Induction motor short circuit diagnosis and interpretation under voltage unbalance and load variation conditions

Avyner L.O. Vitor [a,b,*], Alessandro Goedtel [a], Sylvio Barbon Junior [c], Gustavo H. Bazan [a,b], Marcelo F. Castoldi [a], Wesley A. Souza [a]

[a] Federal University of Technology—Parana, Alberto Carazzai Avenue 1640, Cornelio Procopio, 86300-000, Parana, Brazil
[b] Federal Institute-Parana, S/N Doutor Tito Avenue, Jacarezinho, 86400-000, Parana, Brazil
[c] University of Trieste, Piazzale Europa, 34127, Trieste, Friuli-Venezia Giulia, Italy

ARTICLE INFO

Keywords:
Explainable artificial intelligence
Machine learning
Fault diagnosis
Induction machines
Inter-turn short circuit
Shapley additive explanations

ABSTRACT

Predictions from machine learning algorithms have often supported decision-making in industrial processes. Despite this, complex models can be challenging to interpret, sometimes shrouding the entire prediction process in an undesirable mystery. Understanding how the classifiers' recommendations are made helps human experts understand the phenomenon and develop better data-driven solutions. Therefore, this study takes advantage of Shapley additive explanations to explain the predictions obtained by the classifier and select the most appropriate features for the approaches. The experiments use extreme gradient boosting to evaluate temporal, spectral, and wavelet features of three-phase induction motor current signals. The proposed approach effectively reduces the number of attributes without losing performance, provides an understanding of how each feature affects the model over a wide range of voltage unbalances and torque values, and detects early inter-turn short circuits with severity of 1%. The results show that combining the intelligent model with Shapley explanations improves stator winding fault diagnosis in these highly problematic situations.

1. Introduction

The three-phase induction motor (TIM) is considered a robust and reliable electrical machine, but it is often exposed to unsuitable conditions that can lead to failure. Power quality issues, excessive overloading operation, and natural aging are common factors that lead to insulation degradation and inter-turn short circuit (Gundewar & Kane, 2020; Merizalde et al., 2017). If an inter-turn short is not detected early, it can quickly evolve into coil-to-coil, phase-to-phase, and phase-to-ground faults. In this case, the damage to the machine could be irreversible (Husari & Seshadrinath, 2022). In addition, according to Gundewar and Kane (2020), Gyftakis and Cardoso (2020) and Merizalde et al. (2017), about 28% to 40% of TIM failures are related to stator windings. For these reasons, short circuits in windings are an important problem in the industrial context, and an incipient defect must be detected as soon as possible.

A monitoring strategy is required to prevent motor failures and reduce downtime for maintenance (Gyftakis & Cardoso, 2020). Model-based approaches rely on complex mathematical models that describe the machine's behavior. In this case, the strategy may use equivalent circuits, finite element methods, or analytical models. However, the diagnosis' reliability depends on the quality of the mathematical model that simulates the adverse conditions to which the motor may be subjected. The signal-based method allows the creation of fault indices from various signals collected from TIM. Some typical signals are stator current and voltage, vibration, audio, and electromagnetic flux.

The study by Drif and Cardoso (2014) highlights that most inter-turn short circuit (ITSC) diagnostic strategies are sensitive to phase imbalances in the power grid. The authors stated that distinguishing between a failure and this power quality problem is complex and almost impossible at very early stages. According to Alloui et al. (2022), voltage imbalances lead to similar symptoms as winding faults. The proposed approach to overcome this problem involves the application of short-time least square Prony's (STLSP) in combination with the Fortescue transform to voltage signals. In the recent work of Sonje et al. (2019), the discrete wavelet transform (DWT) and Park's vector modulus (PVM) are applied to current signals. The authors demonstrated the effectiveness of the fault indicator in the face of voltage fluctuations and different load levels. In these works, elaborated strategies have been presented, and their potential for ITSC identification has been

* Corresponding author at: Federal University of Technology—Parana, Alberto Carazzai Avenue 1640, Cornelio Procopio, 86300-000, Parana, Brazil.
E-mail addresses: avyner.vitor@ifpr.edu.br (A.L.O. Vitor), agoedtel@utfpr.edu.br (A. Goedtel), sylvio.barbonjunior@units.it (S. Barbon Junior), gustavo.bazan@ifpr.edu.br (G.H. Bazan), marcastoldi@utfpr.edu.br (M.F. Castoldi), wesleyangelino@utfpr.edu.br (W.A. Souza).

demonstrated. However, for industrial applications, it is critical to implement an automatic process.

In this context, a machine learning (ML) algorithm is usually implemented to detect the fault patterns automatically (Brito et al., 2022). Support vector machine (SVM) is commonly used for classification tasks. Lee et al. (2023) proposed the application of SVM joint to feature variable dimensional coordination to detect failures in TIM. The method uses Mahalanobis distance to extract and coordinate features from massive data, which reduces computational costs. Random forest (RF) is another method used in recent works (Liling et al., 2019; Roy et al., 2020; Tian et al., 2021). It comprises a collection of decision trees (DT) that typically employs a bootstrap-bagging strategy to aggregate individual predictions into a lower variance output (Liling et al., 2019). Roy et al. (2020) proposed detecting bearing faults using features extracted from autocorrelograms of vibration signals. They demonstrated that RF achieved the best accuracy compared to SVM, k-nearest neighbors (kNN), and naive Bayes (NB). Comparatively, Liling et al. (2019) combined RF and wavelet transform (WT) to detect inter-turn short circuits. The approach achieved high accuracy, even with few samples, and outperformed SVM, demonstrating the method's potential.

Boosting is another strategy used to train ensembles. Here, the base models work sequentially, with the adjusted models in the previous steps forming the basis for performing the next ones. This technique aims to apply gradient descent to minimize a loss function at each iteration. In the work of Tian et al. (2021), RF and extreme gradient boosting (XGBoost) are applied to diagnose air gap eccentricity and stator failure. The object of study was the traction motor of a high-speed train, and the dataset was obtained from a simulation platform. The proposed approach combines wavelet-packet decomposition (WPD) and principal component analysis (PCA) with stator current for feature extraction and reduction. They demonstrated that RF and XGBoost method increase detection accuracy compared to a conventional multilayer perceptron (MLP) and SVM. A similar comparison can be found in Wu et al. (2020), where XGBoost, SVM, and adaptive boosting were contrasted for diagnosing a fault in a wind turbine. It was clear that the XGBoost performed the best among the experimental models in this study. The study presented by Ehya et al. (2021) aims to investigate the feasibility of combining many intelligent models to perform a robust diagnosis of the inter-turn short circuit in rotor field windings of a synchronous machine. Various combinations of SVM, kNN, and XGBoost were experimented with others ML algorithms, working together as base classifiers in a stacking scheme. Then, the models' outcomes are aggregated by logistic regression as a meta-classifier into a unique model. The results showed that the stacking classifier often achieves better generalizations than the individual tools. However, it is also more complex and computationally expensive.

The work of Haroun et al. (2018) extracts features from stator currents using the Park transform, zero crossing time, and envelope. The strategy applies SVM and ReliefF to select relevant features. The authors have demonstrated that the self-organizing map correctly identifies winding failures using the proposed approach. The recent work by Kumar et al. (2021) effectively classifies ITSC using Park's vector and MLP. PCA was used for dimension reduction, which improved the predictor performance. Despite the success of ReliefF and other related algorithms in selecting important features, they only focus on improving the performance of ML models. These methods are not able to make the models interpretable and transparent to detect possible bias or improve end-user confidence. On the other hand, although PCA is a powerful tool to make the predictors more accurate, the causal relationship between the defect and the features is lost by creating new orthogonal variables. The present work aims to promote a clear association between the failure event and its indicators. Therefore, Shapley additive explanations (SHAP) are better suited than these groups of techniques to help with this task.

