machines

MDPI

Article

Machine Learning-Based Detection of Inter-Turn Short Circuits in the Stator Windings of 220 V, Three-Phase Induction Motor

Sibusiso Gule, Elsie Fezeka Swana * and Lutendo Muremi

Department of Electrical and Electronic Engineering Technology, University of Johannesburg, Doornfontein Campus, Johannesburg 2028, South Africa; 201702971@student.uj.ac.za (S.G.); lmuremi@uj.ac.za (L.M.)
* Correspondence: eswana@uj.ac.za

Abstract

Induction motors play a vital role in industrial operations; however, stator inter-turn short-circuits faults remain a common and critical source of failure. This paper presents a machine learning-based diagnostic approach for detecting stator inter-turn short-circuit faults in three-phase induction motors operating at 50 Hz. Secondary data from a controlled test bench were processed using Power Spectral Density to determine energy distribution and guide the design of a Butterworth bandpass filter (20–350 Hz). The filtered signals were then analyzed using the Hilbert Transform to extract statistical features, which were ranked using the Minimum Redundancy Maximum Relevance algorithm to identify the most discriminative parameters. Two supervised classifiers, Support Vector Machine and Random Forest, were developed and validated using MATLAB's Classification learner app with 5-fold cross-validation. The Support Vector Machine achieved an accuracy of 94.19%, while the Random Forest model achieved 99.51% with macro and F1-scores of 0.9951 and near-perfect area under the curve values. The results confirm that the Random Forest classifier provides superior generalization, sensitivity, and robustness in fault detection compared to Support Vector Machine. This study successfully demonstrates that combining Power Spectral Density, Hilbert Transform, Minimum Redundancy Maximum Relevance, and ensemble learning yields a highly effective framework for predictive maintenance and reliable fault diagnosis in industrial motor applications.

Keywords: induction motor; inter-turn short-circuit fault; power spectral density; Hilbert transform; support vector machine; random forest; predictive maintenance; fault diagnosis

Academic Editors: Israel Zamudio-Ramirez and Jose Alfonso Antonino-Daviu

Received: 23 June 2026
Revised: 17 August 2026
Accepted: 22 August 2026
Published: 15 September 2026

1. Introduction

Induction motors are the cornerstone of industrial automation, driving pumps, conveyors, compressors, and fans across the mining, manufacturing, and process industries. Their popularity stems from their rugged design, low cost, and high reliability, with estimates suggesting that nearly 90% of industrial mechanical drives rely on three-phase induction motors [1,2]. However, despite their robustness, these machines are vulnerable to electrical and mechanical degradation over time. Among all electrical faults, stator-winding inter-turn short-circuit (ITSC) faults are particularly destructive, accounting for 35–40% of motor failures [3,4].

An ITSC occurs when insulation between adjacent turns in a stator coil deteriorates, forming a shorted loop through which excessive current circulates. Several types of stator-winding faults can occur, including turn-to-turn, coil-to-coil, phase-to-phase, coil-to-ground, and open-circuit faults due to winding breaks. Figure 1 depicts these faults, which all originate from insulation breakdown within the stator windings [5].

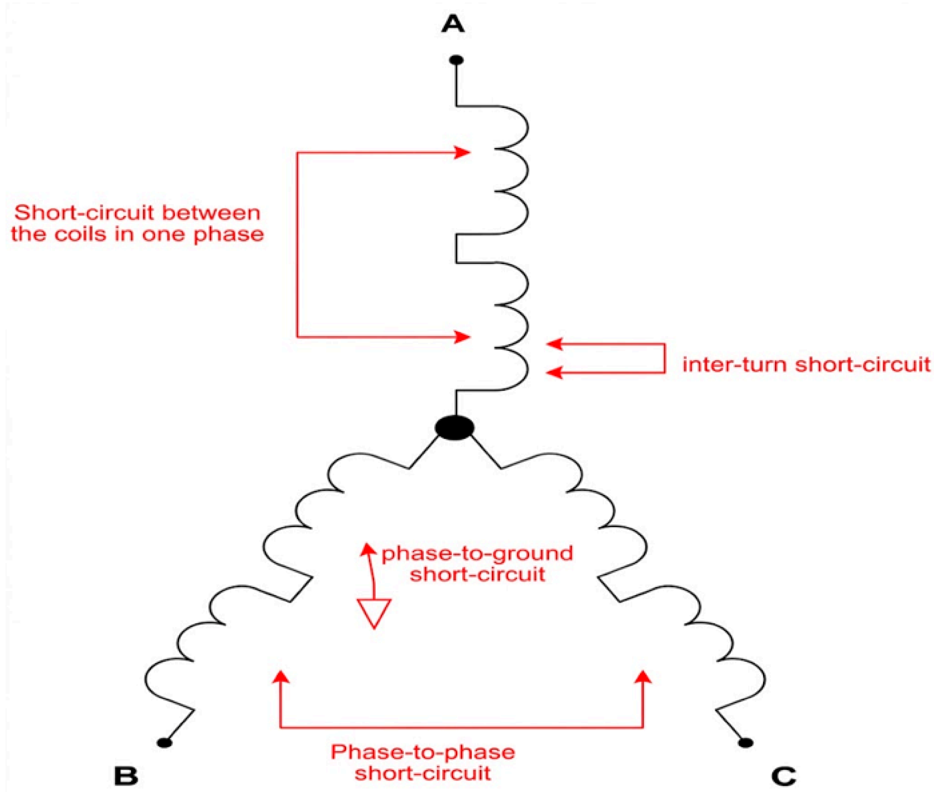

Figure 1. Different types of stator winding faults in a star configuration.

This local heating accelerates insulation aging and distorts the magnetic field, leading to torque imbalance and, if undetected, complete stator burnout [6]. The early detection of ITSCs is therefore critical to prevent extended downtime, high repair costs, and potential safety hazards. In continuous-operation sectors such as coal mining, delayed fault detection can halt production and cause significant financial losses [7]. This motivates the development of robust diagnostic techniques capable of identifying incipient ITSCs before they evolve into severe failures.

In response to these limitations, recent research has shifted towards advanced signal-processing and machine-learning approaches for motor-fault detection. Envelope-based features derived from the Hilbert Transform have been applied to stator-current signals to capture subtle amplitude modulations caused by winding asymmetry [8,9].

Likewise, supervised machine-learning algorithms such as Support Vector Machines (SVMs) and Random Forests (RFs) have demonstrated high accuracy and robustness in identifying motor faults using current-derived features under experimental conditions [10–12]. Comparative studies further show that ensemble methods like RF exhibit superior generalization and noise immunity, while kernel-based approaches such as SVM maintain strong decision-boundary precision in high-dimensional feature spaces [13].

Nonetheless, there remains a need for diagnostic frameworks that combine robust feature-extraction techniques, effective feature-selection algorithms, and comparative classifier evaluation under variable load conditions [14–16]. The present study addresses these gaps by developing a Hilbert-Transform-based, Minimum Redundancy Maximum Relevance (mRMR)-optimized machine-learning framework for stator ITSC detection in three-phase induction motors operating at 50 Hz.

https://doi.org/10.3390/machines14091046

2. Background

Conventional fault-diagnosis techniques such as vibration analysis, infrared thermography, and motor-current signature analysis (MCSA) have been widely used for decades [17–19]. Although MCSA is non-intrusive and cost-effective, it usually depends on Fast Fourier Transform (FFT) analysis to identify characteristic frequency components related to specific faults. The FFT assumes stationary signals, but induction-motor currents often vary with load and supply conditions, causing spectral smearing that obscures weak fault harmonics [20–22]. Consequently, FFT-based MCSA struggles to detect low-severity ITSCs where only a few turns are shorted.

