniques:
  - **Sliding Window:** Divide signal into segments of a fixed length.
  - **Statistical Features:** *Mean* (μ): Average of signal values, *Variance* (σ^2): Measure of signal, *Skewness* (γ_1): Asymmetry of the signal, *Kurtosis* (γ_2): Sharpness of the signal peak.
  - **Power Spectral Density (PSD):** Describes the distribution of signal power over frequency.
  - **Welch Method:** Estimate PSD using overlapped signal segments.
- We found Stacking Classifier to be the best performing with an accuracy of 0.99. Below is the performance report.

	precision	recall	f1-score	support
0	1.00	0.99	0.99	3376
1	0.99	1.00	1.00	3787
accuracy			0.99	7163
macro avg	1.00	0.99	0.99	7163
weighted avg	1.00	0.99	0.99	7163

Vertical - Without Pre-processing

- Classifiers Used: Random Forest, Decision Tree, KNN-Classifier
- We have used 1% of the total available dataset for reduction of computational burden. The shaft misalignment of 0.51 mm, 0.63 mm, 1.27 mm, 1.40 mm, 1.78 mm and 1.90 mm were each considered to be a different class.
- We found Random Forest to be the best performing classifier with an accuracy of 0.36. Below is the performance report for the same.

Model: Random Forest
Accuracy: 0.36
Classification Report:

	precision	recall	f1-score	support
0	0.40	0.36	0.38	38294
1	0.43	0.40	0.41	37225
2	0.35	0.37	0.36	37488
3	0.29	0.26	0.27	37679
4	0.34	0.32	0.33	37634
5	0.36	0.48	0.41	37430
accuracy			0.36	225750
macro avg	0.36	0.36	0.36	225750
weighted avg	0.36	0.36	0.36	225750

Vertical - With Pre-Processing

- Classifiers Used: Random Forest, Decision Tree, KNN-Classifier
- Signal processing technique: Fast Fourier Transform (FFT) was used to find out the frequency components of the dataset.
- All shaft misalignment distances were merged into a single fault class.

```
Training and Evaluating: Random Forest
Cross-Validation Accuracy: 0.96 ± 0.00
Test Accuracy: 0.96
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.91      0.94     50627
           1       0.96      0.99      0.97     99873

    accuracy                           0.96    150500
   macro avg       0.97      0.95      0.96    150500
weighted avg       0.96      0.96      0.96    150500
```

Horizontal - Without Pre-Processing

- Classifiers Used: Random Forest
- We have used 5% of the total available dataset for reduction of computational burden.
- The misalignments of 0.5 mm, 1 mm, 1.5 mm and 2 mm are considered from the dataset, but only 1.5 mm and 2 mm are included for the fault class.
- The performance report is given below -

```
Accuracy: 0.7230162601626017

Classification Report:
              precision    recall  f1-score   support

           0       0.73      0.87      0.79    370248
           1       0.72      0.50      0.59    244752

    accuracy                           0.72    615000
   macro avg       0.72      0.69      0.69    615000
weighted avg       0.72      0.72      0.71    615000
```

Horizontal - With Pre-Processing

- Classifiers Used: Random Forest, Logistic Regression and XGBoost
- We have used 5% of the total available dataset for reduction of computational burden.
- The misalignments of 0.5 mm, 1 mm, 1.5 mm and 2 mm are considered from the dataset, but only 1.5 mm and 2 mm are included for the fault class.
- We found Logistic Regression to be the best performing classifier with an accuracy of 0.6014. The performance report is given below.

Classification Report:

	precision	recall	f1-score	support
0	0.60	1.00	0.75	370248
1	0.00	0.00	0.00	244752
accuracy			0.60	615000
macro avg	0.30	0.50	0.38	615000
weighted avg	0.36	0.60	0.45	615000

Multi-class - With Pre-Processing

- Out of the 45 available sequences, 31 were used as training data, and 14 were used as testing data.
- We classified the machine condition into 4 different classes as given below -
  - Class 0: Normal (healthy) machine
  - Class 1: Imbalance Fault
  - Class 2: Vertical Parallel Misalignment
  - Class 3: Horizontal Parallel Misalignment
- Signal Processing techniques:
  - *Mean* (μ), *Variance* (σ^2), *Skewness* (γ_1), *Kurtosis* (γ_2) and the *PSD* (using Welch Method) was computed.
  - Standardization was carried out.
- Classifiers Used: Logistic Regression, Random Forest, SupportVector Classifier, Decision Tree, K-Nearest Neighbors, Naive Bayes, Gradient Boosting, AdaBoost, Ridge Classifier, XGBoost.
- Random Forest was found to be the best performing classifier.