Frequently, sophisticated models can achieve high accuracy. However, it is not easy to understand what causes the model to make a particular prediction. On the other hand, high-bias ML techniques, such as DT and logistic regression, are comprehensive but lack predictive power (Aas et al., 2021). In Lundberg and Lee (2017), SHAP is introduced to address these inconveniences. It is a unified framework that increases the transparency of classifications and builds on the emerging concept of explainable artificial intelligence (XAI). In general, XAI techniques can effectively take advantage of systems based on computer intelligence without compromising classification performance. Interpretation of a model is essential to debugging it and extracting its full potential. A recent work of Gashi et al. (2022) compared SHAP with ten other XAI methods, including interpretML, explainX, Eli5, and interpretable model-agnostic explanations (LIME). The authors conclude that SHAP excels in computational resources and interactive exploration of model predictions and is the only XAI method with a solid theoretical foundation that ensures a uniform distribution of predictions across feature values for a local explanation. In addition, Lundberg and Lee (2017) proposed SHAP to unify six other XAI methods. Their work showed that SHAP improves computational performance and matches human intuition better than other approaches. For the aforementioned reasons, SHAP is used to interpret the predictions in this work.

Since SHAP is a model-agnostic technique, it analyzes only the input features and the output, so it does not require access to the ML structure (Brito et al., 2022). This means that SHAP can handle any prediction method (Aas et al., 2021). Recently, SHAP has been used to evaluate and interpret predictions in studies from different fields of knowledge. In Dikshit and Pradhan (2021), SHAP is applied to develop a robust drought prediction model. In this study, long short-term memory (LSTM) was employed to generate predictions and SHAP was used to analyze the relationships among variables and their effects on forecasting. The work of Veerappa et al. (2022) investigates the applicability of XAI methods in the maritime domain. A residual neural network (ResNet) is used to classify vessel types using positional information, speed, and other characteristics. The authors demonstrated that methods based on SHAP reveal the truthfulness of the explanations. In the biomedical field, Nohara et al. (2022) adopted SHAP to interpret gradient boost and promote cerebral infarct prediction using real hospital data. The approach of Santos et al. (2022) employed LSTM to extract dynamic features and classify events in power systems, while SHAP was applied to evaluate predictions and help users understand the model's decisions. The effectiveness of combining XGBoost and SHAP was also tested by Smith and Alvarez (2021) for COVID-19 mortality. Analysis of model predictions at local and global levels allowed the assessment of which variables positively or negatively affect the probability of mortality. In Brito et al. (2022), the detection and diagnosis of mechanical failures in rotating machinery are performed by applying vibration signals and different ML models. Then, to explain and compare the results, two XAI techniques are used, SHAP and Local Depth-based Feature Importance for the Isolation Forest (Local-DIFFI). The reports revealed more relevant features to identify an anomaly and indicated which ones are directly or indirectly associated with a specific type of failure. In the work of Zhang et al. (2022), XGBoost and SHAP were used to diagnose faults in oil-filled transformers. The SHAP values explained the model and the influence of attributes in each operating condition, improving the classification performance.

Although advances in TIM fault diagnosis, a robust and reliable identification system is still challenging (Gyftakis & Cardoso, 2020; Husari & Seshadrinath, 2022). The XGBoost training method enables all processor resources to grow learners in parallel, resulting in higher computational speed than neural approaches (Chen & Guestrin, 2016). On the other hand, RF is not a deterministic algorithm. As a result, it can be challenging to obtain coherent explanatory power since the model can produce different results from the same dataset. The accuracy of deterministic prediction is an advantage of XGBoost (Li et al., 2021), and it is a desirable property to support the comprehension of the predictions required in this study. XGBoost is a relatively recent

technique, and few studies have experimented with it in TIM fault diagnosis scenarios. Its iterative training focused on residual error promotes special attention to misclassified observations, which could be suitable for many applications with great prediction power (Chen & Guestrin, 2016). In this context, this work uses the XGBoost to detect ITSC stator windings and SHAP to support feature selection and explain the model's predictions.

Therefore, the main contributions of the study are listed:

- This work considers fault detection in the early stages, with 1% turns in the experiments. According to Garcia-Calva et al. (2022), the stator fault can be considered incipient before exceeding about 3% of damaged turns. Detecting ITSC in the early stages is one of the biggest challenges for TIM diagnostic systems (Gyftakis & Cardoso, 2020). ITSC evolves fast (Mejia-Barron et al., 2021), so the monitoring system must detect the fault as soon as possible (Rangel-Magdaleno, 2021).
- The dataset used considers several levels of supply voltage imbalance. This condition is considered highly problematic for ITSC diagnostics since its effects can be confused with the fault (Gangsar & Tiwari, 2020; Sonje et al., 2019). Moreover, this issue is widespread in industrial facilities (Adekitan & Abdulkareem, 2019).
- Six different loads are considered, including no-load and overload. Typical industrial processes usually require different torques (Mejia-Barron et al., 2021), so the diagnostic system must operate over the entire load range.
- Post-hoc interpretation of the classification results is performed using SHAP to explain the relationship and effects of the input features. According to Gangsar and Tiwari (2020), many studies do not address the description of the behavior of a failure indicator in different situations. Moreover, Zhang et al. (2022) states that research efforts rarely discuss the interpretability of the diagnostic model in selecting failure characteristics. Therefore, this work aims to uncover the correlation between failure and features. In addition, features from temporal, spectral, and wavelet approaches are investigated.
- This work implements a data-driven system that provides automatic fault classification. It is common to find works that propose new methods for fault detection but do not implement an intelligent detection system (Gangsar & Tiwari, 2020).

As many studies report, it is difficult to identify incipient ITSC under adverse conditions caused by the industrial environment, such as unbalanced voltage supply and different torque values. To overcome this problem, the methodology proposed in this work explores different approaches by extracting and combining features calculated in time, frequency, and time–frequency domains. Moreover, the method shows the correlation between the features and the failure, helping to describe how these problematic operating conditions affect the predictions.

2. Technical work preparation

This section presents a theoretical framework and details of data collection and methodology.

2.1. XGBoost

Basically, the XGBoost algorithm uses boosting to train multiple classification and regression trees (CART) (Wu et al., 2020). As described in Chen and Guestrin (2016), the method for estimating the result $\hat{y}_i$ from a database with n samples and m features is shown in (1).

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{1}$$

Here, K is the additive functions (total number of trees), x_i is the dataset of the ith input, f_k is a function corresponding to a particular structure q, and leaf weights w the set of all models is the functional space $F = \{f(x) = w_{q(x)}\}(q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T)$, q represents the structure of each tree, T is the number of leaves in the tree. Thus, the final prediction is calculated by summing up the w values in the leaves.

Additive training in XGBoost changes the objective function by fixing the previous prediction and adding a new tree at each step. Thus, the algorithm trains a base model and uses its prediction to update the training data so that the following model fits better. In each cycle, the additional base model focuses on the pseudo-residuals of the previous rounds and aims to minimize the objective function. The algorithm guides the structure to focus on misclassified instances, using the regularization term to control the complexity of the entire ensemble.

2.2. SHAP

SHAP is a method that assigns each attribute a value corresponding to its importance in a particular prediction. It helps to understand the classification process by quantifying the relevance of predictors. SHAP can be used for global and local explanations. The global interpretation describes which features have the greatest influence on the overall model. However, when the goal is to understand how the combination of input features influenced a particular prediction, it is called a local explanation (Aas et al., 2021).

In 2017, the study of Lundberg and Lee (2017) proposed the calculation of SHAP values ϕ_i using (2), where x' are simplified inputs mapped by $x = h_x(x')$, M is the number of simplified input features, f is the original prediction model, and $z' \in \{0,1\}^M$. Here, h_x maps 1 or 0 to the original input space, indicating if the input is included in the model or not, and $f_x(z') = f(h_x(z'))$.

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!}[f_x(z') - f_x(z' \setminus i)] \tag{2}$$

The strategy of additive feature attribution of SHAP values is shown in (3) and demonstrates local accuracy in approximating the original model for a single input x_i (Lundberg & Lee, 2017).

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i \tag{3}$$

The value of ϕ_0 corresponds to the expected model prediction, i.e., the base value $E[f(z)]$. This means that the SHAP values explain how and why a specific instance prediction $f(x)$ deviates from the base value (Aas et al., 2021). Then the SHAP values comprise a conditional expectation function of an ML model. That is, the SHAP values for each input feature sum the difference between the expected output $E[f(x)]$ and the output of the current model $f(x)$ for a given sample. In this way, it is possible to quantify how far a prediction for a given sample is from the baseline value and how the features interact to obtain this result. The greater the distance between $f(x)$ and $E[f(x)]$, the greater the degree of certainty that a given sample belongs to one category or the other.

