

Novel Hybrid Invasive Weed Optimization and Machine Learning Approach for Fault Detection

Alasmer Ibrahim
Wolfson Centre for Magnetics
Cardiff University
Cardiff CF24 3AA, UK
Ibrahim21@cardiff.ac.uk

Fatih Anayi
Wolfson Centre for Magnetics
Cardiff University
Cardiff CF24 3AA, UK
Anayi@cardiff.ac.uk

Michael Packianather
High-Value Manufacturing Group
Cardiff University
Cardiff CF24 3AA, UK
packianatherms@cardiff.ac.uk

Osama Al-Omari
Computer Science
Istanbul Gelisim University
Istanbul, Turkey
oalomari@gelisim.edu.tr

Abstract— Fault diagnosis of anomalies in induction motors is essential to ensure industry safety. This paper presents a new hybrid Invasive Weed Optimization and Machine Learning approach for fault diagnosis in an induction motor. The vibration signal provides a lot of information about the motor's operating conditions. Therefore, the vibration signal of the motor was chosen to investigate the fault diagnosis. Two identical 400-V, 50-Hz, 4-pole 0.75 HP induction motors were under healthy, mechanical, and electrical faults tested in a laboratory with different loading. A hybrid model was developed using the vibration signal, the Invasive Weed Optimization algorithm (IWO), and machine learning classifiers. Some statistical features were extracted from the signal using Discrete Wavelet Transform (DWT). The invasive weed optimization algorithm (IWO) was utilized to reduce the number of the extracted features and select the most suitable ones. Then, three classification algorithms namely k-Nearest Neighbor neural network (KNN), Support Vector Machine (SVM), and Random Forest (RF), were trained using k-fold cross-validation and tested to predict the true class. The advantage of combining these techniques is to reduce the training time and increase the average accuracy of the model. The performance of the proposed fault diagnosis model was evaluated by measuring the Specificity, Accuracy, Precision, Recall, and F1_score. The experimental results prove that the proposed model has achieved more than 99.90% of accuracy. Furthermore, the other evaluation parameters also show the same representation of performance. The hybrid model has proved successfully its robust for diagnosing the faults under different load conditions.

Keywords—Fault Diagnosis, Induction Motor, Machine Learning Classifiers, Discrete Wavelet Transform (DWT), Invasive Weed Optimization Algorithm (IWO).

I. INTRODUCTION

Induction motors are commonly used as prime movers in the power systems, transportation, and petrochemical industries because of their low cost, easy to control, high reliability. However, they are subjected to many electrical and mechanical stresses due to their long operating times. The initial fault can result the downtime of the motor and increased production loss if it is not diagnosed in the initial stages [1]. Therefore, Condition Monitoring (CM) is very significant to reduce the damage, lengthen machine life, increase machinery availability, and reduce the breakdown maintenance [2].

The common types of motor defects are the bearing, stator, and rotor faults [3]. Among, the possibility of these faults occurring, the bearing fault is the highest, which occurs more than 40% as stated in research by General Electric Co. and IEEE-IGA [4]. Besides, this kind of fault is the main cause of the motor initial failure, which supports the fact that bearing fault diagnosis is important in the condition monitoring of rotating machinery.

Condition monitoring techniques based on machine learning operate by analyzing previous data that gathered from the machine using various sensors, and under various operating conditions. The strategy is to apply feature extraction and feature selection techniques to the raw signals to create a feature matrix, and then the models are trained by these features. The feature matrix is generally generated using different techniques such as time domain, which is simple to implement such as root mean square method[5], it also provides a basic understanding of the signal. Frequency domain which determines the signature function without any prior knowledge such as short time Fourier transform [6]. However, frequency-domain analysis is ineffective for non-stationary signals. Then, time-frequency domain methods such as wavelet transform can overcome the issue of analysing the non-stationary signal [7]. For this benefits, time-frequency analysis has been growing in popularity as it performs well for both stationary and nonstationary signals [8].

Interest in the use of Artificial Intelligence (AI) techniques in the process of fault diagnosis is also recently growing up. Because of its superior pattern recognition capabilities, the artificial neural network (ANN) is the most widely used AI technology [9]. It can also represent non-linear models without any knowledge of the really structure in a short amount of time, which resulting in promising outcomes in defect diagnostics [10].

