Received: 28 August 2022 | Revised: 17 October 2022 | Accepted: 18 October 2022

DOI: 10.1049/elp2.12262

IET Electric Power Applications

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Incipient detection of stator inter-turn short-circuit faults in a Doubly-Fed Induction Generator using deep learning

Ghasem Alipoor[1] | **Seyed Jafar Mirbagheri**[1] | **Seyed Mohammad Mahdi Moosavi**[1] | **Sérgio M. A. Cruz**[2]

[1]Electrical Engineering Department, Hamedan University of Technology, Hamedan, Iran

[2]Department of Electrical and Computer Engineering, University of Coimbra and Instituto de Telecomunicações, Coimbra, Portugal

Correspondence

Ghasem Alipoor, Electrical Engineering Department, Hamedan University of Technology, Hamedan 6516913733, Iran.
Email: alipoor@hut.ac.ir

Abstract

Wind turbines are increasingly expanding worldwide and Doubly-Fed Induction Generator (DFIG) is a key component of most of them. Stator winding fault is a major fault in this equipment and its incipient detection is of vital importance. However, there is a paucity of research in this field. In this study, a novel machine learning-based method is proposed for incipient detection of inter-turn short-circuit fault (ITF) in the DFIG stator based on the current signals of the stator. The proposed method makes use of state-of-the-art deep learning methods along with conventional signal processing tools and general machine learning techniques. More specifically, the incipient fault detection problem is regarded as a multi-class classification problem and a Long Short-Term Memory network, which is more appropriate for time-series data is utilised for inference. Furthermore, a variant of the celebrated Empirical mode Decomposition analysis tool is used to extract some well-known statistical features among which the most informative ones are selected using a new feature selection method. Our tests using experimental data in steady-state conditions show that the proposed method can accurately detect ITF fault at its initial stage when only one turn is shorted. Moreover, its performance is considerably higher than that of a variety of machine learning-based methods.

KEYWORDS

deep learning, doubly-fed induction generator (DFIG), empirical mode decomposition (EMD), fault detection, feature selection, inter-turn short-circuit fault

1 | INTRODUCTION

Climate change is the greatest threat facing humanity and shifting to renewable energy seems to be the inevitable remedy. As a result, the capacity of renewable energy systems has been increasingly expanded over the past decades; and to achieve the net-zero climate goals by 2050, this growth must be even faster. Wind energy, as one of the most cost-competitive and resilient power sources, has always been the leading non-hydro renewable energy source worldwide [1] and even outstripped hydro-power in the US [2]. Wind is now an affordable and beneficial source of energy; partly owing to the efficient and reliable condition monitoring services that are absolutely vital for the efficiency, reliability, and safety of wind power plants. The kinetic energy of the wind is generally converted to electricity using a wind turbine whose key element is a generator.

A study carried out on more than 1500 wind turbines operating in Germany over a period of 15 years showed that among all components of a wind turbine, damage to the generator results in the longest downtime [3]. On the other hand, with a 48.62% share, Doubly-Fed Induction Generator (DFIG) is the most commonly used generator in wind turbines [4], mainly because utilising partial converters and induction machine makes it more cost-effective when compared to other technologies. A survey conducted by the Motor Reliability Working Group of the Institute of Electrical and Electronics Engineers Industrial Application Society revealed that winding failures account for more than 29% of the faults in electric motors [5]. Another study done under the sponsorship of the

Electric Power Research Institute suggested that 36% of failures in powerhouse motors pertain to the stator, mostly winding faults [6]. With the increase in the rated power of the electric machines, one can surmise that these percentages can be even higher.

Among the various winding faults, the inter-turn short-circuit fault (ITF) is the most incipient one and the main cause of the others. An ITF, if not detected in an early stage, results in an increase in the circulating current and therefore overheating that in turn can deteriorate the insulation and lead to more severe faults, such as phase-to-phase and phase-to-ground faults. This fault can also propagate to other parts, for example, rotor and even mechanical components. Timely fault detection is more critical for equipment used for power generation since any interruption has both financial and social consequences. On the other hand, it is extremely difficult to detect the ITF incipiently, since it has an almost imperceptible symptom.

Fault detection and diagnosis (FDD) in induction machines has always been a hot topic of research and technology. Various methods proposed for FDD can be categorised into three main classes: model-based, signal-based, and data-based approaches. In the first approach, a specific machine fault is mathematically modelled based on the physics of the fault and its impacts on the observed signals. Some mathematical models derived for ITF in DFIGs can be found in refs. [7–12]. Model-based methods are susceptible to the problem characteristics and even the parameter variation and suffer from the lack of flexibility. On the other hand, in signal-based methods signals are analysed to extract symptoms or signatures of the fault. This can be done in time or/and frequency domains by signal processing tools such as Discrete Fourier Transform and its short-time version known as the Short Time Fourier Transform, Discrete Wavelet Transform, Hilbert–Huang transform, and Empirical mode Decomposition (EMD). Among numerous signal-based FDD methods proposed in the literature, a small number addressed the ITF detection problem in DFIGs [13–16]. One of the most popular techniques in this category that has been widely used for ITF detection is the Motor Current Signature Analysis [17].

In both aforementioned categories, interpretation and final inference must be done by an expert. In addition to affecting the detection performance, this fact increases the decision time which is very crucial in many applications. Nevertheless, the methods within the third category employ machine learning algorithms to make the FDD process more intelligent and automated. The main idea of the data-based FDD approaches is that all necessary information is embedded in the data and all one needs to do is to let a machine (that is an algorithm) learn from the data how to detect the fault. Furthermore, in some cases, it may be useful to first extract discriminative features from the signals using signal processing tools, before applying them to the machine.