Vibration and thermal-imaging techniques are also inadequate for early fault detection because the mechanical or thermal symptoms of ITSCs appear only at advanced stages. These limitations underscore the need for a more sensitive signal-processing approach that can extract subtle features directly from current and flux signals, without relying on the FFT [23,24].

In this project, Hilbert Transform is employed instead of FFT as the analytical tool for extracting instantaneous amplitude and phase information from stator-current signals. The Hilbert Transform yields the signal envelope, which captures low-amplitude modulations caused by winding asymmetry. From this envelope, a set of statistical time-domain features, including mean, RMS, variance, skewness, kurtosis, energy, and crest factor, are computed to quantitatively describe current variations under different fault conditions.

Machine-learning (ML) algorithms have recently emerged as powerful tools for data-driven fault detection because they can learn complex, non-linear relationships directly from measured data [25–27]. Unlike physics-based analytical models that require detailed motor parameters, ML techniques can adapt to noisy, real-world signals and varying operating conditions.

Previous studies have successfully applied Artificial Neural Networks (ANNs), k-Nearest Neighbors (kNN), Support Vector Machines (SVMs), and ensemble methods such as Random Forests (RFs) to motor-fault classification [7,28–30]. Deep-learning architecture, such as CNNs, often achieve high accuracy but require large datasets and heavy computation, which are impractical for many maintenance teams. In contrast, SVM and Random Forest classifiers provide strong accuracy with modest data requirements and excellent generalization [28,30]. The summary of the reviewed methods has been presented in Table 1 [17–22,28–30].

Table 1. Summary of reviewed methods.

Method	Advantage	Limitation	Reported Performance
MCSA with FFT	Non-intrusive, low cost, widely adopted in industry.	Assumes stationary signals; suffers from spectral smearing under varying loads; struggles to detect low-severity ITSCs.	Effective for mature faults; poor sensitivity for early or low-impedance ITSCs under dynamic load conditions.
Deep Learning	High classification accuracy without manual feature engineering.	High computational complexity; requires massive training datasets; impractical for edge deployment or field maintenance teams.	High theoretical accuracy but hindered by heavy computational overhead and resource demands.

In this project, all modeling is performed using MATLAB's Classification Learner App R2025a, which enables the structured training, validation, and comparison of supervised models. Two algorithms, Support Vector Machine (SVM) with a Quadratic kernel and Random Forest (Bagged Trees), are intentionally chosen for comparison because of their complementary properties. SVM excels at handling non-linear boundaries in high-dimensional feature spaces, while RF ensembles combine multiple decision trees to improve robustness and interpretability [31].

To further optimize classification performance, the Minimum Redundancy Maximum Relevance (mRMR) feature-selection algorithm is used before training. mRMR identifies the subset of features that carry the most diagnostic information while minimizing redundancy, thereby improving model accuracy and computational efficiency. This integration of Hilbert-based feature extraction, mRMR optimization, and comparative SVM–RF classification forms a comprehensive framework for reliable ITSC detection under real industrial conditions.

Despite significant progress in intelligent motor diagnostics, three gaps remain evident in the published literature.

Limited experimental validation under variable loads: Many studies rely on simulation data or tests conducted at a single load condition. In reality, motor-load variation significantly affects current signatures, and algorithms trained on fixed conditions often fail when generalized to new loads [32].
Lack of comparative evaluation of diverse algorithms: Most previous research focuses on a single model—typically SVM or ANN—without benchmarking ensemble methods like RF on the same dataset. There is limited comparative evidence of how SVM and RF perform under identical experimental and preprocessing settings for low-impedance ITSCs [33].
Underutilization of hybrid signal processing: Few works combine Power Spectral Density (PSD), Butterworth band-pass filtering, and Hilbert Transform in a single pipeline. PSD analysis can reveal the frequency range where ITSC-related sideband components appear; this information can then be used to design a Butterworth band-pass filter that isolates those fault-sensitive frequencies before envelope extraction. Additionally, studies seldom integrate mRMR feature selection to systematically reduce feature redundancy [34].

To address these shortcomings, this research proposes a data-driven, experimentally validated diagnostic framework that leverages PSD-guided band-pass filtering, Hilbert-based statistical feature extraction, and mRMR-optimized SVM and RF classifiers.

The primary aim of this research is to develop and experimentally validate a machine-learning-based framework for detecting and classifying inter-turn short-circuit (ITSC) faults in three-phase induction motors using Power Spectral Density (PSD) analysis, Hilbert Transform-derived statistical features, and comparative SVM and Random Forest classifiers implemented in MATLAB.

The specific objectives are:

- To analyze the stator-current signals using Power Spectral Density (PSD) to identify frequency bands containing ITSC-related components.
- To design a Butterworth band-pass filter based on the PSD results to isolate the fault-sensitive frequency range and remove noise and supply harmonics.
- To apply the Hilbert–Huang Transform to the filtered current signals to obtain the analytical signal and extract its envelope.
- To compute time-domain statistical features such as mean, RMS, variance, skewness, kurtosis, energy, and crest factor from the Hilbert envelope to represent the fault characteristics.
- To perform feature selection using the Minimum Redundancy Maximum Relevance (mRMR) algorithm to retain the most discriminative features.
- To train and evaluate two classifiers, namely SVM (Quadratic kernel) and Random Forest (Bagged Trees), using MATLAB's Classification Learner App, and to compare their performance using accuracy, precision, recall, and computational time.

This study presents an integrated, experimentally validated methodology for the early detection of stator-winding inter-turn short-circuit faults. The key contributions are as follows:

- PSD-guided signal preprocessing introduces the use of Power Spectral Density to identify fault-sensitive frequency bands, enabling the design of an effective Butterworth band-pass filter before Hilbert analysis.
- Hilbert-based envelope and feature extraction replaces FFT-based methods with Hilbert Transform to capture instantaneous amplitude modulation directly from time-domain signals, providing richer diagnostic information.
- Statistical-feature optimization extracts multiple time-domain descriptors from the Hilbert envelope and applies mRMR selection to minimize redundancy while maximizing fault sensitivity.
- Comparative machine-learning evaluation implements and compares SVM (Quadratic kernel) and Random Forest (Bagged Trees) classifiers using MATLAB's Classification Learner App R2025a on the same dataset, revealing their respective advantages in accuracy and interpretability.
- Experimental validation across load conditions validates the proposed approach on real measured data under varying loads and fault severities, bridging the gap between simulation and real industrial environments.
- Practical impact demonstrates a low-cost, scalable, and non-intrusive diagnostic technique suitable for integration into existing predictive-maintenance systems, particularly in mining and manufacturing sectors where motor reliability is vital.

In essence, this project advances induction-motor fault diagnosis by combining PSD-driven filtering, Hilbert-based feature extraction, and machine learning classification within a single, coherent framework. The approach enhances early fault detectability and supports the transition toward intelligent, condition-based maintenance in industrial systems.

3. Materials and Methods

3.1. Overview of Methodology

The overall methodology adopted in this study integrates experimental data acquisition, signal processing, feature extraction, feature selection, and supervised machine-learning classification. The workflow is shown conceptually in Figure 2.