It has been proved in many research studies that feature selection stage is a important process when utilising a machine learning approaches for building successful model. The reason for selecting some specific features is to achieve a better learning performance, simplicity, and lower measurement costs [8].

In practise, there are a variety of feature selection techniques such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO). In [11] GA was successfully implemented as feature selection tool to reduce the data

dimension. Furthermore, Multi-objective PSO-based approach for ranking features based on their frequency (RFP SOFS) in the archive set was proposed [12]. These rankings are then utilised to fine-tune the archive set.

A new artificial neural network (ANN)-based model with invasive weed optimization (IWO) was developed to estimate rock tensile strength [13]. The proposed hybrid of the IWO-ANN model has provided a higher level of accuracy in the prediction. Invasive weed optimization (IWO) was also used to provide a simple and innovative variable selection technique using FTIR and NIR dataset [14]. The results showed that the performance of IWO is a simple and robust optimization technique.

Invasive Weed Optimization algorithm (IWO) based feature selection as depicted in Fig.1 is implemented in few applications to reduce the original number of the extracted features and to ensure both the learning process of the used algorithms and the classification error minimization.

The structure of this work is arranged as follows: Section I presents the introduction; Section II presents the principle of the proposed technique; Section III presents experimental setup and dataset acquisition; Section IV presents the proposed model; Sections V and VI discuss the results, and the conclusion, respectively.

II THE PRINCIPLE OF THE PROPOSED TECHNIQUE

A. Invasive Weed Optimization (IWO)

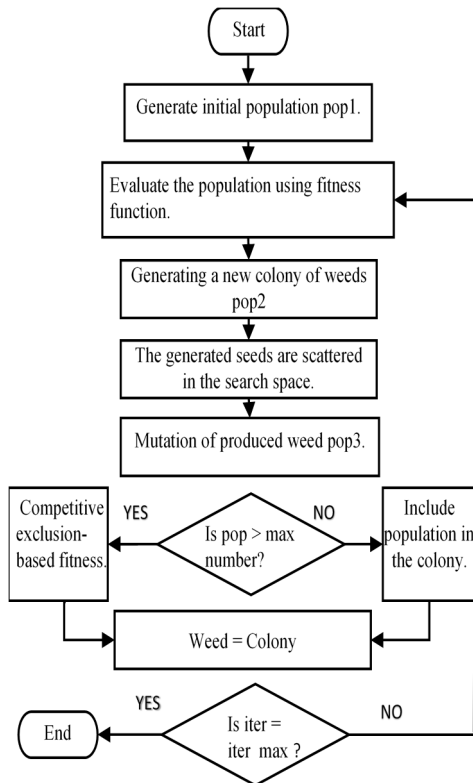


Fig.1 A flow chart of invasive weed optimization.

Invasive Weed Optimization (IWO) is a swarm intelligence metaheuristic algorithm that is inspired by the invasive weeds

colonization behaviour along their journey to find an appropriate place for growth and reproduction. It is very simple and powerful for variable selection.

The overall process of IWO-based feature selection is provided in the following steps:

- **Step 1:** Generate a population of initial solution (weeds), which is referred to as pop1.
- **Step 2:** Each weed's in the population will be evaluated using a certain fitness function.
- **Step 3:** Every weed reproduces new seeds. The weeds will generate seeds ($Seed_i$) that is calculated as given in this formula:

$$Seed_i = S_{min} + (S_{max} + S_{min}) \times \frac{fit_{worst} - fit_i}{fit_{worst} - fit_{best}} \quad (1)$$

Where S_{min} , and S_{max} are the minimum and maximum possible numbers of produced seeds for the worst and best weeds. fit_{worst} and fit_{best} are the fitness values for the worst and best weeds in the colony. And fit_i refers to the fitness value of the weeds in the colony.

- **Step 4:** Spatial dispersal where randomness and adaptation are incorporated into the algorithm. The generated seeds are scattered in the search space, spatial dispersion performance can be given by the following equation:

$$SD_{iter} = \frac{(iter_{max} - iter)^n \times (SD_{initial} - SD_{final})}{(iter_{max})^n} + SD_{final} \quad (2)$$

Where $iter_{max}$ is the maximum number of iterations, and n is the index of nonlinear modulation.

