A Project Report

On

CLSSIFICATION OF DEGREE OF FAULT IN INDUCTION MACHINE

BY

SIDDHARTH SINGH 2020B5A32029H

Under the supervision of

DR. SUDHA RADHIKA

SUBMITTED IN COMPLETE FULFILLMENT OF THE REQUIREMENTS OF

EEE F366: LABORATORY PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(DECEMBER 2023)

ACKNOWLEDGMENTS

I would like to express my profound gratitude to the individuals and organizations whose unwavering support and invaluable contributions made the realization of this project possible.

My heartfelt thanks go to my supervisor Dr. Sudha Radhika for their sagacious guidance, encouragement, and profound insights throughout the journey. Your mentorship has been instrumental in shaping our understanding and refining our approach in the intricate field of prognosis of fault in induction machine.

My appreciation extends to the participants in my research, whose willingness to engage, share insights, and provide critical feedback enriched our understanding and strengthened the empirical foundation of this project. Your participation was invaluable.

I would also like to acknowledge the broader scientific community whose groundbreaking work and discoveries continue to inspire and inform our research. Your pioneering efforts have laid the groundwork for the advancements I have made.

Finally, to my family and loved ones, who stood by me with unwavering support and patience during the long hours and intense moments of this project, I offer my heartfelt thanks. Your understanding and encouragement were my pillars of strength.

In conclusion, this project stands as a testament to the power of collaboration, dedication, and innovation. With deep gratitude, I acknowledge the collective efforts and commitment that have brought me to this milestone.

Birla Institute of Technology and Science-Pilani,

Hyderabad Campus

Certificate

This is to certify that the project report entitled "**Classification of degree of fault in induction machine"** submitted by Mr. Siddharth Singh (2020B5A32029H) in complete fulfillment of the requirements of the course EEE F366, Laboratory Project Course, embodies the work done by him under my supervision and guidance.

Date: **(Dr. Sudha Radhika)**

7th DECEMBER, 2023 BITS- Pilani, Hyderabad Campus

ABSTRACT

The goal of this research is to improve the degree of fault categorization in induction machines through the integration of cutting-edge machine learning techniques. In order to provide a solid basis for fault diagnosis, the Generalized Hurst Exponent and R/S analysis approach were initially used in the feature extraction phase to examine motor current signatures. Expanding upon this, the primary algorithm designed integrates a Random Forest classifier with a Recurrent Neural Network (RNN). It is anticipated that this innovative method will outperform conventional models in precisely detecting fault states. Comparative analysis was conducted using conventional machine learning algorithms, both with and without Principal Component Analysis (PCA), to benchmark the effectiveness of the proposed RNN-Random Forest model. Results from Logistic Regression, Support Vector Machine, K-Nearest Neighbors, Gaussian Naive Bayes, Decision Trees, and Random Forest demonstrate varying degrees of accuracy and precision, with the Random Forest model showing particularly promising results. This project's outcomes not only contribute significantly to fault diagnosis in induction machines but also open avenues for further research in combining neural networks with ensemble learning for industrial applications.

CONTENTS

CHAPTER - 1

Introduction

For a variety of industrial applications, induction motor dependability and efficiency are essential. These motors can, however, develop a variety of problems that can cause serious disruptions to operations and financial losses. For these reasons, prompt and precise fault diagnosis is essential to keeping these devices healthy and operating properly. This study uses cutting-edge machine learning approaches to improve detection accuracy and reliability while addressing the problem of fault classification in induction motors.

Within the field of fault diagnostics, conventional techniques frequently find it difficult to handle the intricacy and diversity of motor defect signals. This constraint emphasizes the requirement for more advanced and flexible methods. A potential remedy is provided by machine learning, which can learn from and predict data. The Generalized Hurst Exponent and R/S analysis approach were employed in the first phase of this project to extract features. In order to properly create the foundation for defect identification, these methods were utilized to examine motor current signatures.