- Stator-current and flux signals are acquired from a laboratory test bench at 10 kHz sampling frequency for 5 s under both healthy and faulty conditions.
- Power Spectral Density (PSD) is used to identify frequency bands containing ITSC-related components.
- Based on PSD results, a Butterworth band-pass filter is designed to isolate the dominant fault frequencies.
- The filtered signals are processed using Hilbert Transform to obtain their analytical envelopes.
- Thirteen (13) time-domain statistical features are extracted from the Hilbert envelope.
- Minimum Redundancy Maximum Relevance (mRMR) feature selection is applied to identify the most discriminative features.
- Selected features are used to train and evaluate Support Vector Machine (SVM) and Random Forest (RF) classifiers in MATLAB's Classification Learner App.
- The classifiers' performance is compared using accuracy, precision, recall, F1-score, and computational time.

https://doi.org/10.3390/machines14091046

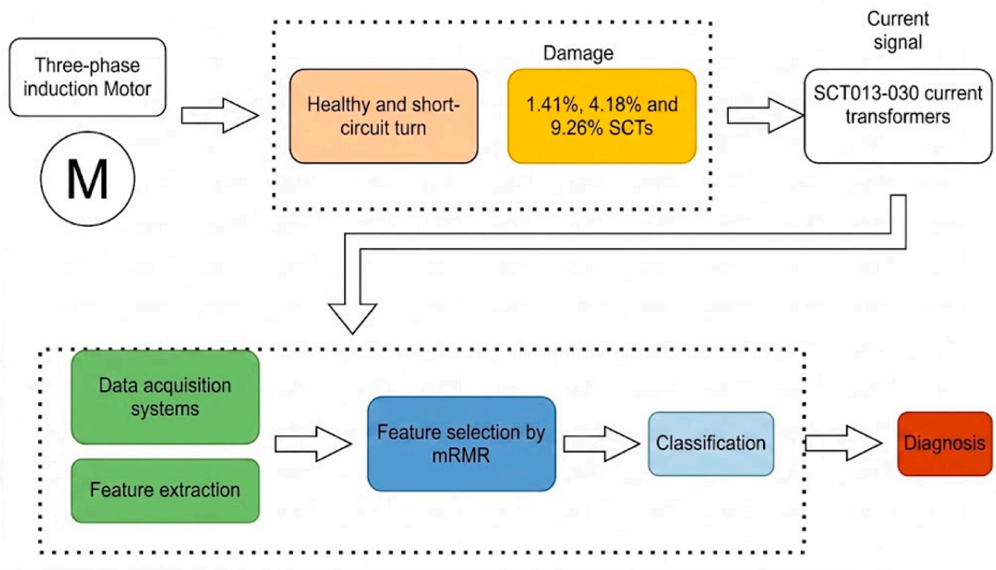

Figure 2. Methodology overview.

3.2. Experimental Setup

This study uses a secondary, publicly available dataset collected on a controlled laboratory test bench and released by R. Cunha on Kaggle (Inter-turn Short-Circuit in Induction Motor) [15]. The original bench enables precise emulation of stator inter-turn short-circuit (ITSC) conditions at multiple severities and under varying loads, with synchronized acquisition of stator currents and axial leakage flux. In the present work, only the subset recorded at a fixed electrical frequency of 50 Hz is considered, focusing on low-impedance (LI) ITSCs with three severities, namely: LI-1: 1.41%, LI-2: 4.81%, and LI-3: 9.26% of the stator winding and the healthy condition. Where LI-1 presents five shorted turns, LI-2 presents 17 shorted turns and LI-3 presents 32 shorted turns. The fault severity can be determined by Equation (1):

$$S_{ITSC} = \frac{N_{sc}}{N_{turn}} \times 100\% \quad (1)$$

where S_{ITSC} is the inter-turn short-circuit fault severity, N_{sc} is the number of shorted windings, and N_{turn} is the number of turns.

The data were generated on a twin-machine rig comprising two identical, three-phase squirrel-cage induction motors coupled via a shaft. One machine operated as the drive motor and the second as a controllable mechanical load. Both machines were supplied through WEG CFW-08 frequency converters, enabling precise control of the operating point as presented in Table 2 [35]. The drive motor stator was rewound to expose coil taps, allowing for insertion of inter-turn short circuits through a dedicated fault-emulation panel.

Two sensing modalities were acquired in the source bench:

- Three-phase stator currents, measured by SCT013-030 current transformers (CTs).
- Axial leakage flux, measured by a 100-turn coil wound with 24 AWG copper around the motor shaft.

Signals were routed through analog anti-aliasing and conditioning stages and digitized by a data acquisition (DAQ) system.

To monitor electrical and magnetic behavior, current transformers measured the stator currents in all three phases, while a 100-turn, 24 AWG copper coil wound around the motor shaft measured the axial leakage flux. The analog signals were filtered and digitized using a 16-bit data acquisition card at a sampling frequency of 10 kHz for 5 s per measurement, producing 50 k samples per channel. The sampling frequency was based on the Nyquist Theorem and to ensure that all the fault harmonic attributes were captured without distortion.

Table 2. Twin-machine bench and motor specifications (source: Ref. [35]).

Parameter	Value
Machines	2 × three-phase squirrel-cage IM, mechanically coupled
Rated power	1 HP (0.746 kW) each
Poles	4
Rated voltage/connection	220 V, Δ (delta)
Rated current	3 A
Supply and drive	WEG CFW-08 frequency converters
Fault mechanism	Rewound stator with accessible taps for ITSC insertion
Fault mode used (this study)	Low-impedance (LI)—full short path in turns
Fault severities	LI-1 (1.41%), LI-2 (4.81%), LI-3 (9.26%)
Load control	Resistive load bank via coupled machine/drive
Operating frequency	50 Hz
Machines	2 × three-phase squirrel-cage IM, mechanically coupled

The data comprise three healthy modes of low-impedance severity of inter-turn fault short-circuit conditions. The data used in this work include a total of 5999 extracted windows. The data are split across the four modes, where healthy data contain 1499 and 1500, respectively, for each fault, where each class has 1050 windows for training and 450 for testing.

3.3. Signal Preprocessing and Power Spectral Density Analysis

Before feature extraction, the raw stator current signals are analyzed using Power Spectral Density (PSD) to identify frequency regions with significant spectral energy associated with ITSC faults. The PSD represents how signal power is distributed across frequencies and is computed using the Welch method, as follows:

$$P_{xx}(f) = \frac{1}{Nf_s}\left|\sum_{n=0}^{N-1} x(n)e^{-j2\pi fn/f_s}\right|^2 \tag{2}$$

where $P_{xx}(f)$ is the PSD, $x(n)$ is the signal, f_s = 10 kHz, and N is the number of samples.

Fault signatures are primarily observed between 40 and 300 Hz, consistent with sideband frequencies around the 50 Hz fundamental. These results guide the design of a Butterworth band-pass filter, which isolates fault-sensitive frequency bands while suppressing high-frequency noise and supply harmonics.

The Butterworth filter transfer function is:

$$H(s) = \frac{1}{\sqrt{1+\left(\frac{s}{\omega_c}\right)^{2n}}} \tag{3}$$

where n is the filter order and ω_c represents the cutoff angular frequency derived from PSD analysis.

3.4. Hilbert Transform and Analytic Signal Construction

Filtered signals are processed using the Hilbert Transform, defined as:

$$\hat{x}(t) = \frac{1}{\pi}P.V.\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{4}$$

The analytic signal $z(t)$ is expressed as:

$$\mathrm{z(t)} = \mathrm{x(t)} + \mathrm{j}\,\hat{\mathrm{x}}(\mathrm{t}) = \mathrm{A(t)}\,\mathrm{e}^{\mathrm{j}\phi(\mathrm{t})} \quad (5)$$

The instantaneous amplitude (envelope) $a(t)$ is then obtained as:

$$a(t) = |z(t)| = \sqrt{x^2(t) + \widehat{x^2}(t)} \quad (6)$$

This envelope captures low-frequency amplitude modulation caused by inter-turn asymmetry and serves as the foundation for statistical feature extraction.

3.5. Statistical Feature Extraction

From each analytic signal envelope, thirteen time-domain statistical features are computed. These features characterize the amplitude, shape, and statistical distribution of the current waveform, as presented in Table 3 [12].

Table 3. Computed selected features from Hilbert Transform algorithms.