- **Step 5:** Then performing the local search method where the mutation can be applied to prevent premature convergence. The population resulted from mutation operation is named pop3 that its size is equal to that of pop2.

- **Step 6:** An affinity function was applied to avoid premature convergence and extend the diversification.

- **Step 7:** The competitive step is implemented to eliminate the weeds with poor fitness values when the maximum population size is reached in the colony.

Repeat the process will continue until the maximum iterations number is achieved.

B. Wavelet transforms (WT)

Wavelet Transform (WT) is a powerful tool that analysis the signal in time-frequency domain. It has gained a lot of attention in many fields of engineering, particularly in analysing the rotary machines for fault detection and diagnosis [15]. Wavelet analysis has been developed to overcome the non-stationary signal problems in some traditional methods such as Short Time Fourier Transform. WT is not only having the ability in processing the non-stationary signals, but it also has other distinct advantages. Discrete Wavelets Transform (DWT) is a signal processing tool that decomposes the raw signal into multiple coefficients. The procedure of DWT at n level decomposition is summarized as follows:

Firstly, the raw signal is decomposed into approximation coefficients a_1 and detail coefficients d_1 .

At the second decomposition level, a_1 is further decomposed into a_2 and d_2 , and so on. Ddecomposed process will be continued until the desired level n is reached (a_n , d_n).

This can be formulated as given in the equations below:

$$Z_{high} = \sum x[k].g[2k - n] \quad (3)$$

$$Z_{low} = \sum x[k].h[2k - n] \quad (4)$$

Where Z_{low} represents the approximation coefficients and Z_{high} is the detail coefficients, $x[k]$ is the raw signal, and $g[2k-n]$, and $h[2k-n]$ represent the high pass filter and the low pass filter, respectively.

II. EXPERIMENTAL SET UP

A. Test rig and Equipments

Test rig of the present work is as illustrated in Fig.2

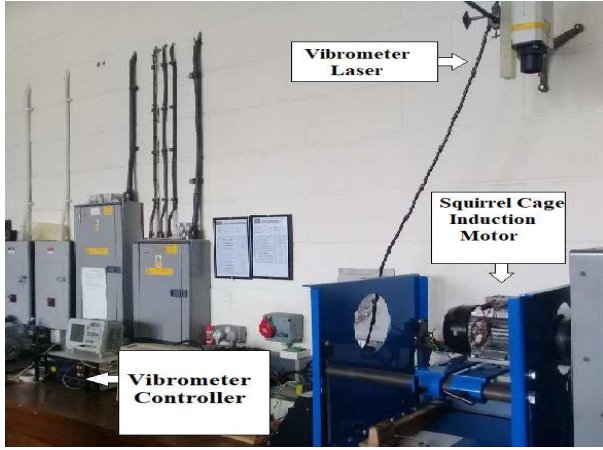


Fig 2: The experimental test rig

The proposed work was executed at Wolfson Centre for Magnetics, Machines and Transformers Laboratory, Cardiff University, UK. Induction motor vibration signal was chosen to be recorded. 15000 data samples in each fault class were collected from a 4-pole, 0.75-HP, 80B-230/400 V, 1380r/min three-phase squirrel cage induction motor (Model Clarke 6430439) using a Polytec OFV 303 vibrometer laser sensor (model OFV-3001). vibrometer laser sensor was connected to vibrometer controller, and its output signal was stored in flash memory by using a digital storage oscilloscope. Three types of faults were developed bearing fault, broken rotor bar fault, and stator fault. The bearing fault was done by drilling a 0.25 cm hole into the outer race as demonstrated in Fig.3(a). Ball bearing fault was made by removing one ball with its cage then the remaining balls are crushed during the load operation, Fig.3(b) describe it.

Broken Rotor Bar fault was produced by drilling a hole with 4.2mm diameter, and a 16mm depth dimension to cut the bar resistance for the one broken rotor bar fault (1BRB) as displayed in Fig.4(a).

The five-broken rotor bars fault was created with the same diameter and depth dimension used before to make the one-broken rotor bar fault, and the holes were separated as displayed in Fig.4(b). Stator fault was an open circuit in one phase turn of the motor stator.