Developing a hybrid model that combines a Random Forest classifier and a Recurrent Neural Network (RNN) is at the heart of this research, building on this basis. This novel method seeks to combine the decision-making prowess of Random Forest classifiers with the temporal pattern recognition capabilities of RNNs to produce superior fault classification performance. It is anticipated that this model will perform better than the conventional machine learning techniques, which were also assessed in order to provide comparisons. These common techniques can be used with or without Principal Component Analysis (PCA), and they include Gaussian Naive Bayes, Random Forest, K-Nearest Neighbors, Support Vector Machines, and Logistic Regression.

This research is significant not only because it has the potential to enhance induction motor fault diagnosis, but also because it can be applied more broadly to other domains where machine health monitoring is essential. This project advances the state-of-the-art in fault detection techniques, which helps to achieve the overarching objective of improving operational effectiveness and safety in industrial settings.

CHAPTER 2: METHODOLOGY

2.1 Data Preprocessing and Feature Extraction

2.1.1 Data Collection: Data for this project is sourced from a specific directory, where historical records of machine performance are stored. The code reads data files and proceeds with further analysis.

2.1.2 Data Preprocessing: Data preprocessing is essential to ensure data quality. The code implements steps for cleaning and formatting raw data, addressing missing values, and handling outliers if present.

2.1.3 Feature Extraction: Feature extraction is a critical step in fault classification. The project employs two methods, the Generalized Hurst Exponent and R/S analysis, to transform raw data into meaningful features that capture patterns in machine behavior.

2.1.4 Code Implementation

- H_Genhurst calculates the Generalized Hurst Exponent, while the R/S analysis method is applied to estimate the Hurst exponent.
- These functions read data files, extract relevant columns, and perform the necessary calculations for feature extraction.

2.1.5 Results and Interpretation: The results of feature extraction, specifically the Hurst exponent values, are critical for understanding the behavior of the induction machines. These values are analyzed to provide insights into the degree of fault severity.

2.2 RNN with Random Forest Classifier

The innovative approach of this study integrates a Recurrent Neural Network (RNN) with a Random Forest classifier to enhance the fault classification accuracy in induction motors.

RNN Model Architecture: TensorFlow and Keras libraries are used in the design of the RNN model. It consists of 64 units in the SimpleRNN layer, each unit specifically designed to process a sequence of 18 features. A Dense layer with a sigmoid activation function for binary classification comes after this layer. The binary cross-entropy loss function and the Adam optimizer are used to compile the model. A ModelCheckpoint callback is used to save the best model based on accuracy in order to maximize performance.

Training Process: Preprocessed data is used to train the RNN model, which is then reshaped to meet the RNN's input specifications. Batch sizes and epochs are used in the training process, with the best accuracy and epoch being recorded. After that, the model is saved and ready to be loaded again for more forecasts.

Random Forest Integration: Features taken from the residuals of the RNN model are used to train the Random Forest Classifier, which is implemented using the sklearn library. The residuals' mean and standard deviation for every class are among these characteristics. The goal of the RNN and Random Forest integration is to take advantage of Random Forests' strong classification capabilities and RNNs' strength in sequential data processing.

Preliminary Results: *The current implementation of this hybrid model has demonstrated a promising accuracy of* ***95.61%.*** However, it is important to note that this model is yet to undergo fine-tuning. Further optimization of hyperparameters and model structure is expected to enhance its performance and reliability in fault classification.

Ongoing and Future Work: We are currently fine-tuning the model to maximize its performance. This involves modifying the training procedure, hyperparameters, and network architecture. By striking a balance between computational efficiency and accuracy, the model will be more reliable and useful for real-world applications.

RESULTS

Results without PCA

Out[4]:

	Model	Accuracy	Precision	F1 Score
0	Logistic Regression	0.947500	0.951368	0.949068
1	Support Vector Machine	0.933750	0.938184	0.933894
2	K-Nearest Neighbors	0.836250	0.834111	0.834529
3	Gaussian Naive Bayes	0.879375	0.881181	0.873689
4	Decision Tree	0.965625	0.967029	0.966210
5	Random Forest	0.990000	0.990190	0.990134

Optimized Result Without PCA.