2.3. Data collection and organization

For the development of this work, stator current signals were acquired from induction motors with different numbers of shorted turns, 5%, 3%, 1%, and 0% (no-fault). The percentage of shorted turns depends on the severity of the fault that the approach should detect. The lower the percentage, the more incipient the fault and the more difficult it is to detect. Since higher degrees of fault severely damage the coil, a TIM monitoring system must detect the fault early to avoid a critical situation. The method used to emulate ITSC in the laboratory was to intentionally damage the insulation at specific points on the coil and connect it to external taps. It is worth noting that the coils of all three

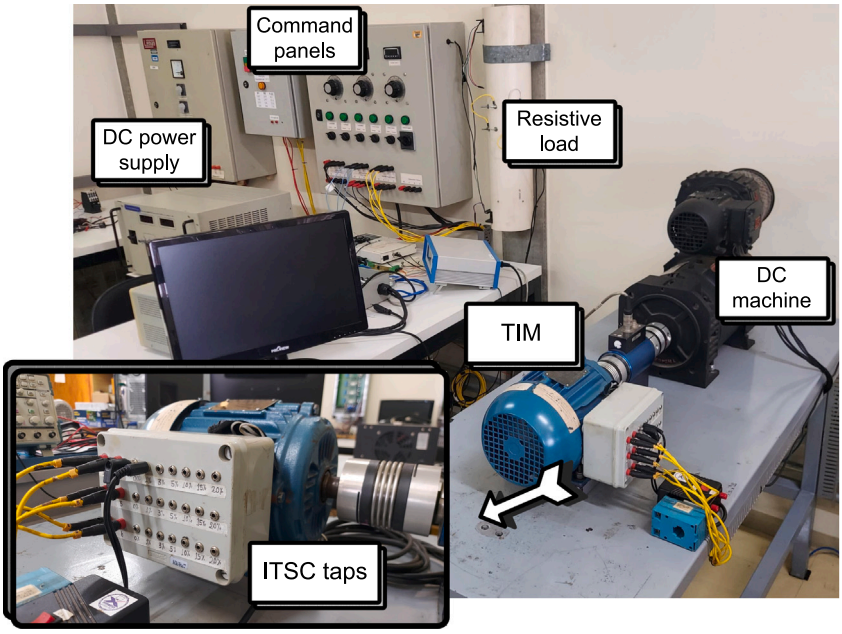

Fig. 1. General scheme for experiments.

Table 1
Number of damaged turns and corresponding severity level.

Severity level	TIM 1	TIM 2
1%	3 turns	8 turns
3%	9 turns	23 turns
5%	15 turns	39 turns

phases were rewound in this procedure. This ensures the symmetry of the assembly. This procedure was performed for two induction motors, TIM 1 and TIM 2. Both machines are four-pole, 60 Hz, squirrel-cage type. However, TIM 1 is a 1 HP, 4.06 N m, 1730 rpm, while TIM 2 is a 2 HP, 8.02 N m, 1750 rpm motor. The stators of the TIMs have six coils, with 306 turns per phase for TIM 1 and 780 turns for TIM 2. The fault severity corresponds to the percentage of turns that are shorted (Table 1), which is defined using the taps (Fig. 1). Therefore, this method does not require resistors to simulate different fault intensities.

Fig. 1 provides a comprehensive overview of the test bench. A direct current (DC) machine coupled to the induction motor is controlled by a DC power supply that allows variation of the resistive load. A resistor bank is connected to the armature to dissipate the electrical power generated. The TIM 1 database contains examples with mechanical loads ranging from $0.1T_n$ to $1.25T_n$, where T_n is the nominal torque. For TIM 2, the range is from $0.1T_n$ to $1.125T_n$. Three independent single-phase variable voltage transformers are used in the panels to generate the various combinations of voltage unbalance employed in this work. For all fault and load conditions, 60 samples of undervoltage and overvoltage were collected for the three phases as described in the diagram of Fig. 2. The undervoltage in phase A reached 10% of the nominal voltage (V_n) in $0.02V_n$ steps. Overvoltages and undervoltages were applied simultaneously in phases B and C, limited to ±8% of V_n. The signals have a duration of 1.5 s at a sampling frequency of 15.5 kHz. A set of Hall-effect current sensors and a DAQ6221 connected to a microcomputer via USB are responsible for measuring and storing the data. It is worth noting that each of the 60 samples considers a different operating condition, either due to load variations or imbalances.

2.4. Methodology

Fig. 3 shows the proposed methodology used in the fault diagnosis. The strategy explores the XGBoost classification potential using features computed in the temporal, spectral, and wavelet domains. The strategy in the time domain is to extract features from the raw signal. Consequently, no signal processing techniques and less computational effort are required. For this reason, this approach serves as the basis for comparison. In addition, since each procedure promotes analysis from a different perspective, using different analysis domains can help distinguish authentic ITSC-related patterns from those created by other operating conditions to which the TIM is exposed.

Table 2 exhibits the full list of features used for each approach. Details of each feature can be found in Barandas et al. (2020). In Approach 1, the twenty listed features were computed directly from the time series. On the other hand, Approach 2 performs the fast Fourier transform (FFT) of the current signal and splits the obtained spectrum into segments of 30 Hz, from 0 to 600 Hz (s_1 to s_{20}). Ten features were obtained from each spectral segment, for a total of over two hundred features. It should be noted that other segment sizes from 10 to 50 Hz were also tested, up to a limit of 1000 Hz, but this configuration proved more promising. Similarly, sixteen attributes were computed from the detail coefficients (cD) of the discrete wavelet transform compound of the wavelet feature set in Approach 3, yielding a total of 160 features. The Daubechies wavelet $db10$ is used to perform the decomposition up to the tenth level, cD_{10}, which is the maximum level of signal decomposition. In the preliminary analysis, other filter orders were also tested, but $db10$ was better for the predictions of this approach. The experiments were performed with the current of phase A, whose winding was subjected to failure.

Before classification, the input features must be preprocessed. Removing highly correlated features is important to obtain a consistent interpretation of SHAP. According to Aas et al. (2021), the explanation by SHAP can be misleading and unrealistic if the features are highly correlated. In this step of fault diagnosis, pairwise correlation is calculated using Pearson's method to determine the collinearity of pairs of features. This procedure can also minimize the complexity of the model.

The recursive feature elimination (RFE) procedure is employed to eliminate redundant features, which quantifies the importance of attributes using SHAP scores. The algorithm discards less relevant

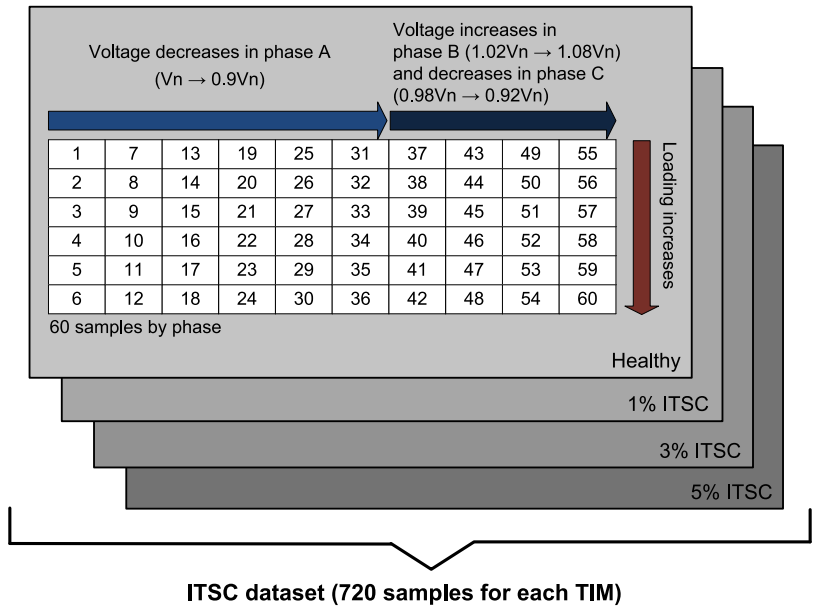

Fig. 2. Arrangement of the dataset.

Table 2
List of features grouped by domain.