Features	Equations	
Root Mean Square	$x_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	(7)
Standard Deviation (StdDev):	$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}$	(8)
Mean	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	(9)
Kurtosis	$K = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^4$	(10)
Skewness	$S = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^3$	(11)
Peak Value	$x_{\mathrm{peak}} = \max_{1\le i\le N} \lvert x_i\rvert$	(12)
Shape Factor:	$SF = \frac{x_{\mathrm{RMS}}}{\lvert\mu\rvert}$	(13)
Crest Factor	$CF = \frac{x_{\mathrm{peak}}}{x_{\mathrm{RMS}}}$	(14)
Clearance Factor	$CLF = \frac{x_{\mathrm{peak}}}{\left(\frac{1}{N}\sum_{i=1}^{N} 2\sqrt{\lvert x_i\rvert}\right)^2}$	(15)
Impulse Factor	$IF = \frac{x_{\mathrm{peak}}}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$	(16)
Signal Energy	$E = \sum_{i=1}^{N} x_i^2$	(17)
Spectral Energy	$H = -\sum_{k=1}^{M} p_k \log_2(p_k),\ p_k$	(18)
Signal-to-Noise Ratio (SNR)	$\mathrm{SNR} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right)$	(19)

Let x_i denote the i^{th} sample, N the total number of samples, and $\bar{x}$ the mean of the signal.

3.6. Feature Selection Using mRMR

Feature selection was conducted using the Minimum Redundancy Maximum Relevance (mRMR) algorithm, a filter-based approach introduced by Peng, Long, and Ding [36]. The method aims to select features that have maximum dependency and relevance to the target class while simultaneously ensuring minimum redundancy among themselves. It is grounded in mutual information theory, which quantifies the statistical dependency between features and the target variable.

The mutual information between a feature hi and the class label c is expressed as:

$$I(h_i, c) = \int p(h_i, c)\log\left(\frac{p(h_i, c)}{p(h_i)p(c)}\right) dh_i\, dc \quad (20)$$

The algorithm seeks to maximize the relevance of selected features with respect to the class:

$$\max_{S} D(S,c) = \frac{1}{|S|}\sum\nolimits_{h_i \in S} I(h_i, c) \tag{21}$$

while minimizing redundancy among them:

$$\min_{S} R(S) = \frac{1}{|S|^2}\sum\nolimits_{h_i, h_j \in S} I\left(h_i, h_j\right) \tag{22}$$

Hence, the overall mRMR optimization objective is formulated as:

$$\max_{S}[D(S,c) - R(S)] \tag{23}$$

In this project, the mRMR algorithm was implemented in MATLAB's Classification Learner App after the feature dataset was partitioned using a 70/30 stratified split. Thirty percent of the holdout test set was set aside prior to model training to serve as a completely unseen test dataset for final model evaluation. The remaining 70% of the dataset was used to train the model via 5-fold cross-validation after implementing the mRMR algorithm. In each iteration, four folds were used to train the model, with one fold held out for validation, ensuring the training sample was validated exactly once. This selection approach was employed to retain only the most discriminative and non-redundant features before training the SVM and Random Forest classifiers.

Recent work demonstrated the effectiveness of mRMR in diagnosing inter-turn winding faults in induction motors. Their findings revealed that mRMR achieved higher accuracy and computational efficiency than other feature selection methods, such as Fisher Score and Relief, particularly for multi-dimensional time-domain data. This supports its use in the present study, where mRMR was applied to enhance classifier performance and reduce model complexity [26].

3.7. Classification Models

The classification stage of this study employed Support Vector Machine (SVM) and Random Forest (RF) algorithms.

3.7.1. Support Vector Machine

SVM constructs an optimal hyperplane that maximizes the margin between the classes. The decision function is defined as:

$$f(x) = \mathrm{sign}\left(\sum\nolimits_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \tag{24}$$

where $K(x_i, x)$ is the Quadratic kernel, α_i are Lagrange multipliers, y_i- are class labels, and $b-$ is the bias term [37,38].

The data points are represented as vectors belonging to one of two classes. To achieve optimal classification accuracy, the algorithm finds a decision boundary—the hyperplane—that maximizes the separation between these distinct classes. For non-linear data, SVM uses a kernel function instead of raw data points. While kernels are computationally simple and do not require prior domain knowledge, they can be slow when processing large-scale datasets.

3.7.2. Random Forest

The RF model combines multiple decision trees trained on random subsets of features and data. The final class prediction is determined by majority voting.

$$\hat{y} = \mathrm{mode}\{T_1(x), T_2(x), \ldots, T_M(x)\} \tag{25}$$

where $T_m(x)$ is the prediction from the m^{th} decision tree, and M is the total number of trees [39,40].

Both classifiers are trained and validated using 5-fold cross-validation in MATLAB's Classification Learner App. The RF model was trained using MATLAB's Classification Learner app with 30 decision trees and a maximum of 109 splits per tree. Total training time of 4.80 s.

3.7.3. Performance Evaluation Metrics

The performance of the proposed machine learning-based stator inter-turn fault detection framework was quantitatively assessed using accuracy, precision, recall, and F1 score. These metrics provide insight into the predictive reliability and robustness of the classifiers across both healthy and faulted operating conditions. A 5-fold cross-validation (5-CV) strategy was applied to minimize overfitting and ensure generalization, following a methodology consistent with Ali et al. [17] and Sharma et al. [41].

The performance of the classifiers was evaluated using the following standard definitions as presented in Table 4 [42]:

Table 4. Classification performance metrics.

Features	Equations	
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	(26)
Precision	$\frac{TP}{TP+FP}$	(27)
Recall	$\frac{TP}{TP+FN}$	(28)
F1-Score	$2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$	(29)

where TP, TN, FP, and FN represent true-positives, true-negatives, false-positives, and false-negatives, respectively.

The developed framework successfully integrates signal preprocessing, Hilbert-based feature extraction, mRMR feature optimization, and machine learning classification. The inclusion of PSD ensures the filter is tuned to ITSC-relevant frequencies, while mRMR reduces computational complexity by prioritizing the most informative features.

This methodology provides a reproducible approach for the early detection of stator-winding inter-turn faults under variable-load conditions and serves as the foundation for the subsequent results and discussion sections.

4. Results

4.1. Power Spectrum Density and Bandpass Filter

The frequency-domain representation of the stator current signals was analyzed using Power Spectral Density (PSD) to identify frequency regions sensitive to inter-turn short-circuit faults. Figure 3 presents the PSD for four operating conditions. In the healthy condition, the signal energy is concentrated around the fundamental frequency (50 Hz) and its harmonics, with minimal sideband activity.

As the fault severity increases, distinct sideband frequencies appear between approximately 45 Hz–49 Hz and 51 Hz–55 Hz, corresponding to modulation effects introduced by the rotor slip frequency ($f_r \approx$ 1–5 Hz).

Additional low-frequency components (below 40 Hz) arise from magnetic field asymmetry and rotor–stator coupling caused by shorted turns.

Furthermore, high-frequency spikes are visible between 2 kHz and 3 kHz across all faulted conditions. These peaks arise mainly from inverter switching harmonics and high-frequency transients, which intensify with increasing fault severity.