B. Data preparation

A total of four datasets from the vibration signal were

generated as described in the table I.



Fig.3 (a) Outer bearing, (b) Ball bearing faults



Fig.4 (a) One-broken rotor bar, (b) Five-broken rotor bars faults

TABLE I. DESCRIPTION OF THE PROPOSED DATASET

case	Signal Type	N. of Total Samples	class	Load Condition (RPM)	No. of Features
1	Vibration	1800	6	1480	14
2	Vibration	1800	6	1450	14
3	Vibration	1800	6	1400	14
4	Vibration	1800	6	Mixed speed	14

The healthy case and the three different induction motor faults were considered. All gathered data is termed in cases. For example, in case1 data was acquired with a speed of 1480 rpm. While, in case 4, the data are combined for all speeds together (1480, 1450, and 1400 rpm). Table II explains the motor condition features with its class label. Fourteen statistical feature parameters feature as shown in Fig.5 have been extracted in time-frequency domain using DWT with six decomposed levels.

C. Proposed model

The schematic diagram for proposed algorithms that suggested in research are described in Fig.6. It is implemented to show how can the faults of the induction motor to be diagnosed.

The extracted features are fed to Invasive Weed Optimization (IWO) to reduce training time and increase the average accuracy of the model. These features were taken to form four datasets with a feature matrix dimension of 1800×14. Then will be applied to the proposed algorithm IWO to investigate feature selection stage.

Only selected features then are utilized to train three machine learning classifiers namely, k-Nearest Neighbor (KNN), Support Vector Machine (SVM) Random Forest (RF) classifiers applying k-fold cross-validation technique to identify the faults. The fitness function is the most essential component of feature selection algorithm; it must be defined

TABLE II. MACHINE CONDITIONS WITH RELATED LABELS

Machine Conditions	Vibration Signal	
	Class label	No. of Testing Samples
Healthy Motor	Label 1	300
One-Broken Rotor Bar Fault	Label 2	300
Five-Broken Rotor Bar Fault	Label 3	300
Stator Fault	Label 4	300
Inner Bearing Fault	Label 5	300
Outer Bearing Fault	Label 6	300
Sample total number		1800

properly for assessing the discriminative capabilities for each subset of features.

The proposed fitness function in this application is KNN (K-Nearest Neighbor)-based classification error which is described as the following equation:

$$\text{Fit fun} = \frac{\alpha}{N_f} - \exp\left(-\frac{1}{N_f}\right) \quad (5)$$

where α = k-NN-Based classification error and N_f is cardinality of the selected features.

From the total of fourteen features, IWO has selected a minimum of four features which were combined to make different input subsets. This resulted in a suitable input based on the predicted accuracy of the proposed models. The feature matrix that obtained after executing the feature selection process has been used with a cross-validation to train and validate the proposed machine learning algorithms. Some of the different evaluation parameters, namely overall

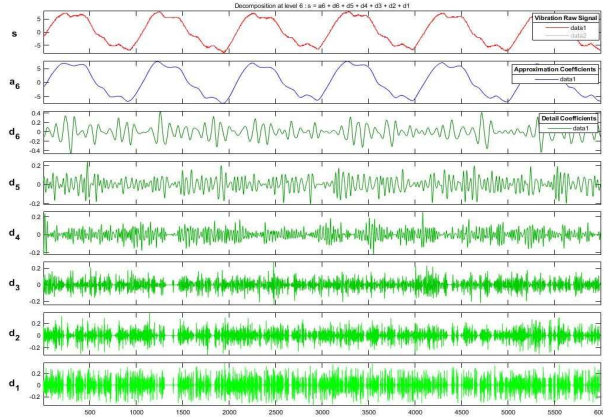


Fig. 5: Processed vibration signal using wavelet packet transform for the motor having one broken bar fault at speed of 1480 rpm

accuracy, sensitivity, specificity, and the F1-score were calculated to identify the performance of the proposed model. The following Equations provide these parameters:

$$\text{Specificity (SPEC)} = \frac{TN}{FP+TN} \quad (6)$$

$$\text{Accuracy (ACC)} = \frac{TP+TN}{TP+FP+TN+FN} \quad (7)$$

$$\text{Precision (PRE)} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall (REC)} = \frac{TP}{TP+FN} \quad (9)$$