Multi-class - With Pre-Processing

Model: Random Forest

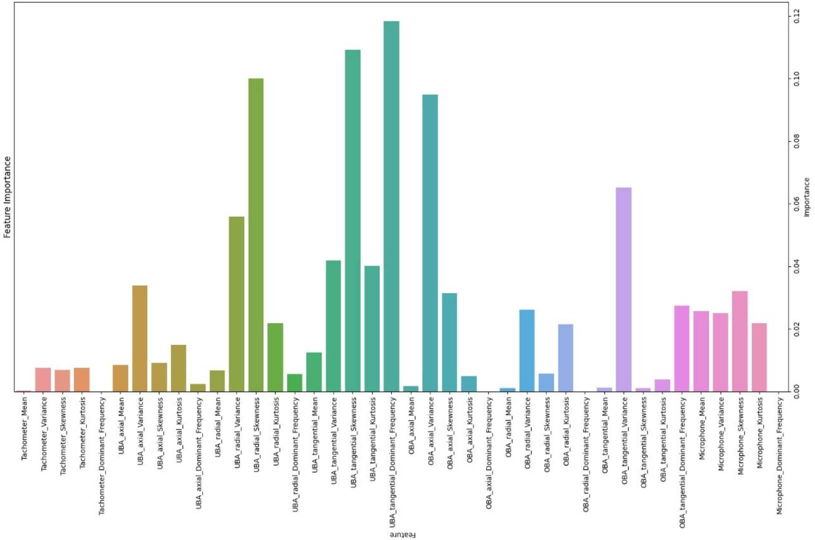
Accuracy: 0.98

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	3376
1	0.99	0.99	0.99	3787
2	0.96	0.99	0.97	3246
3	0.98	0.97	0.97	3604
accuracy			0.98	14013
macro avg	0.98	0.98	0.98	14013
weighted avg	0.98	0.98	0.98	14013

Figure 3: Performance Report for Random Forest Classifier in our Multi-class model

Feature Importance



ROC Curves

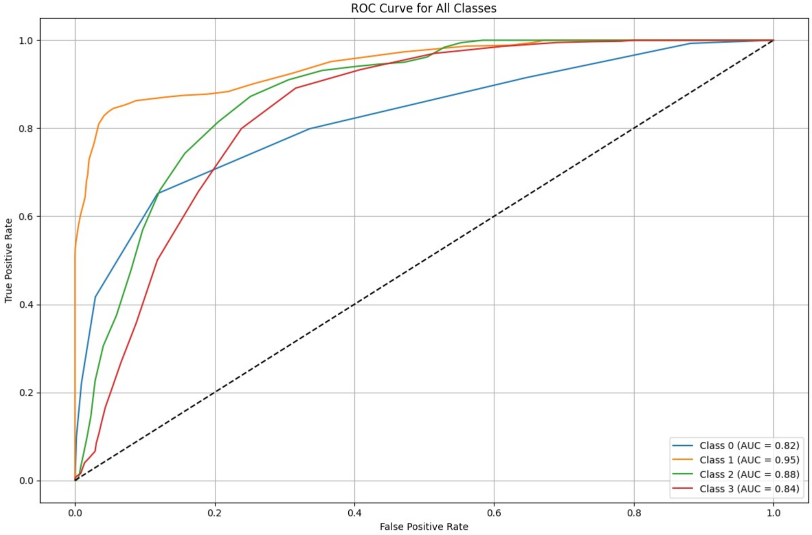


Figure 4: ROC Curves

Observations and Results

Observations

- In the multi-class model, the features consisting of the mean parameters of different signals have lower significance, and can be dropped for dimensionality reduction.
- Generally, we see that pre-processing improves the performance of our models significantly (by *upto 60%*). This is because we were able to better interpret the hidden characteristics of the available features.

Results

- Even though Imbalance, and Vertical and Horizontal misalignment faults may co-exist for more than 80% of the cases, our multi-class model accurately classifies the faults in more than 98% of cases.

Future Work

- Expand model to detect additional motor faults (e.g., underhang, overhang, broken rotor bar).
- Incorporate online datasets with diverse fault types and based on diverse simulation models for broader and more robust fault classification. Validate these results against performance of in-house obtained data from data-acquisition setups.
- Enhance model's versatility, reducing industrial downtime and operational costs through comprehensive predictive maintenance.

References

1. *MAFAULDA (Motor Faults Database), Felipe M.L. Ribeiro (https://www02.smt.ufrj.br/ offshore/mfs/page$_0$1.htmlSEC1).*
2. *Terron, Carla Martinez-Roman, J. Puche-Panadero, Ruben Sapena-Bañó, Angel. (2021). A Review of Techniques Used for Induction Machine Fault Modelling. Sensors. 21. 4855. 10.3390/s21144855.*
3. *Chaudhuri, S. Introduction to Machine Learning Assignment. GitHub repository. Retrieved from (https://github.com/Soham-Chaudhuri/Introduction-to-Machine-Learning-Assignment).*