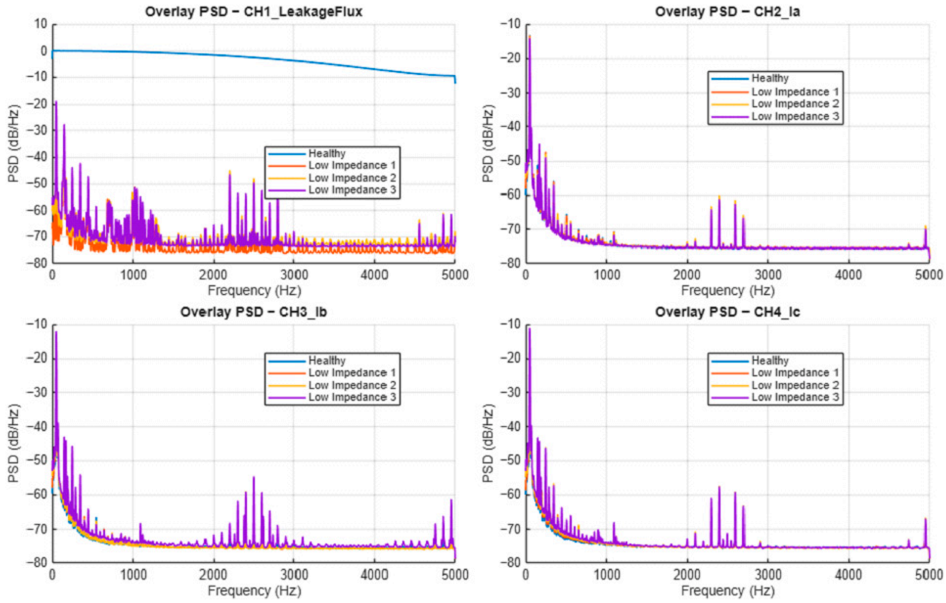

Figure 3. Power Spectrum Density analysis (PSD).

Figure 4 focuses on the most informative spectral content; the Butterworth bandpass filter was designed with a passband of 20 Hz–350 Hz, effectively preserving the fundamental and fault-related frequency components while suppressing switching noise and DC offsets.

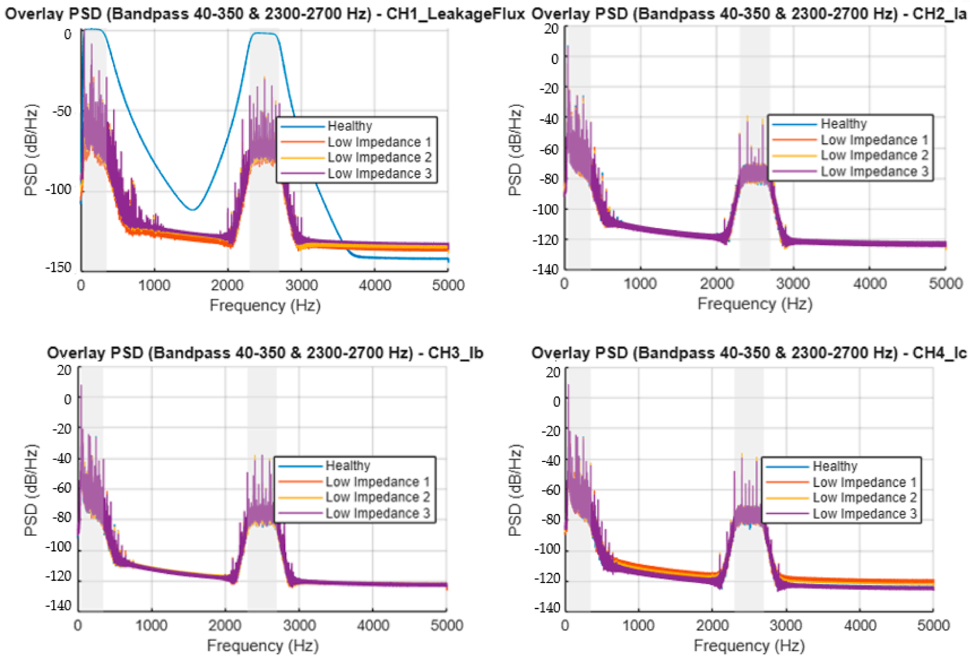

Figure 4. Butterworth bandpass filter.

The filtered signal was then processed through the Hilbert Transform to extract instantaneous amplitude envelopes, forming the basis for time-domain statistical feature extraction and subsequent machine-learning classification.

4.2. Feature Importance Ranking Using the mRMR Algorithm

Table 5 illustrates the ranking of the extracted features based on their importance scores computed using the Minimum Redundancy Maximum Relevance (mRMR) algorithm.

Table 5. Feature importance scores using mRMR.

Rank	Feature	Importance Score
1	Mean	0.6119
2	Shape Factor	0.5193
3	SNR	0.4297
4	Kurtosis	0.4202
5	Peak Value	0.3978
6	Energy	0.3455
7	Impulse Factor	0.3455
8	RMS	0.3421
9	Clearance Factor	0.2783
10	Crest Factor	0.2163
11	StdDev	0.2096
12	Spectral Entropy	0.1950
13	Skewness	0.0761

This analysis identifies the most discriminative statistical parameters for detecting inter-turn short-circuit faults.

As shown, the mean, Shape Factor, Signal-to-Noise Ratio (SNR), and kurtosis exhibit the highest relevance, indicating their strong influence in distinguishing between healthy and faulted conditions.

Meanwhile, features such as energy, Impulse Factor, and RMS also contribute significantly, capturing variations in amplitude and waveform distortion as a function of fault severity. Lower-ranked features, such as Spectral Entropy and Standard Deviation, provided less discriminative power but still enhanced model generalization.

These results validate the effectiveness of the mRMR algorithm in reducing redundancy and highlighting features with high diagnostic value prior to model training.

4.3. SVM Results

The SVM classifier demonstrated impressive performance in detecting inter-turn short-circuit faults, achieving an overall accuracy of 94.19% in a 5-fold cross-validation. As shown in the confusion matrix in Figure 5, the model correctly classified most samples across all four modes of operation with minimal misclassifications.

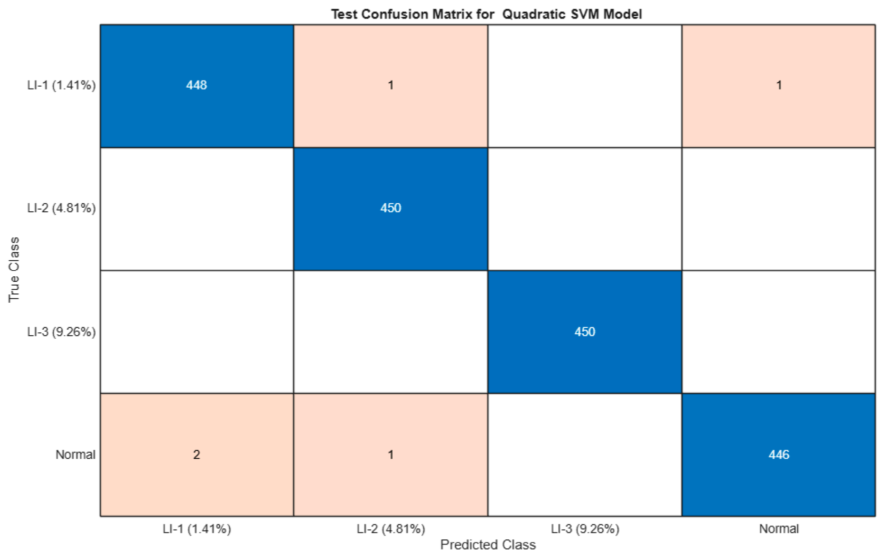

Figure 5. Test confusion matrix for Quadratic SVM model.

Table 6 shows that the highest precision and recall values were observed for Fault 2 (96.12%, 95.97%) and Fault 3 (98.00%, 97.83%), confirming the SVM's high reliability in identifying moderate and severe low-impedance faults. A slightly lower recall for Fault 1 (89.17%) indicates the challenge of recognizing incipient faults, where signal variations are subtle and overlap with healthy conditions.

Table 6. SVM classification performance metrics.