Temporal domain (20 features): Absolute Energy (AbsEn), Area Under the Curve (ArUC), Autocorrelation (AutoCorr), Centroid (Ctd), Entropy (Ent), Kurtosis (Kurt), Interquartile Range (IQRange), Maximum (Max), Mean, Mean Abs. Deviation (MeanAD), Median, Median Absolute Deviation (MedAD), Minimum (Min), Number of Peaks (NPeaks), Peak to Peak Distance (PPD), Root Mean Square (RMS), Skewness (Skn), Standard Deviation (StD), Variance (Var), Zero Crossing Rate (ZCR)
Spectral domain (10 features from each spectral segment): Absolute Energy (AbsEn_s_x), Autocorrelation (AutoCorr_s_x), Centroid (Ctd_s_x), Entropy (Ent_s_x), Kurtosis (Kurt_s_x), Maximum (Max_s_x), Mean (Mean_s_x), Number of Peaks (NPeaks_s_x), Skewness (Skn_s_x), Standard Deviation (StD_s_x)
Wavelet domain (16 features from each detail coefficient): Absolute Energy (AbsEn_cD_x), Area Under the Curve (ArUC_cD_x), Autocorrelation (AutoCorr_cD_x), Entropy (Ent_cD_x), Interquartile Range (IQRange_cD_x), Kurtosis (Kurt_cD_x), Maximum (Max_cD_x), Mean (Mean_cD_x), Minimum (Min_cD_x), Number of Peaks (NPeaks_cD_x), Root Mean Square (RMS_cD_x), Skewness (Skn_cD_x), Standard Deviation (StD_cD_x), Variance (Var_cD_x), Zero Crossing Rate (ZCR_cD_x)

features in each cycle and XGBoost performs a new classification. The best feature set is selected based on its size and the model performance criterion: if one or more feature sets provide mean values for predictions with equivalent performance metrics, the smallest feature set is chosen.

In the experiments, a binary classification was developed for each fault severity. The diagnostic system indicates whether an observation is indexed to a healthy or faulty signal. As shown in Fig. 2, the dataset consists of 60 samples for each fault severity, plus 60 healthy samples. Therefore, there are 120 instances for each test. The operating condition in which each sample was collected is unique, so the intelligent model must learn the general fault patterns to classify them correctly.

Labeling was performed by assigning 0 (healthy) and 1 (failure) values to the instances. For this purpose, a target vector was created with the same size as the attribute dataset, presented to the ML models only during the training phase.

Splitting data into training and testing subsets is often used to evaluate the performance of supervised learning models and avoid overfitting. In this case, the training subsets consist of 75% of the data, i.e., 90 samples, while the other 30 samples have been reserved for testing. A validation method is also recommended to improve the reliability of the model. In this study, k-fold cross-validation is applied, i.e., the training dataset is divided into k subsets, and training/testing is performed by interchanging the subsets k-folds. Here, 10-fold cross-validation is applied, i.e., $k = 10$. The algorithm repeats this process ten times, shuffling the samples. The average values of the performance metrics and the respective standard deviations are then calculated. The k-fold procedure is a well-established method considered less biased than hold-out because it uses all available data for training and testing, and it is less variable than leave-one-out because it averages over multiple runs.

Accuracy, receiver operating characteristic (ROC), the area under the ROC curve (AUC), precision, recall, F1-score, kappa, and the confusion matrix are commonly used performance metrics for evaluating ML algorithms. AUC can summarize the information provided by the ROC curve, while F1-score is the harmonic mean of precision and recall. Therefore, AUC, F1-score, and kappa are used in this study because they provide comprehensive information about many other performance metrics and allow a concise and reliable assessment of the results. All possible values for these parameters are limited from 0 to 1.

Finally, tuning the ML parameters can also significantly affect classification results. Among other benefits, adjusting the parameters can help achieve more robust and accurate predictions, reduce the impact of outliers in the data, and minimize overfitting. The XGBoost hyperparameters are tuned for each experiment using grid search. It consists of testing multiple combinations of hyperparameters within specified limits. This makes it more likely to find the optimal hyperparameters

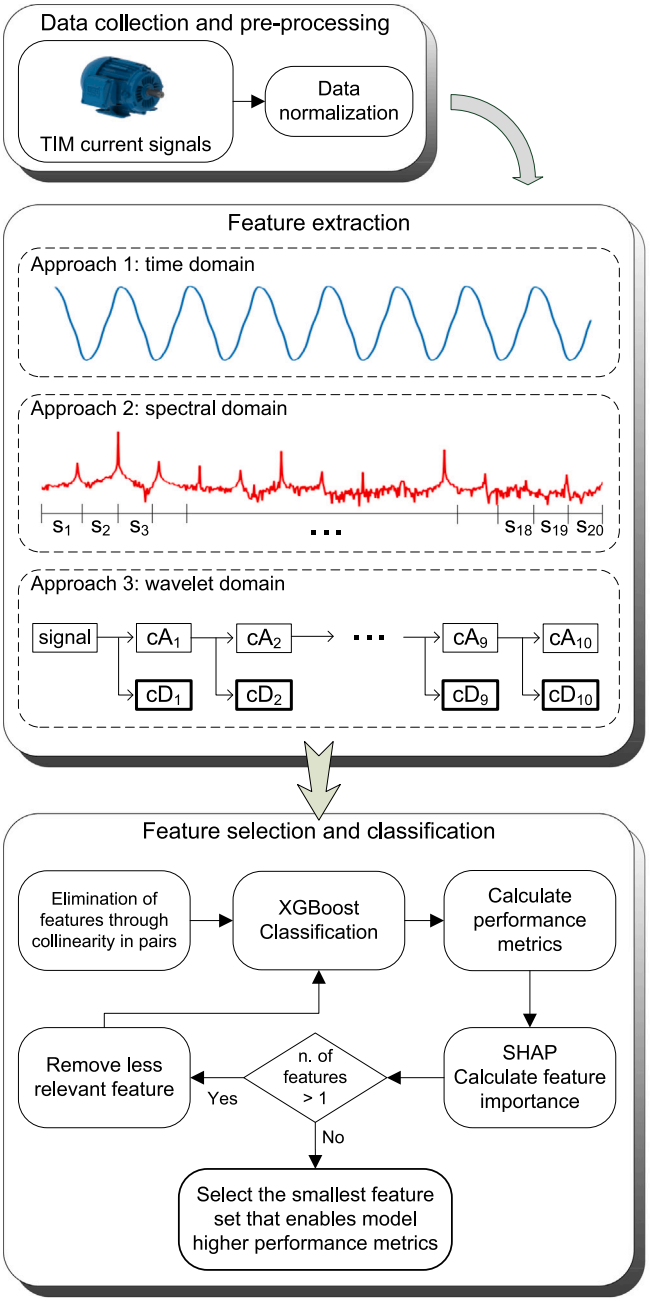

Fig. 3. The proposed methodology is divided into three stages: data collection and pre-processing, feature extraction, and, concomitantly, feature selection and fault diagnosis.

than some alternative methods, such as random search and Bayesian optimization. Some bounds have been set, as follows:

- The learning rate determines the step size at which the model updates other parameters for the next cycle. If it is too large, the model may not converge, and if it is too small, the model may get stuck in a suboptimal solution. So the values tested are 0.05, 0.1, 0.3 and 0.5;
- The estimators are the weak learners, or trees, that compound the ensemble. The number of estimators tested is 20, 50, 70, 100, 150, and 200;
- The maximum depth of a tree impacts the model complexity. Large values must be avoided to prevent overfitting, while smaller values can lead to underfitting. Thus, the values tested are 2, 3, 4 and 5;
- The Lagrangian multiplier, also called gamma, is a regularization parameter that prevents overfitting. However, the higher the gamma value, the more difficult for the model to find an optimal local variance/bias. Thus, the gamma values tested were 0, 0.1, 0.5, 1, 3, and 5.

Considering the combination of these parameters, 2304 possibilities were tested to achieve the optimal solution in each experiment.

Table 3
Detection of 5% ITSC in TIM 1, using temporal, spectral, and wavelet approaches.