	Model	Accuracy	Precision	F1 Score	Best Parameters	Best Score
0	Logistic Regression	0.9475	0.951368	0.94907	C': 10, 'penalty': 'l2', 'solver': 'lbfgs'	0.95999997
1	Support Vector Machine	0.93375	0.938184	0.93389	C': 10, 'gamma': 'scale', 'kernel': 'linear'	0.972343
2	K-Nearest Neighbors	0.83625	0.834111	0.83453	algorithm': 'ball_tree', 'n_neighbors': 7, 'weights': 'distance'	0.841095
3	Gaussian Naive Bayes	0.879375	0.881181	0.87369	var_smoothing': 1e-07	0.868438
4	Decision Tree	0.9675	0.968635	0.96795	"criterion': 'gini', 'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 5	0.963281
5	Random Forest	0.99	0.990078	0.99012		

Results With PCA

Out[4]:

	Model	Accuracy	Precision	F1 Score
0	Logistic Regression	0.886250	0.887736	0.885044
1	Support Vector Machine	0.908125	0.911292	0.907132
2	K-Nearest Neighbors	0.853125	0.851751	0.852365
3	Gaussian Naive Bayes	0.800625	0.805016	0.782217
4	Decision Tree	0.881875	0.883606	0.883245
5	Random Forest	0.915000	0.915151	0.914255

Optimized Result With PCA

Model	Accuracy	Precision	F1 Score	Best Parameters	Best Score
Logistic Regression	0.88625	0.887736	0.88504	'C': 10, 'penalty': 'l2', 'solver': 'lbfgs'	0.89250063
Support Vector Machine	0.908125	0.911292	0.90713	'C': 10, 'gamma': 'scale', 'kernel': 'rbf'	0.93234387
K-Nearest Neighbors	0.853125	0.851751	0.85237	'algorithm': 'ball_tree', 'n_neighbors': 5, 'weights': 'distance'	0.84999992
Gaussian Naive Bayes	0.800625	0.805016	0.78222	var_smoothing': 1e-09	0.80593756
Decision Tree	0.880625	0.882165	0.8817	'criterion': 'gini', 'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2	0.85718781
Random Forest	0.914375	0.913715	0.9137		

FUTURE PLANS

The model is in the process of being fine-tuned to optimize its performance. This includes adjustments to the network architecture, hyperparameters, and training process. The aim is to achieve a balance between accuracy and computational efficiency, making the model more robust and practical for real-world applications.

CONCLUSION

This project represents a significant stride in the field of fault classification in induction motors. By innovatively integrating a Recurrent Neural Network (RNN) with a Random Forest classifier, the study demonstrates a promising approach to enhancing diagnostic accuracy and reliability. The RNN model, specifically designed for sequence processing, effectively captures temporal patterns in motor current signatures, while the Random Forest classifier contributes to robust decision-making based on the features extracted from RNN residuals.

With a preliminary accuracy of 95.61%, the model shows considerable potential, though it is still in the stage of fine-tuning and optimization. This continuous improvement process aims not only to increase accuracy but also to refine the model for practical, real-world applications, balancing computational efficiency and predictive power.

The implications of this research extend beyond the immediate application of fault detection in induction motors. It sets a precedent for the application of hybrid machine learning models in industrial diagnostics, opening new avenues for research and development in this domain. As the project moves forward, further enhancements and validations of the model are anticipated, contributing to the broader goal of achieving operational excellence and safety in industrial environments.

References

An integrated condition monitoring scheme for health state identification of a multi-stage gearbox through Hurst exponent estimates

- Vamsi Inturi1 , Sai Venkatesh Balaji2 , Praharshitha Gyanam2 , Brahmini Priya Venkata Pragada2 , Sabareesh Geetha Rajasekharan2 and Vikram Pakrashi3

Fault Detection and Isolation of Induction Motors Using Recurrent Neural Networks and Dynamic Bayesian Modeling

- Hyun Cheol Cho, Jeremy Knowles, M. Sami Fadali, and Kwon Soon Lee