Faults	Precision	Recall	F1-Score	Support
Fault 1	0.9145	0.8917	0.9030	600
Fault 2	0.9612	0.9597	0.9605	620
Fault 3	0.9800	0.9783	0.9791	600
Healthy	0.9132	0.9375	0.9252	640
MacroAvg	0.9410	0.9360	0.9370	2460
MicroAvg	0.9360	0.9360	0.9360	2460
Accuracy	0.9419	-	-	2460

Figure 5 presents a confusion matrix that also confirms a more balanced and consistent performance across all classes and indicates few misclassified classes. The ROC analysis supports these findings, with AUC values above 0.99% for all fault classes, indicating excellent separability between healthy and faulty states, as presented in Figure 6.

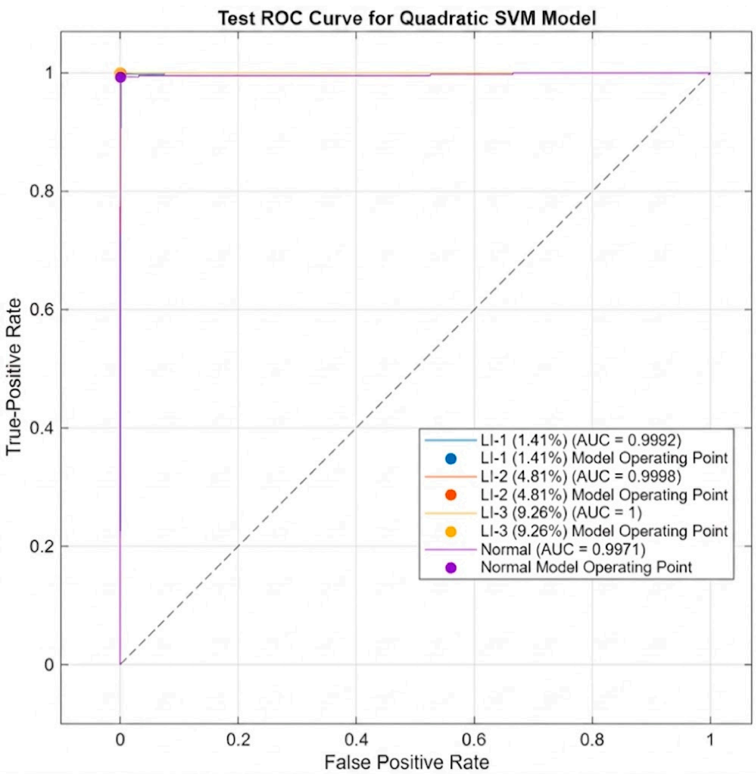

Figure 6. Test ROC curve for Quadratic SVM model.

The ROC analysis shows the highest AUCs for LI-3 to be 1.00, reflecting the classifier's strong sensitivity to fault progression, while the slightly lower healthy AUC presents 0.9971 highlights the difficulty in detecting early-stage insulation degradation. Overall, the SVM model exhibited robust diagnostic performance, demonstrating high sensitivity, precision, and generalization in distinguishing healthy from faulty motor conditions across varying load levels at 50 Hz.

4.4. Random Forest Results

Table 7 presents RF performance metrics; the classifier achieved exceptional performance in detecting inter-turn short-circuit faults, with an overall accuracy of 99.51% under 5-fold cross-validation. The model achieved precision and recall values above 99% for all classes, indicating consistent and reliable classification across all fault levels. Fault 3 presented 100% recall, confirming the RF model's strong capability to identify severe low-impedance faults with no false-negatives. The slightly lower precision for Fault 1 of 99.33% suggests minor overlap between the incipient fault and healthy features, but this has a negligible effect on overall accuracy. The macro and micro F1-scores present 0.9951, which further validates the model's uniform and unbiased performance. The confusion matrix clearly shows that all four operating conditions were classified with near-perfect accuracy and minimal misclassification, as presented in Figure 7.

Table 7. Performance evaluation metrics of RF.

Faults	Precision	Recall	F1-Score	Support
Fault 1	0.9933	0.9900	0.9917	600
Fault 2	0.9936	0.9952	0.9944	620
Fault 3	0.9983	1.0000	0.9992	600
Healthy	0.9953	0.9953	0.9953	640
MacroAvg	0.9410	0.9951	0.9951	2460
MicroAvg	0.9951	0.9951	0.9951	2460
Accuracy	0.9951	-	-	2460

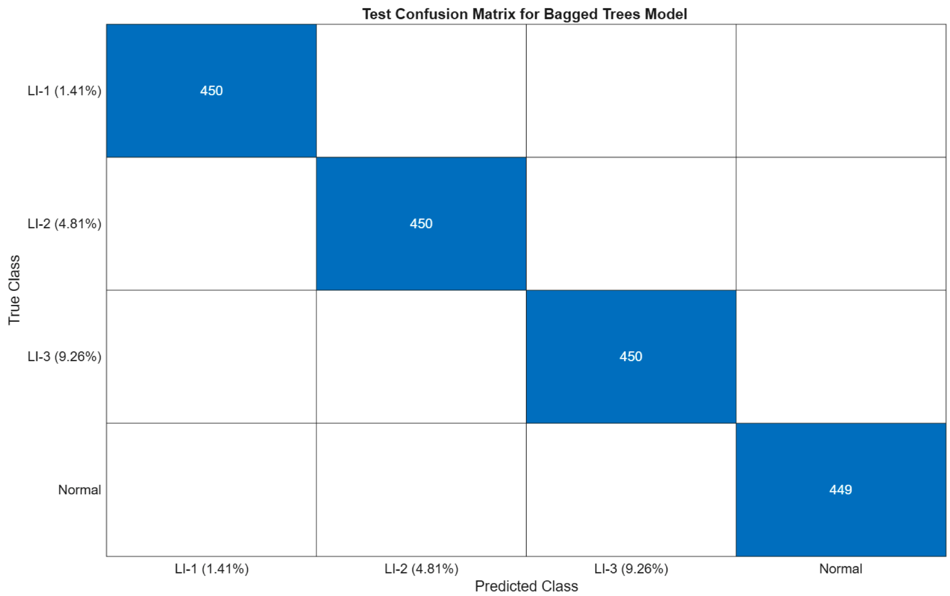

Figure 7. Test confusion matrix for Bagged Trees model.

The ROC curves in Figure 8 reinforce these results, showing AUC values of 0.9999 for Fault 1 and 1.0 for all other classes, indicating near-perfect class separability and outstanding discriminative power.

These findings confirm that the Random Forest model offers superior generalization and robustness compared to SVM, effectively capturing nonlinear feature interactions derived from PSD- and Hilbert-Transform-based signal analysis. Overall, the results highlight the Random Forest's exceptional reliability, sensitivity, and diagnostic precision in identifying inter-turn short-circuit faults under varying load conditions at 50 Hz.