Domain	Feature set	AUC	F1	kappa
Temporal	All	0.730 (0.102)	0.770 (0.113)	0.730 (0.102)
	Selected	0.818 (0.083)	0.820 (0.079)	0.746 (0.058)
Spectral	All	0.976 (0.051)	0.960 (0.049)	0.919 (0.101)
	Selected	0.983 (0.030)	0.978 (0.018)	0.972 (0.034)
Wavelet	All	0.980 (0.029)	0.979 (0.030)	0.959 (0.061)
	Selected	0.986 (0.023)	0.986 (0.023)	0.972 (0.047)

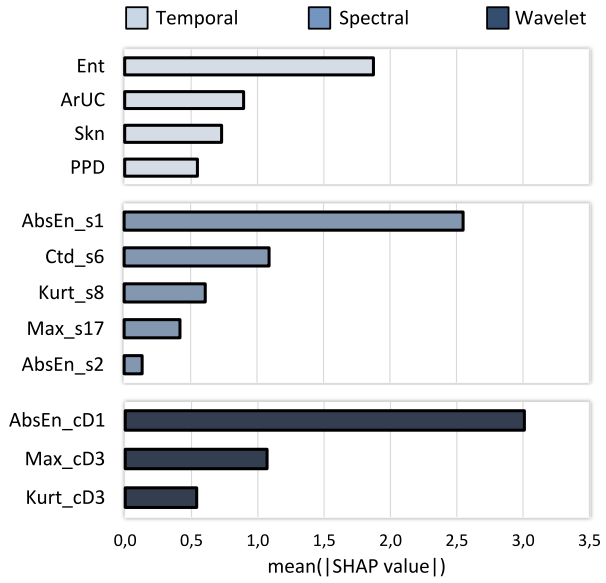

Fig. 4. Importance of features selected from temporal, spectral, and wavelet domains.

3. Fault diagnosis and interpretability

In the following experiments, the predictions are made using the TIM 1 dataset stratified by the fault severity. Table 3 demonstrates the average AUC, F1, and kappa values for 5% of ITSC detection. The standard deviation is given in parentheses. In all cases, the classification performance with the selected features is higher than the original feature set. In general, this is because there are redundant predictors in the initial stage, and the classification algorithm tries to fit all of them into the model. By removing the unnecessary attributes, the algorithm focuses on learning only the important patterns. The increment is small for spectral and wavelet approaches, but more substantial for temporal approaches. Fig. 4 shows the mean of the absolute SHAP values for the selected feature sets. It allows the interpretation of the model result by quantifying the importance of each predictor.

The metrics demonstrated that the spectral and wavelet approaches are equivalent for 5% ITSC. However, the performance observed for the attributes obtained from the signal in the time domain is far below. Moreover, the standard deviation is about 0.08, indicating a high dispersion of the results. Clearly, the different combinations of torque

Table 4
Detection of ITSC with spectral features.

ITSC	Data	AUC	F1	kappa
3%	TIM 1	0.860 (0.041)	0.861 (0.042)	0.719 (0.083)
	TIM 2	0.800 (0.074)	0.876 (0.070)	0.599 (0.148)
1%	TIM 1	0.819 (0.045)	0.822 (0.044)	0.641 (0.090)
	TIM 2	0.660 (0.072)	0.675 (0.077)	0.319 (0.146)

Table 5
Detection of ITSC with wavelet features.

ITSC	Data	AUC	F1	kappa
3%	TIM 1	0.847 (0.054)	0.846 (0.053)	0.692 (0.107)
	TIM 2	0.895 (0.051)	0.897 (0.058)	0.790 (0.102)
1%	TIM 1	0.781 (0.078)	0.777 (0.082)	0.562 (0.156)
	TIM 2	0.784 (0.087)	0.784 (0.085)	0.568 (0.176)

on the shaft and disturbances in the three-phase input voltages mask the current changes caused by the fault. Consequently, the features extracted from the raw signal needed to be more comprehensive to promote a distinction between categories in the classifications. Therefore, this strategy is discarded in the following experiments.

Tables 4 and 5 present the classification reports for 3% and 1% ITSC, using spectral and wavelet attributes, respectively. In addition, these classifications also use the TIM 2 dataset. Table 4 shows that AUC and F1 for 1% of short in TIM 2 reach 0.675, and kappa is 0.319. However, for other classifications, AUC and F1 remain above 0.8, with a standard deviation of about 0.04 for 3% ITSC and 0.07 for 1% ITSC. The prediction for 3% of fault in TIM 1 has the highest concordance rate, with a kappa value of 0.719. Fig. 5 presents insight into how the features interact to achieve these results. One can observe that kurtosis recurs in the predictions: $Kurt\_s_{14}$ and $Kurt\_s_{12}$ are important features for TIM 1, while $Kurt\_s_1$ and $Kurt\_s_{10}$ are important for the TIM 2 dataset. Other features that stand out are the absolute energy and the mean value. The recurrence of segments s_1, s_2, s_5, s_{14}, and s_{19} indicates a high relevance of these specific spectral intervals.

From Table 5, both AUC and F1 scores indicate that XGBoost learned more efficiently from the patterns in the TIM 2 dataset. However, the difference cannot be considered significant for failures in 1% of turns. The degree of agreement, expressed by the kappa coefficient, is higher for 3% ITSC, TIM 2, and smaller for 1% ITSC, TIM 1. Fig. 6 depicts the importance of the attribute for these predictions. It can be seen that $Kurt$, $Npeaks$, and $AbsEn$, related to the DWT coefficients cD_1 and cD_5, strongly influenced the TIM 1 classifications for both short circuit levels. For the 3% ITSC of the TIM 2 dataset, the attributes from coefficients cD_2 and cD_3 are the most important, and for 1% ITSC the relevance is spread over many DWT levels. In general, kurtosis and absolute energy are more common in all classifications, but with different levels of importance.

Analyzing the computational cost of an algorithm is essential to make fair comparisons. The asymptotic analysis expresses the upper bound of the complexity of an algorithm. Among the techniques used, FFT is associated with a complexity of $\mathcal{O}(p \log p)$, while DWT is associated with $\mathcal{O}(p)$, where p is the number of data points of the analyzed signal (Mörchen, 2003). Although the linear logarithm function is more complex than the linear one, neither of such algorithms can be considered particularly computationally intensive compared to other processing methods used in this context (e.g., MUSIC and ESPRIT have cubic complexity, $\mathcal{O}(p^3)$ Riera-Guasp et al., 2015).

The following experiment checks if a fusion of the most important features from spectral and wavelet approaches can better assist XGBoost in learning healthy and faulty signal patterns. For each case, the features shown in Figs. 5 and 6 are used jointly as input sets. Table 6 has shown that this strategy improves the average model performance for all tests performed compared to Tables 4 and 5, some to a greater, some to a lesser degree. The local interpretation provided by the SHAP values

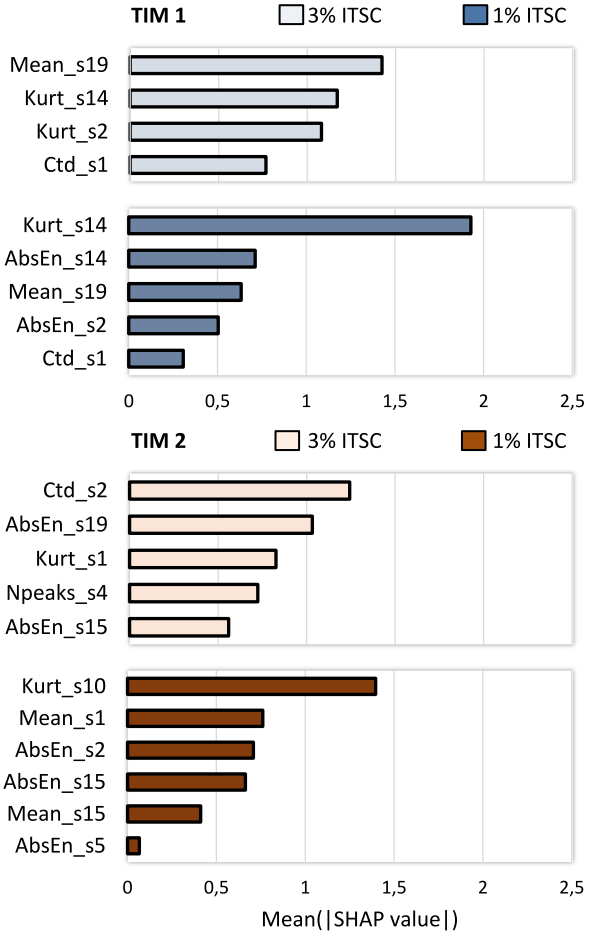

Fig. 5. Importance of selected features (spectral domain).

Table 6
Detection of different levels of ITSC with the fusion of features in spectral and wavelet domains.