$$F1_score (F1_SC) = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

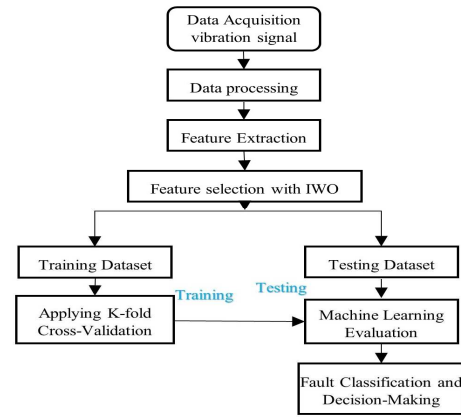


Fig.6: Schematic diagram of the proposed model

In the upcoming section, the results of the hybrid models are in brief presented.

IV. RESULT AND DISCUSSION

The performance of employing IWO with the of the proposed machine learning classifiers for diagnosing the induction motor faults is presented.

The efficient performance to select the optimum features needs to set the feature selection algorithm parameters to tuning values. IWO parameters have been set as follows: Number of initial populations=10, Maximum weed population=25, Maximum number of seeds=10, The final

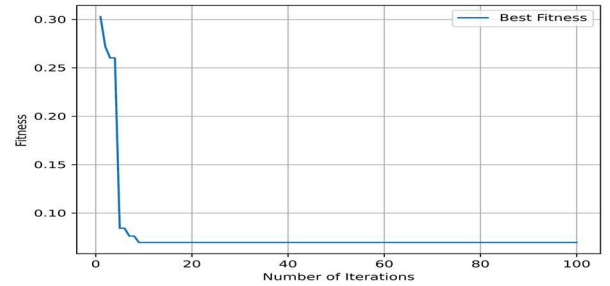


Fig. 7: Convergence diagram for fitness functions of Invasive Weed Optimization

value of standard deviation=0.001, and the Maximum number of iterations=100.

When applying IWO using the dataset that was described in table I (case 4), which combined three motor speeds, a minimum of four features were selected instead of the fourteen original ones. The best cost function that was obtained from the convergence diagram of IWO is depicted in Fig.7, which shows the cost function reaching a value of 0.0480 in 100 iterations.

The classifiers performance was assessed using the evaluation parameters that given in Equations (6) – (10), and Receiver Operating Characteristic curve (ROC).

The area under the curve (AUC) was employed in ROC because it gives an excellent indication of how well a classification model performed on a dataset. The AUC

curve's is ranged between 0 to 1. If it's around or less 0.5, the classifier hasn't done well. When the nearest value is close to 1 that represents the best model.

The combined ROC curves is presented in Fig.8. The overall classification results of employing IWO with the three proposed classifiers are measured to evaluate the model.

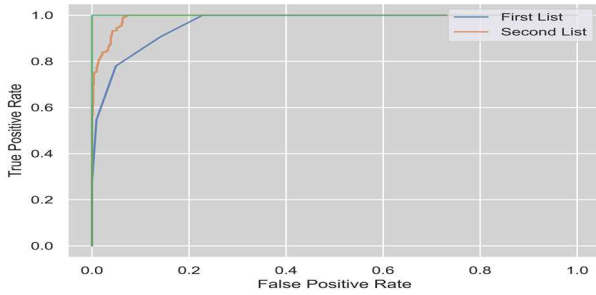


Fig.8: Combined ROC curve for the applied classifiers

The predicted classification curves of KNN, SVC, and RF classifiers are given in Fig. 9, 10, and 11, respectively.

Equations (6)–(10) were used to calculate the evaluation parameters. Machine learning classifiers have been trained using cross-validation. The obtained results that achieved by applying 5-fold cross-validation, and 10-fold cross-validation techniques are given in table III, and table IV, respectively.