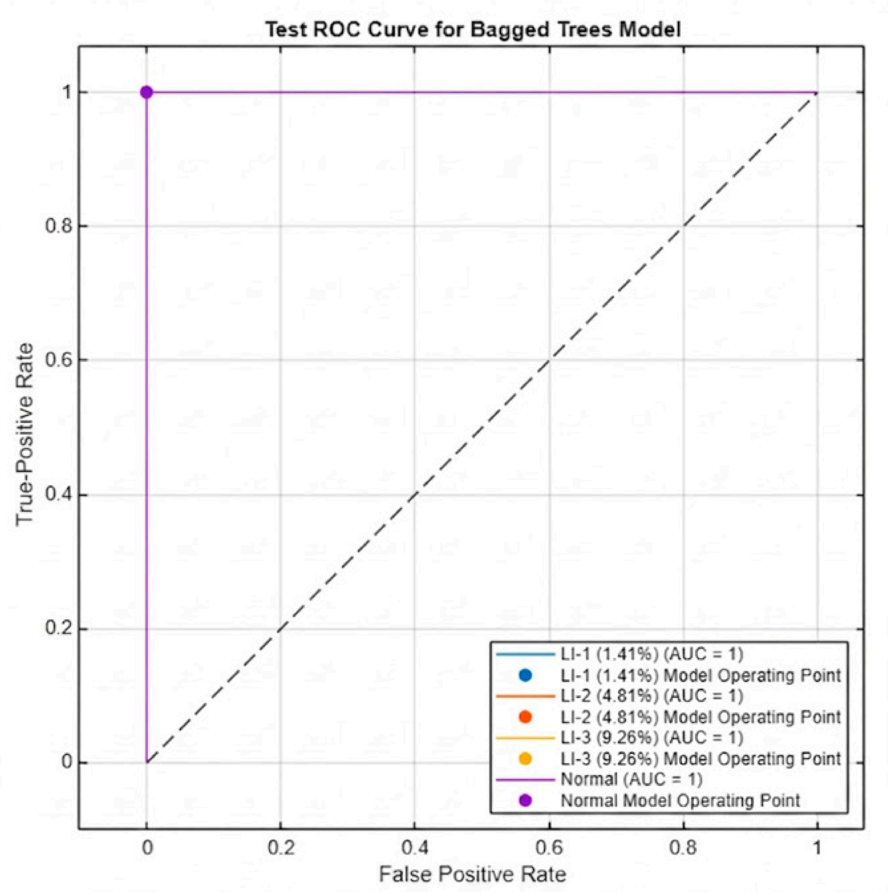

Figure 8. Test ROC curve for Bagged Trees model.

5. Conclusions

This project aimed to develop and evaluate a machine-learning-based approach to detect inter-turn short-circuit faults in three-phase induction motors under varying load conditions at 50 Hz. The primary objectives were to preprocess the acquired current signals using PSD and Hilbert Transform techniques, extract relevant statistical features, apply the minimum mRMR algorithm for feature selection, and compare the performance of two supervised classifiers, namely SVM and RF, in identifying and classifying low-impedance stator faults: LI-1, LI-2, LI-3, and healthy conditions.

The experimental framework was successfully implemented using secondary data from a controlled test bench comprising two coupled squirrel-cage induction machines, enabling realistic emulation of ITSC faults at varying severity levels. Signal processing and feature extraction were executed in MATLAB, where both classifiers were trained and validated using 5-fold cross-validation in the Classification Learner App.

The results demonstrated that both models achieved strong classification accuracy, confirming the effectiveness of the proposed methodology. The SVM classifier achieved an overall accuracy of 94.19%, with high precision and recall values above 90% across all classes. It showed robust performance in identifying moderate and severe faults (LI-2 and LI-3) but exhibited minor limitations in detecting early-stage faults (LI-1) due to overlapping spectral features between healthy and incipient faulty signals. In contrast, the Random Forest classifier achieved an outstanding accuracy of 99.51%, with macro and micro F1-scores of 0.9951 and AUC values approaching 1.0 for all fault classes. The confusion matrices and ROC curves indicated near-perfect classification capability, highlighting RF's superior feature discrimination and ability to model nonlinear relationships in the data.

A critical comparison of the two models shows that the Random Forest classifier outperformed the SVM on every evaluation metric, particularly recall and overall generalization. This superiority can be attributed to RF's ensemble learning structure, which reduces variance and captures complex feature interactions derived from PSD and Hilbert Transform-based inputs. The SVM, while effective, is more sensitive to overlapping decision boundaries and kernel parameter tuning, which affects its performance for incipient faults. Therefore, the Random Forest model is identified as the more reliable and robust choice for real-world ITSC fault detection and diagnosis in induction motors.

Overall, the project successfully met its objectives, demonstrating that integrating signal processing techniques with feature selection and machine learning can enable accurate, automated fault classification. The proposed framework offers a practical, data-driven diagnostic approach to enhance predictive maintenance systems for industrial motor applications. Future work may focus on extending the methodology to multi-fault scenarios, incorporating real-time data acquisition, and evaluating model performance under variable supply and environmental conditions.

Author Contributions: S.G. was responsible for the concept, methodology, and simulations, as well as the formal analysis and original draft. E.F.S. provided supervision and conceptual guidance while assisting with analysis, results validations, and critical editing. L.M. handled administrative oversight, supervision, and provided critical revisions to the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: https://www.kaggle.com/datasets/rebecacunha/mit-short-circuit-flux-and-current-signals (accessed on 16 August 2026).

Acknowledgments: During the preparation of this manuscript/study, the author(s) used Claude Pro, v1.260820.0 for the purposes of paraphrasing the content and Gemini 3.6 to improve graphics. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Thomson, W.T.; Fenger, M. Current Signature Analysis to Detect Induction Motor Faults. *IEEE Ind. Appl. Mag.* **2001**, *7*, 26–34. [CrossRef]
2. Benbouzid, M.E.H. A review of induction motors signature analysis as a medium for faults detection. *IEEE Trans. Ind. Electron.* **2000**, *47*, 984–993. [CrossRef]
3. Bonnett, A.H.; Soukup, G.C. Cause and analysis of stator and rotor failures in three-phase squirrel-cage induction motors. *IEEE Trans. Ind. Appl.* **1992**, *28*, 921–937. [CrossRef]
4. Nandi, S.; Toliyat, H.A.; Li, X. Condition Monitoring and Fault Diagnosis of Electrical Motors—A Review. *IEEE Trans. Energy Convers.* **2005**, *20*, 719–729. [CrossRef]
5. Raj, K.K.; Tamata, J.; Waisale, L.; Assaf, M.H.; Groza, V.; Kumar, R.R. Enhancing Reliability in Induction Machines: A Focus on Stator Inter-Turn Fault Diagnosis. In *Proceedings of the 2024 IEEE 18th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 21–24 May 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 000449–000454. [CrossRef]
6. Arkan, M.; Perović, D.K.; Unsworth, P. Online stator fault diagnosis in inductionmotors. *IEE Proc. Electr. Power Appl.* **2001**, *148*, 537–547. [CrossRef]
7. Xu, Z.; Hu, C.; Yang, F.; Kuo, S.H.; Goh, C.K.; Gupta, A.; Nadarajan, S. Data-Driven Inter-Turn Short Circuit Fault Detection in Induction Machines. *IEEE Access* **2017**, *5*, 25055–25068. [CrossRef]
8. Ünsal, A.; Kabul, A. Diagnosis of multiple faults of an induction motor based on Hilbert envelope analysis. *Metrol. Meas. Syst.* **2022**, *29*, 191–205. Available online: https://journals.pan.pl/dlibra/publication/138541/edition/122790 (accessed on 11 November 2025). [CrossRef]