ITSC	Data	AUC	F1	kappa
3%	TIM 1	0.863 (0.065)	0.871 (0.067)	0.705 (0.130)
	TIM 2	0.903 (0.031)	0.904 (0.030)	0.807 (0.061)
1%	TIM 1	0.845 (0.033)	0.837 (0.067)	0.660 (0.129)
	TIM 2	0.846 (0.052)	0.850 (0.044)	0.695 (0.103)

can help better understand how the intelligent model performs the presented diagnosis. This analysis considers the influence of features on the predictions for each sample in the test subset. Since the presentation of the local explanation is not appropriate for all experiments, the following section examines the classification for TIM 1 with 1% ITSC.

To compare the performance of XGBoost with other intelligent models, RF, SVM, and MLP scored AUC of 0.828, 0.841, and 0.749, respectively, on the TIM 1 dataset for 1% ITSC. Although SVM and RF achieved results close to XGBoost, their complexity can be estimated as $\mathcal{O}(m^3 n)$ for SVM (Geron, 2019) and $\mathcal{O}(Kmn^2 \log n)$ for RF (Louppe, 2014). On the other hand, the XGBoost algorithm requires fewer instructions to execute, which is maintained by the function $\mathcal{O}(rKd + r \log B)$, where r is the number of non-missing entries and B is the block length of the XGBoost structure (Chen & Guestrin, 2016).

3.1. Local interpretability

We have selected the most important features for the experiments based on the global interpretation of the model. The influences of the

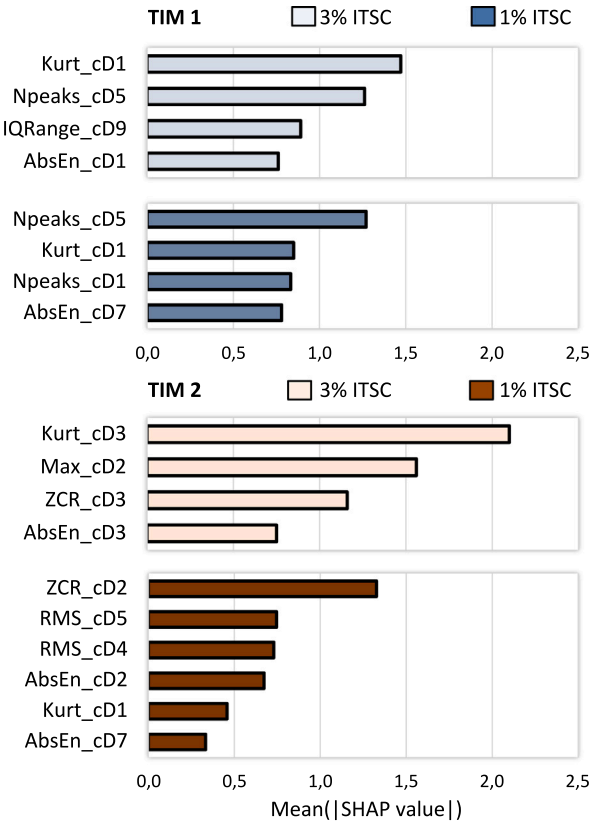

Fig. 6. Importance of selected features (wavelet domain).

features were measured by the mean of the SHAP absolute values of all the samples provided. On the other hand, this subsection focuses on understanding how the interaction of the features impacts each sample to build the final model.

The local interpretation allows the comparison of a result for a given sample with the global expected baseline model. Fig. 7 shows the local explanation for specific samples in the test subset. In each chart, starting from the global baseline, E[$f(X)$], a value SHAP is added for some variables and subtracted for others. In this way, the final prediction is formed. This baseline represents the expected value of the model output for all observations in that classification. For a single sample, the final observation value $f(x)$ indicates whether the model assigns this instance to one class or another: If the model considers the sample healthy, $f(x)$ <E[$f(X)$], otherwise as faulty, $f(x)$ >E[$f(X)$]. This means that positive SHAP values contribute to the fault indication for each feature and vice versa. And the further $f(x)$ remains away from the global baseline value, the more likely the prediction is correct.

In Fig. 7, sample 02, sample 06, and sample 34 are healthy, while sample 61, sample 120, and sample 110 are faulty. The observed $f(x)$ agrees with the labels, indicating that the model correctly classified these samples. In Fig. 7a, some features indicate the wrong class, and the corresponding SHAP values are low (SHAP value for Ctd_s_1 is +0.26 and for $AbsEn\_s_{14}$ is +0.06). Consequently, $f(x)$ for this observation is −4.570. In Fig. 7b, $f(x)$ reaches a value of 4.159. In both samples, the load on the shaft is small, and the three-phase voltage is balanced. In contrast, for the representations in Fig. 7c and d, one can notice that more features indicate high misclassification. For example, $Kurt\_s_{14}$ is an important predictor in the previous context (samples 02 and 06), but it pushes $f(x)$ into the wrong category with a considerable module value. Therefore, the $|f(x)|$ for these predictions is 2. Consequently, these classifications are correct but have been associated with lower confidence. Both samples correspond to an overload condition. A similar analysis for Fig. 7e and f suggest that voltage imbalance in phases B and C are more problematic for this approach than an imbalance in phase A.

It is interesting to note that all features sometimes shift the prediction into the incorrect category. Otherwise, a single feature would be enough for a perfect classification. Nevertheless, it can be seen that if one predictor contributes to a wrong classification, the other features must compensate. Since the dataset used in this work covers multiple loadings and unbalanced situations, the model has learned multiple patterns depending on the combination of features for a given situation. This approach can be employed to determine the feature or group of features that are more important for classification in each tested condition. However, some features that seem irrelevant to one sample may be the key to correctly classifying another sample. These insights, provided by local explanations, can increase the reliability of the diagnostic system since they enable one to understand how features interact to induce the model to classify a single sample as faulty or not.

3.2. Comparison with related works

Next, the performances presented in this article are compared with other methods in recent studies. The work of Haroun et al. (2018) reported a classification accuracy of 100%. However, the dataset used includes about 3.8% to 12.5% of the shorted turns. In addition, the authors found that the power supply voltage was not exactly balanced during the experiments, but no intentional phase imbalance was applied. In Kumar et al. (2021), the methodology considered up to 10% shorted turns and did not address voltage imbalances in the supply.

In contrast, the works of Alloui et al. (2022) and Drif and Cardoso (2014) consider voltage imbalance at certain levels in the experiments. However, no intelligent system was implemented, making it difficult to compare objectively.

Among works that used the same dataset as the present study, Bazan et al. (2017) and Palácios et al. (2017) reported 99% and 96% of accuracy, respectively. However, the experiments include up to 10% shorted turns and an unbalanced supply limited to 4% of nominal. Vitor et al. (2023) presented an accuracy of 100%, but the fault severity is 7%, and no voltage unbalances are considered.

Table 7 shows the summarized information about related works. It is worth noting that voltage unbalances are a major challenge for stator fault classification, and the present work covers up to 10% of phase unbalance levels. In addition, most of the work uses data from a single machine or equivalent machines of the same size. The present work demonstrates its robustness using experimental signals collected from two TIM of different sizes, with 1 HP and 2 HP, including overload conditions.

4. Conclusion

The objective of this work was to identify ITSC in TIM, which are exposed to typical industrial environments. To overcome the combination of power quality problems and different loads, the strategy is based on the extraction of several features from the stator current time series, the FFT spectrum and the DWT coefficients. XGBoost and SHAP were combined with the RFE algorithm to create a small set of meaningful attributes. The results show that this compact set of attributes provides equivalent or even higher performance metrics than the original set.

It was found that features collected in the time domain did not provide enough information to correctly classify incipient faults, and there was no evident advantage to using FFT or DWT features. However, there is an increase of the system performance when spectral and wavelet features were used simultaneously.