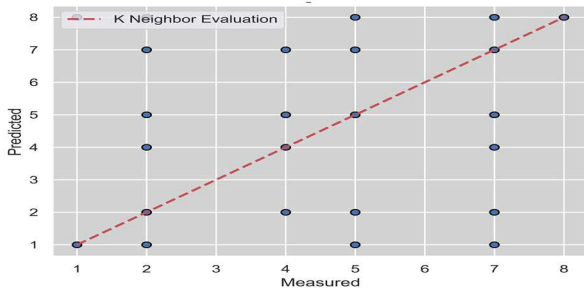


Fig.9: Predicted curve of the K Neighbour classifier

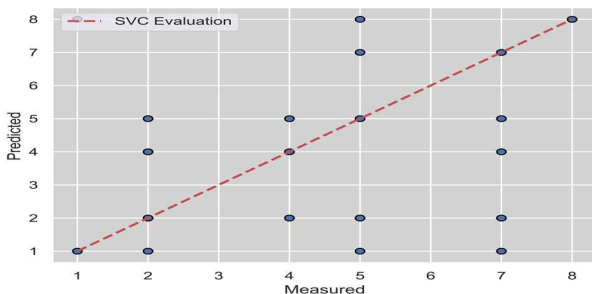


Fig. 10: Predicted curve of Support Vector Classifier

The classification accuracy of RF is 99.96% when the model is trained with 5-fold cross validation, and it is further raised to 99.98 with model is trained using 10-fold cross validation. Furthermore, when SVC model is trained with 5-fold cross-

validation, the accuracy is 95.01%, and it is 95.53% when the model trained with 10-fold cross-validation. In addition, the classification accuracy using KNN classifier, which is less.

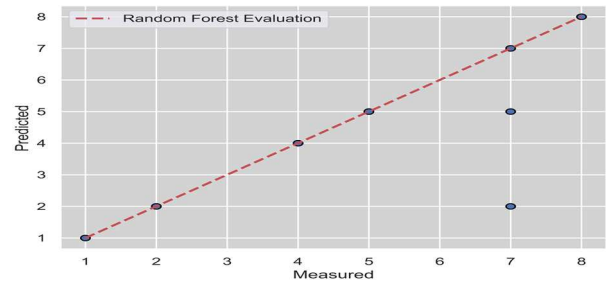


Fig. 11: predicted curve of Random Forest classifier

TABLE III. CLASSIFICATION RESULT APPLYING 5-FOLD CROSS-VALIDATION

Model	SPE %	ACC %	PRE %	RECA %	F1_SC %
KNN	95.17	91.96	78.32	75.88	74.89
SVM	97.00	95.01	86.19	85.11	84.73
RF	99.97	99.96	99.89	99.88	99.88

TABLE IV. CLASSIFICATION RESULT APPLYING 10-FOLD CROSS-VALIDATION

Model	SPEC %	ACCU %	PREC %	RECA %	F1_SC %
KNN	95.40	92.33	79.45	77.00	76.11
SVM	97.32	95.53	87.85	86.61	86.28
RF	99.98	99.98	99.94	99.94	99.94

The accuracy is 91.96% and 92.33% with the utilize of 5-fold and 10-fold cross-validation, respectively. The evaluation parameters such as specificity, precision, recall, and F1-score are given the outcome in the same representations. The AUC score for RF is 1, while it is 0.99 for SVC, and it is 0.96 for KNN, which indicates that RF and SVC models perform well in comparison with the KNN classifier. It can be said that Random Forest classifier performs better than that of KNN and SVM in predicting the true class. Training the proposed models using 10-fold cross validation can improve the classification accuracy for fault diagnosis in induction motor.

VI CONCLUSION

A novel hybrid model-based motor-vibration data-driven approach has been proposed in work. The main goal of this model is precisely to reduce the dimension of the original data and raise the average accuracy. The model

combines statistical features that extracted by discrete wavelet transform (DWT), Invasive Weed Optimization (IWO), and three machine learning classifiers for induction motor fault diagnosis. The machine learning classifiers KNN, SVM, and RF have been trained with cross validation techniques. A minimum of four features out of fourteen time-frequency domain features were efficiently selected o by IWO. The fault classification average accuracy has been further improved and indicated that the novel hybrid model is successful investigates the requirement. It has been gained an accuracy value of 99.98% when the model uses Random Forest classifier that trained with 10-fold cross-validation method.

Concisely, the proposed model using vibration signal of the motor with invasive weed optimization algorithm has been proved to be valid for diagnosing the induction motor faults with different load conditions.