9. Chen, L.; Shen, J.; Xu, G.; Chi, C.; Feng, Q.; Zhou, Y.; Deng, Y.; Wen, H. Induction Motor Stator Winding Inter-Tern Short Circuit Fault Detection Based on Start-Up Current Envelope Energy. *Sensors* **2023**, *23*, 8581. [CrossRef] [PubMed]
10. dos Santos, T.; Ferreira, F.; Pires, J.; Damásio, C. Stator winding short-circuit fault diagnosis in induction motors using random forest. In *Proceedings of the 2017 IEEE International Electric Machines & Drives Conference (IEMDC), Miami, FL, USA, 21–24 May 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 1–8. [CrossRef]
11. Panigrahy, P.S.; Santra, D.; Chattopadhyay, P. Decent fault classification of VFD fed induction motor using random forest algorithm. *AI EDAM* **2020**, *34*, 492–504. [CrossRef]
12. Haroun, S.; Seghir, A.N.; Touati, S. Multiple features extraction and selection for detection and classification of stator winding faults. *IET Electr. Power Appl.* **2018**, *12*, 339–346. [CrossRef]
13. Rengifo, J.; Moreira, J.; Vaca-Urbano, F.; Alvarez-Alvarado, M.S. Detection of Inter-Turn Short Circuits in Induction Motors Using the Current Space Vector and Machine Learning Classifiers. *Energies* **2024**, *17*, 2241. [CrossRef]
14. Jiang, Y.; Ji, B.; Zhang, J.; Yan, J.; Li, W. An Overview of Diagnosis Methods of Stator Winding Inter-Turn Short Faults in Permanent-Magnet Synchronous Motors for Electric Vehicles. *World Electr. Veh. J.* **2024**, *15*, 165. [CrossRef]
15. Rajagopalan, S.; Habetler, T.G.; Harley, R.G.; Sebastian, T.; Lequesne, B. Current/Voltage-Based Detection of Faults in Gears Coupled to Electric Motors. *IEEE Trans. Ind. Appl.* **2006**, *42*, 1412–1420. [CrossRef]
16. Awadallah, M.A.; Morcos, M.M. Application of AI tools in fault diagnosis of electrical machines and drives-an overview. *IEEE Trans. Energy Convers.* **2003**, *18*, 245–251. [CrossRef]
17. Ali, M.Z.; Shabbir, M.N.S.K.; Liang, X.; Zhang, Y.; Hu, T. Machine Learning-Based Fault Diagnosis for Single- and Multi-Faults in Induction Motors Using Measured Stator Currents and Vibration Signals. *IEEE Trans. Ind. Appl.* **2019**, *55*, 2378–2391. [CrossRef]
18. Smart, E.; Brown, D.; Axel-Berg, L. Comparing one and two class classification methods for multiple fault detection on an induction motor. In *Proceedings of the 2013 IEEE Symposium on Industrial Electronics & Applications, Kuching, Malaysia, 22–25 September 2013*; IEEE: Piscataway, NJ, USA, 2013; pp. 132–137. [CrossRef]
19. Kawady, T.A.; Khater, W.M.; Abu-Faty, H.G.; Izzularab, M.A.; Ibrahim, M.E. Induction Motor Stator inter-turn Fault Diagnosis Approach based on Machine Learning Models. In *Proceedings of the 2024 25th International Middle East Power System Conference (MEPCON)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–7. Available online: https://ieeexplore.ieee.org/abstract/document/10850370/ (accessed on 28 October 2025).
20. Yoo, Y.J. Fault Detection of Induction Motor Using Fast Fourier Transform with Feature Selection via Principal Component Analysis. *Int. J. Precis. Eng. Manuf.* **2019**, *20*, 1543–1552. [CrossRef]
21. Romeral, L. Fault Detection in Induction Machines Using Power Spectral Density in Wavelet Decomposition. *IEEE Trans. Ind. Electron.* **2008**, *55*, 633–643. [CrossRef]
22. Hussain, M.; Ahmed, R.R.; Kalwar, I.H.; Memon, T.D. Multiple faults detection and identification of three phase induction motor using advanced signal processing techniques. *3C Tecnol.* **2020**, 93–117. [CrossRef]
23. Glowacz, A.; Glowacz, Z. Diagnosis of the three-phase induction motor using thermal imaging. *Infrared Phys. Technol.* **2017**, *81*, 7–16. [CrossRef]
24. Goh, Y.J.; Kim, O. Linear Method for Diagnosis of Inter-Turn Short Circuits in 3-Phase Induction Motors. *Appl. Sci.* **2019**, *9*, 4822. [CrossRef]
25. Kim, M.-C.; Lee, J.-H.; Wang, D.-H.; Lee, I.-S. Induction Motor Fault Diagnosis Using Support Vector Machine, Neural Networks, and Boosting Methods. *Sensors* **2023**, *23*, 2585. [CrossRef] [PubMed]
26. Samiullah, M.; Ali, H.; Zahoor, S.; Ali, A. Fault Diagnosis on Induction Motor using Machine Learning and Signal Processing. *arXiv* **2024**, arXiv:2401.15417. [CrossRef]
27. Pohakar, P.; Gandhi, R.; Hans, S.; Sharma, G.; Bokoro, P.N. Analysis of multiple faults in induction motor using machine learning techniques. *e-Prime Adv. Electr. Eng. Electron. Energy* **2025**, *12*, 101007. Available online: https://www.sciencedirect.com/science/article/pii/S2772671125001147 (accessed on 28 October 2025). [CrossRef]
28. Palaniappan, R. Comparative Analysis of Support Vector Machine, Random Forest and k-Nearest Neighbor Classifiers for Predicting Remaining Usage Life of Roller Bearings. *Informatica* **2024**, *48*, 39–52. [CrossRef]
29. Abdulkareem, A.; Anyim, T.; Popoola, O.; Abubakar, J.; Ayoade, A. Prediction of induction motor faults using machine learning. *Heliyon* **2025**, *11*, e41493. [CrossRef] [PubMed]
30. Kumar, R.; Anand, R.S. Machine Learning Approach with Multiple Feature Selection Techniques to Diagnose the Inter-Turn Winding Faults in Induction Motor. *Arab. J. Sci. Eng.* **2025**, *50*, 5945–5961. [CrossRef]
31. Leon-Olvera, A.N.; Rangel-Rodriguez, A.H.; Granados-Lieberman, D.; Amezquita-Sanchez, J.P.; Valtierra-Rodriguez, M. A machine learning strategy for detection of stator winding short-circuit faults in induction motors. In *Proceedings of the 2023 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 11–13 October 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7. [CrossRef]
32. Yakhni, M.F.; Cauet, S.; Sakout, A.; Assoum, H.; Etien, E.; Rambault, L.; El-Gohary, M. Variable speed induction motors' fault detection based on transient motor current signatures analysis: A review. *Mech. Syst. Signal Process.* **2023**, *184*, 109737. [CrossRef]

33. Boucerredj, L.; Benalia, N. A comparative study of machine learning classifiers for intelligent fault diagnosis of electric vehicles based on FMECA data. *Adv. Mech. Eng.* **2025**, *17*, 16878132251342413. [CrossRef]
34. Gangsar, P.; Tiwari, R. Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review. *Mech. Syst. Signal Process.* **2020**, *144*, 106908. [CrossRef]
35. Cunha, R. Inter-Turn Short-Circuit in Induction Motor: Leakage Flux and Current Signals Dataset. *Kaggle* 2022. Available online: https://www.kaggle.com/datasets/rebecacunha/mit-short-circuit-flux-and-current-signals (accessed on 16 August 2026).
36. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]
37. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
38. Suykens, J.A.K.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [CrossRef]
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Shi, H.; Yuan, Z.; Zhang, Y.; Zhang, H.; Wang, X. A New Ensemble Strategy Based on Surprisingly Popular Algorithm and Classifier Prediction Confidence. *Appl. Sci.* **2025**, *15*, 3003. [CrossRef]
41. Sharma, P.; Cirrincione, G.; Kumar, R.R.; Mohammadi, A.; Cirrincione, M. A Comparative Study for the Detection of Stator Inter-Turn Faults in Induction Motors Using Shallow Neural Networks and Non-Neural Based Techniques. In *Proceedings of the 2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET), Hammamet, Tunisia, 29 April–1 May 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [CrossRef]
42. Sharma, N.R.; Bhalja, B.R.; Malik, O.P. Machine Learning-Based Severity Assessment and Incipient Turn-to-Turn Fault Detection in Induction Motors. *IEEE Trans. Energy Convers.* **2025**, *40*, 557–567. [CrossRef]