The local interpretation allowed us to understand why a single case receives its prediction. It has been shown that some features are important fault indicators when there is no voltage unbalance and light loads are applied. However, these same features are irrelevant in other situations or can disrupt correct prediction. It was concluded that the

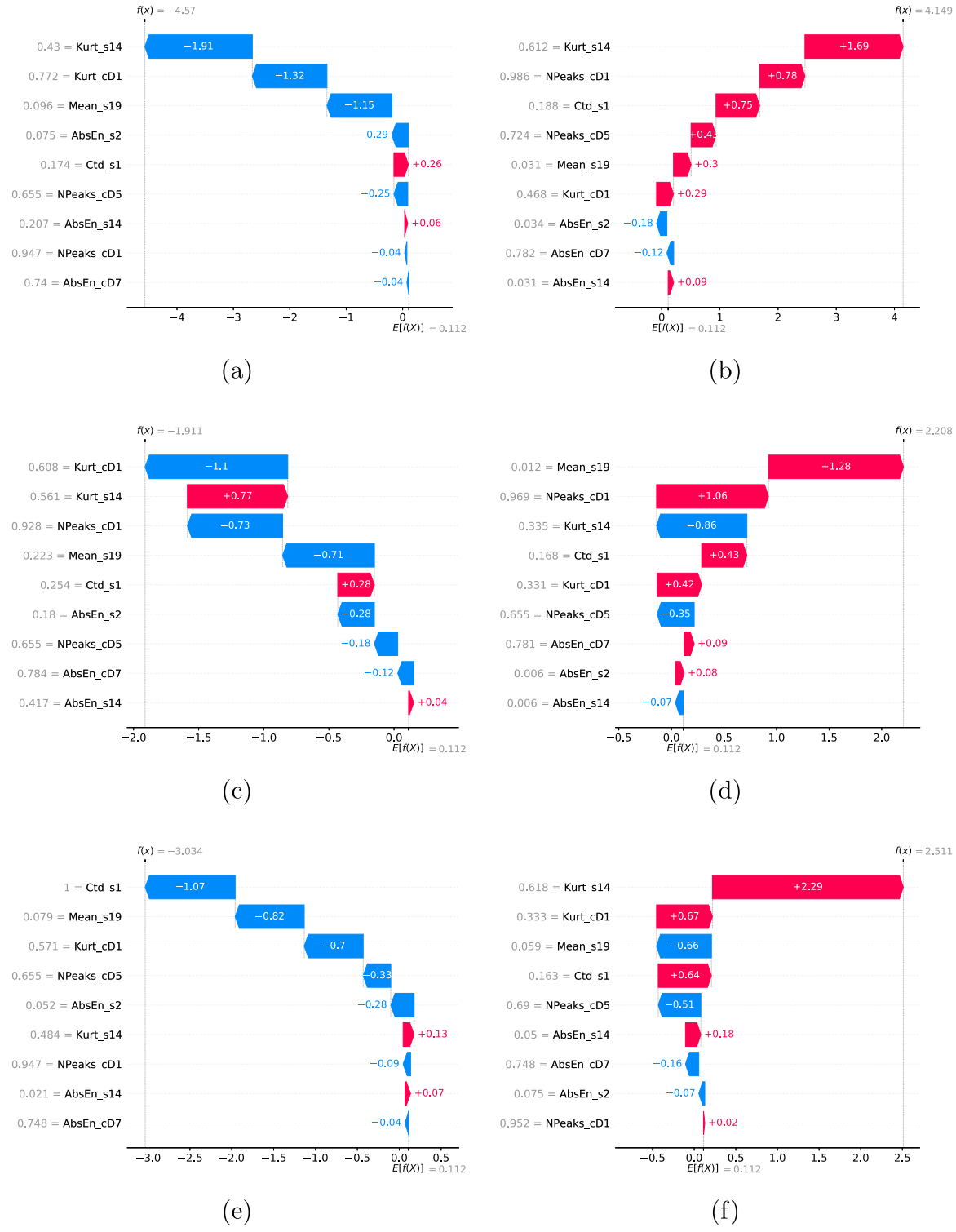

Fig. 7. Local interpretation of the classification for 1% of ITSC in TIM 1 with combined spectral and wavelet features. (a) Sample 02 ($0.25T_n$, $V_a = V_b = V_c = V_n$), (b) Sample 61 ($0.1T_n$, $V_a = V_b = V_c = V_n$), (c) Sample 06 ($1.25T_n$, $V_a = V_b = V_c = V_n$), (d) Sample 120 ($1.25T_n$, $V_a = V_n$, $V_b = 1.08V_n$, $V_c = 0.92V_n$), (e) Sample 34 ($0.75T_n$, $V_a = 1.1V_n$, $V_b = V_c = V_n$), (f) Sample 110 ($0.25T_n$, $V_a = V_n$, $V_b = 1.06V_n$, $V_c = 0.94V_n$).

Table 7
Related works for comparison.

Ref.	Loading levels	Unbalanced Supply[a]	ITSC severity	Success rate
Haroun et al. (2018)	No load to full load	No	3.8% to 12.5% (mixed)	100%
Kumar et al. (2021)	NS[c]	No	Up to 10% (mixed)	99%
Drif and Cardoso (2014)	No load to full load	Up to 1%	1% to 12%	NS[c]
Alloui et al. (2022)	No load to 30% of nominal	Up to 4%	1.5% and 3%	NS[c]
Bazan et al. (2017)[b]	No load to overload	Up to 4%	1%, 3%, 5%, and 10% (mixed)	93%–99%
Palácios et al. (2017)[b]	No load to overload	Up to 4%	1%, 3%, 5%, and 10% (mixed)	89%–96%
Vitor et al. (2023)[b]	No load to full load	No	7%	86%–100%
This work	No load to overload	Up to 10%	1%, 3%, and 5%	98% (5% ITSC) 90% (3% ITSC) 85% (1% ITSC)

[a]Intentionally applied.
[b]Same dataset as this work.
[c]NS: not specified.

features interact differently depending on TIM operating conditions and their respective contributions are unique to each case.

In this work, an effective method for detecting incipient stator winding short circuits under problematic operating conditions has been proposed, taking into account a wide range of voltage unbalances and mechanical load levels. Current signals from two different TIMs are used in the experiments to demonstrate the robustness of the method. While this study has demonstrated the feasibility of the method on direct-fed machines, further research could apply the proposed approach to diagnose TIMs controlled by frequency converters and investigate other TIM failures.

CRediT authorship contribution statement

Avyner L.O. Vitor: Conceptualization, Methodology, Software, Writing – original draft, Investigation, Visualization. **Alessandro Goedtel:** Supervision, Data curation, Writing – reviewing and editing, Funding acquisition. **Sylvio Barbon Junior:** Software, Supervision, Writing – original draft. **Gustavo H. Bazan:** Investigation, writing – reviewing and editing. **Marcelo F. Castoldi:** Supervision, Data curation, Writing – reviewing and editing, Funding acquisition. **Wesley A. Souza:** Software, Writing – reviewing and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential

Acknowledgments

This work was supported in part by the National Council for Technological and Scientific Development (CNPq), Brazil under Grant 474290/2008-5, Grant 473576/2011-2, Grant 552269/2011-5, Grant 201902/2015-0, and Grant 405228/2016-3; in part by the Coordination for the Improvement of Higher Level Personnel (CAPES), Brazil; in part by the Araucária Foundation for the Support of the Scientific and Technological Development of the State of Paraná under Grant 338/2012 and Grant 06/56093-3; and in part by the Federal University of Technology—Paraná.

References

Aas, K., Jullum, M., & Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, *298*, http://dx.doi.org/10.1016/j.artint.2021.103502.

Adekitan, A., & Abdulkareem, A. (2019). Prediction of the voltage status of a three-phase induction motor using data mining algorithms. *SN Applied Sciences*, http://dx.doi.org/10.1007/s42452-019-1720-9.

Alloui, A., Laadjal, K., Sahraoui, M., & Marques Cardoso, A. J. (2022). Online inter-turn short-circuit fault diagnosis in induction motors operating under unbalanced supply voltage and load variations, using the STLSP technique. *IEEE Transactions on Industrial Electronics*, 1. http://dx.doi.org/10.1109/TIE.2022.3172751.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., & Gamboa, H. (2020). TSFEL: Time series feature extraction library. *SoftwareX*, *11*, http://dx.doi.org/10.1016/j.softx.2020.100456.

Bazan, G. H., Scalassara, P. R., Endo, W., Goedtel, A., Godoy, W. F., & Palácios, R. H. C. (2017). Stator fault analysis of three-phase induction motors using information measures and artificial neural networks. *Electric Power Systems Research*, *143*, 347–356. http://dx.doi.org/10.1016/j.epsr.2016.09.031.

Brito, L. C., Susto, G. A., Brito, J. N., & Duarte, M. A. (2022). An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. *Mechanical Systems and Signal Processing*, *163*, http://dx.doi.org/10.1016/j.ymssp.2021.108105.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM international conference on knowledge discovery and data mining*. ACM, http://dx.doi.org/10.1145/2939672.2939785.

Dikshit, A., & Pradhan, B. (2021). Explainable AI in drought forecasting. *Machine Learning with Applications*, *6*, http://dx.doi.org/10.1016/j.mlwa.2021.100192.