ACKNOWLEDGMENT

Alasmer Ibrahim acknowledges the sponsorship from the High Ministry of Education in Libya.

REFERENCES

- [1] M. Abd-el-Malek, A. K. Abdelsalam, and O. E. Hassan, "Induction motor broke rotor bar fault location detection through envelope analysis of start-up current using Hilbert transform," *Mech. Syst. Signal Process.*, vol. 93, pp. 332–350, 2017, doi: 10.1016/j.ymssp.2017.02.014.
- [2] R. N. Toma, A. E. Prosvirin, and J. M. Kim, "Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers," *Sensors (Switzerland)*, vol. 20, no. 7, 2020, doi: 10.3390/s20071884.
- [3] C. P. Mboo and K. Hameyer, "Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection," *IEEE Trans. Ind. Appl.*, vol. 52, no. 5, pp. 3861–3868, 2016, doi: 10.1109/TIA.2016.2581139.
- [4] G. K. Singh and S. A. S. Al Kazzaz, "Induction machine drive condition monitoring and diagnostic research - A survey," *Electr. Power Syst. Res.*, vol. 64, no. 2, pp. 145–158, 2003, doi: 10.1016/S0378-7796(02)00172-4.
- [5] D. Siegel, C. Ly, and J. Lee, "Methodology and Framework for Predicting Helicopter Rolling Element Bearing Failure," vol. 61, no. 4, pp. 846–857, 2012.
- [6] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez, and R. A. Osornio-Rios, "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," *IEEE Trans. Ind. Informatics*, vol. 9, no. 2, pp. 760–771, 2013, doi: 10.1109/TII.2012.2221131.
- [7] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, 2016.
- [8] C. M. Andersen and R. Bro, "Variable selection in regression-a tutorial," *J. Chemom.*, vol. 24, no. 11–12, pp. 728–737, 2010, doi: 10.1002/cem.1360.
- [9] S. Altaf, M. W. Soomro, and M. S. Mehmood, "Fault diagnosis and detection in industrial motor network environment using knowledge-level modelling technique," *Model. Simul. Eng.*, vol. 2017, 2017, doi: 10.1155/2017/1292190.
- [10] M. Skowron, M. Wolkiewicz, T. Orlowska-Kowalska, and C. T. Kowalski, "Application of self-organizing neural networks to electrical fault classification in induction motors," *Appl. Sci.*, vol. 9, no. 4, 2019, doi: 10.3390/app9040616.
- [11] R. N. Toma, A. E. Prosvirin, and J. M. Kim, "Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers," *Sensors (Switzerland)*, vol. 20, no. 7, 2020, doi: 10.3390/s20071884.
- [12] M. Amoozegar and B. Minaei-Bidgoli, "Optimizing multi-objective PSO based feature selection method using a feature elitism mechanism," *Expert Syst. Appl.*, vol. 113, pp. 499–514, 2018, doi: 10.1016/j.eswa.2018.07.013.
- [13] L. Huang, P. G. Asteris, M. Koopialipoor, D. J. Armaghani, and M. M. Tahir, "Invasive weed optimization technique-based ANN to the prediction of rock tensile strength," *Appl. Sci.*, vol. 9, no. 24, 2019, doi: 10.3390/app9245372.
- [14] S. Sheykhizadeh and A. Naseri, "An efficient swarm intelligence approach to feature selection based on invasive weed optimization: Application to multivariate calibration and classification using spectroscopic data," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 194, pp. 202–210, 2018, doi: 10.1016/j.saa.2018.01.028.
- [15] M. Aktas and V. Turkmenoglu, "Ten Lectures on Wavelets," *IET Sci. Meas. Technol.*, vol. 4, no. 6, pp. 303–310, 2010, doi: 10.1049/iet-smt.2009.0121.
- [16] M. Z. Ali, M. N. S. K. Shabbir, X. Liang, Y. Zhang, and T. Hu, "Machine learning-based fault diagnosis for single- and multi-faults in induction motors using measured stator currents and vibration signals," *IEEE Trans. Ind. Appl.*, vol. 55, no. 3, pp. 2378–2391, 2019, doi: 10.1109/TIA.2019.2895797.