Drif, M., & Cardoso, A. J. M. (2014). Stator fault diagnostics in squirrel cage three-phase induction motor drives using the instantaneous active and reactive power signature analyses. *IEEE Transactions on Industrial Informatics*, *10*(2), 1348–1360. http://dx.doi.org/10.1109/TII.2014.2307013.

Ehya, H., Skreien, T., & Nysveen, A. (2021). Intelligent data-driven diagnosis of incipient inter-turn short circuit fault in field winding of salient pole synchronous generators. *IEEE Transactions on Industrial Informatics*, 1. http://dx.doi.org/10.1109/TII.2021.3054674.

Gangsar, P., & Tiwari, R. (2020). Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. *Mechanical Systems and Signal Processing*, *144*, http://dx.doi.org/10.1016/j.ymssp.2020.106908.

Garcia-Calva, T., Morinigo-Sotelo, D., Fernandez-Cavero, V., & Romero-Troncoso, R. (2022). Early detection of faults in induction motors: A review. *Energies*, *15*(21), http://dx.doi.org/10.3390/en15217855.

Gashi, M., Vuković, M., Jekic, N., Thalmann, S., Holzinger, A., Jean-Quartier, C., & Jeanquartier, F. (2022). State-of-the-art explainability methods with focus on visual analytics showcased by glioma classification. *BioMedInformatics*, *2*(1), 139–158. http://dx.doi.org/10.3390/biomedinformatics2010009.

Geron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.

Gundewar, S., & Kane, P. (2020). Condition monitoring and fault diagnosis of induction motor. *Journal of Vibration Engineering and Technologies*, *2020*, http://dx.doi.org/10.1007/s42417-020-00253-y.

Gyftakis, K., & Cardoso, A. (2020). Reliable detection of stator inter-turn faults of very low severity level in induction motors. *IEEE Transactions on Industrial Electronics*, *68*(4), 3475–3484. http://dx.doi.org/10.1109/TIE.2020.2978710.

Haroun, S., Seghir, A. N., & Touati, S. (2018). Multiple features extraction and selection for detection and classification of stator winding faults. *IET Electric Power Applications*, *12*(3), 339–346. http://dx.doi.org/10.1049/iet-epa.2017.0457.

Husari, F., & Seshadrinath, J. (2022). Incipient interturn fault detection and severity evaluation in electric drive system using hybrid HCNN-SVM based model. *IEEE Transactions on Industrial Informatics*, *18*(3), 1823–1832. http://dx.doi.org/10.1109/TII.2021.3067321.

Kumar, R. R., Cirrincione, G., Cirrincione, M., Tortella, A., & Andriollo, M. (2021). Induction machine fault detection and classification using non-parametric, statistical-frequency features and shallow neural networks. *IEEE Transactions on Energy Conversion*, *36*(2), 1070–1080. http://dx.doi.org/10.1109/TEC.2020.3032532.

Lee, Y., Park, B., Jo, M., Lee, J., & Lee, C. (2023). A quantitative diagnostic method of feature coordination for machine learning model with massive data from rotary machine. *Expert Systems with Applications*, *214*, Article 119117. http://dx.doi.org/10.1016/j.eswa.2022.119117.

Li, X., Ma, L., Chena, P., Xu, H., Xing, Q., Yan, J., Lua, S., Fan, H., Yanga, L., & Cheng, Y. (2021). Probabilistic solar irradiance forecasting based on XGBoost. *Energy Reports*, *8*, http://dx.doi.org/10.1016/j.egyr.2022.02.251.

Liling, S., Kuankuo, Z., & Xiangdong, L. (2019). Stator inter-turn fault diagnosis of induction motor based on wavelet packet decomposition and random forest. In *8th renewable power generation conference* (pp. 1–6). http://dx.doi.org/10.1049/cp.2019.0667.

Louppe, G. (2014). *Understanding random forests: from theory to practice* (Ph.D. thesis), France: University of Liège.

Lundberg, S., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. Cornell University, arXiv:1705.07874.

Mejia-Barron, A., Tapia-Tinoco, G., Razo-Hernandez, J. R., Valtierra-Rodriguez, M., & Granados-Lieberman, D. (2021). A neural network-based model for MCSA of inter-turn short-circuit faults in induction motors and its power hardware in the loop simulation. *Computers & Electrical Engineering*, (93), http://dx.doi.org/10.1016/j.compeleceng.2021.107234.

Merizalde, Y., Hernández-Callejo, L., & Duque-Perez, O. (2017). State of the art and trends in the monitoring, detection and diagnosis of failures in electric induction motors. *Energies*, *10*, http://dx.doi.org/10.3390/en10071056.

Mörchen, F. (2003). *Time series feature extraction for data mining using DWT and DFT*. Philipps University.

Nohara, Y., Matsumoto, K., Soejima, H., & Nakashima, N. (2022). Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Computer Methods and Programs in Biomedicine*, *214*, http://dx.doi.org/10.1016/j.cmpb.2021.106584.

Palácios, R. H. C., Godoy, W. F., Goedtel, A., da Silva, I. N., Moríñigo-Sotelo, D., & Duque-Perez, O. (2017). Time domain diagnosis of multiple faults in three phase induction motors using inteligent approaches. In *2017 IEEE 11th international symposium on diagnostics for electrical machines, power electronics and drives* SDEMPED, (pp. 85–89). http://dx.doi.org/10.1109/DEMPED.2017.8062338.

Rangel-Magdaleno, J. (2021). Induction machines fault detection: An overview. *IEEE Instrumentation & Measurement Magazines*, *24*(7), 63–71. http://dx.doi.org/10.1109/MIM.2021.9549228.

Riera-Guasp, M., Antonino-Daviu, J. A., & Capolino, G.-A. (2015). Advances in electrical machine, power electronic, and drive condition monitoring and fault detection: State of the art. *IEEE Transactions on Industrial Electronics*, *62*(3), 1746–1759. http://dx.doi.org/10.1109/TIE.2014.2375853.

Roy, S. S., Dey, S., & Chatterjee, S. (2020). Autocorrelation aided random forest classifier-based bearing fault detection framework. *IEEE Sensors Journal*, *20*(18), 10792–10800. http://dx.doi.org/10.1109/JSEN.2020.2995109.

Santos, O. L. D., Dotta, D., Wang, M., Chow, J. H., & Decker, I. C. (2022). Performance analysis of a DNN classifier for power system events using an interpretability method. *International Journal of Electrical Power and Energy Systems*, *136*, http://dx.doi.org/10.1016/j.ijepes.2021.107594.

Smith, M., & Alvarez, F. (2021). A machine learning research template for binary classification problems and shapley values integration. *Software Impacts*, *8*, http://dx.doi.org/10.1016/j.simpa.2021.100074.

Sonje, D., Kundu, P., & Chowdhury, A. (2019). A novel approach for sensitive inter-turn fault detection in induction motor under various operating conditions. *Arabian Journal for Science and Engineering*, 6887–6900.

Tian, R., Chen, F., & Dong, S. (2021). Compound fault diagnosis of stator interturn short circuit and air gap eccentricity based on random forest and XGBoost. *Mathematical Problems in Engineering*, *2021*, http://dx.doi.org/10.1155/2021/2149048.

Veerappa, M., Anneken, M., Burkart, N., & Huber, M. F. (2022). Validation of XAI explanations for multivariate time series classification in the maritime domain. *Journal of Computational Science*, *58*, http://dx.doi.org/10.1016/j.jocs.2021.101539.

Vitor, A., Scalassara, P., Goedtel, A., & Endo, W. (2023). Patterns based on clarke and park transforms of wavelet coefficients for classification of electrical machine faults. *Journal of Control, Automation and Electrical Systems*, *34*, 230–245. http://dx.doi.org/10.1007/s40313-022-00946-7.

Wu, Z., Wang, X., & Jiang, B. (2020). Fault diagnosis for wind turbines based on relieff and extreme gradient boosting. *Applied Sciences*, *10*, http://dx.doi.org/10.3390/app10093258.

Zhang, D., Li, C., Shahidehpour, M., Wu, Q., Zhou, B., Zhang, C., & Huang, W. (2022). A bi-level machine learning method for fault diagnosis of oil-immersed transformers with feature explainability. *International Journal of Electrical Power & Energy Systems*, *134*, Article 107356. http://dx.doi.org/10.1016/j.ijepes.2021.107356.