Employing classical machine learning algorithms, such as Support Vector Machine (SVM), shallow neural networks, genetic algorithm, fuzzy logic, and decision trees for ITF detection in induction machines has a long history [18–25]. However, over the last 2 decades, the new concept of Deep Learning (DL) brought about an evolution in the field of machine learning [26]. In many fields, for example, natural language, speech, image, and vision, DL methods have met successes far beyond the classical methods and eclipsed them. Within the past 5 years, several attempts have also been made to use DL techniques for fault detection, but the conducted study on ITF detection in induction machines is very few [27–32] and to the best of our knowledge, there is no report on employing machine learning methods (whether shallow or deep) for ITF detection in DFIGs.

Due to the importance of the ITF detection problem in DFIGs and in light of the paucity of research in the field, this study is dedicated to incipient detection of stator ITF in DFIGs using DL techniques. For this aim, detection of the ITF at various levels is considered as a multi-class classification problem and a deep model is trained to discover the occurrence of the fault and detect it at its initial stage when only one turn is shorted. Owing to the higher performance of the Long Short-Term Memory (LSTM) network in coping with time-series data, this deep model is adopted in the current study. Additionally, among the various modalities used for FDD, we decided to use the three-phase stator current signals. The main reasons are that these signals contain more information about and have a higher sensitivity to the target fault, while no dedicated sensor is used as these signals are usually measured for other purposes. On the contrary, vibration signal that is vastly used for fault detection suffers from two problems of the uncertainty of sensor location and the high vulnerability to noise.

To further improve the capability of the LSTM, the celebrated EMD signal processing tool is also employed in the proposed scheme, based on which some well-known statistical features are extracted and a new feature selection approach is also devised. Empirical mode Decomposition [33] is a data-adaptive time-frequency analysis method that has proven to be useful in the field of FDD. In this analysis method, a signal is modelled as the superposition of some components that are the natural oscillatory modes embedded in the signal. Moreover, in contrast to the harmonic functions, the frequency and the amplitude of these components are generally time-varying. These facts make EMD useful for our study since electrical signals are originally sinusoidal where any fault superimposes some disturbances, with variable amplitudes and frequencies, on it. On the other hand, EMD decomposes a signal into some components that are all in the time domain and of the same length as the original signal. This allows for varying frequencies in time to be preserved and hence makes EMD a useful tool for extracting information inherent to the signal. Specifically, in the proposed method, each signal is first decomposed into its main components using the EMD and some statistical features are then extracted from each component. Subsequently, the most useful features are selected using the proposed feature selection technique, that are then used as the inputs of an LSTM.

The novelties of the work can be summarised as follows:

- Studying the stator ITF detection problem using DL techniques for the first time in DFIGs

- Incorporating EMD tool from signal processing realm and feature extraction and selection techniques from classical machine learning field into the proposed scheme to devise a method that can effectively detect the ITF in its incipient stages
- Devising a new feature selection approach that efficiently and effectively selects the most suitable features extracted from the signals
- Applying the proposed method to experimental data in steady-state operation and conducting extensive investigations and evaluations in comparison with classical machine learning methods as well as state-of-the-art methods proposed for ITF detection in induction machines

The rest of the manuscript is organised as follows. Literature background, including a short introduction to the LSTM model and the EMD as well as a concise review of the research done in the field of ITF detection using DL, is presented in the next section. The following two sections are dedicated to describing the proposed method and reporting the test results respectively. Finally, some concluding remarks are drawn up in the last section.

2 | BACKGROUND

2.1 | Empirical mode Decomposition and its variants

Empirical mode Decomposition is a robust time-frequency analysis method first proposed by Huang et al. to adaptively decompose non-stationary signals into their intrinsic oscillatory modes [33]. These components are called intrinsic mode functions (IMF), each with slowly varying amplitude and phase. IMFs satisfy the generalised alternating and zero-mean properties and relaxes the amplitude and frequency from being constant. In other words, EMD is based on a simple premise that each signal is made of a number of IMF each with the following two conditions:

- **Alternating Property:** The number of extrema and the number of zero-crossings must be either equal or differ at most by one, that is, IMFs have alternating stationary points and zeroes.
- **Zero-Mean Property:** At any point, the mean value of the envelopes defined by the local maxima and local minima is zero, that is, the maxima and the minima of the IMFs are opposite in sign.

These IMFs are extracted through an algorithm known as the sifting process that can effectively sift the complex signals in the time domain. IMFs provide valuable information about the signal. Empirical mode Decomposition suffers from a problem known as the mode-mixing and Ensemble EMD (EEMD) was introduced to tackle this problem [34]. In the EEMD algorithm, IMFs are obtained by averaging over modes resulting from applying the EMD sifting procedure on several noisy versions of the original signal, each with different realisations of the white noise. As a result of adding noise to the signal, the signal reconstructed from the IMFs of the EEMD algorithm contains some residual noise. Furthermore, adding different samples of the white noise may lead to different IMFs. These issues are addressed in a modified version of the EEMD, called complementary EEMD (CEEMD) [35]. Figure 1 demonstrates an example of decomposing a frame of faulty stator current signal into its first five IMFs, using the CEEMD algorithm.

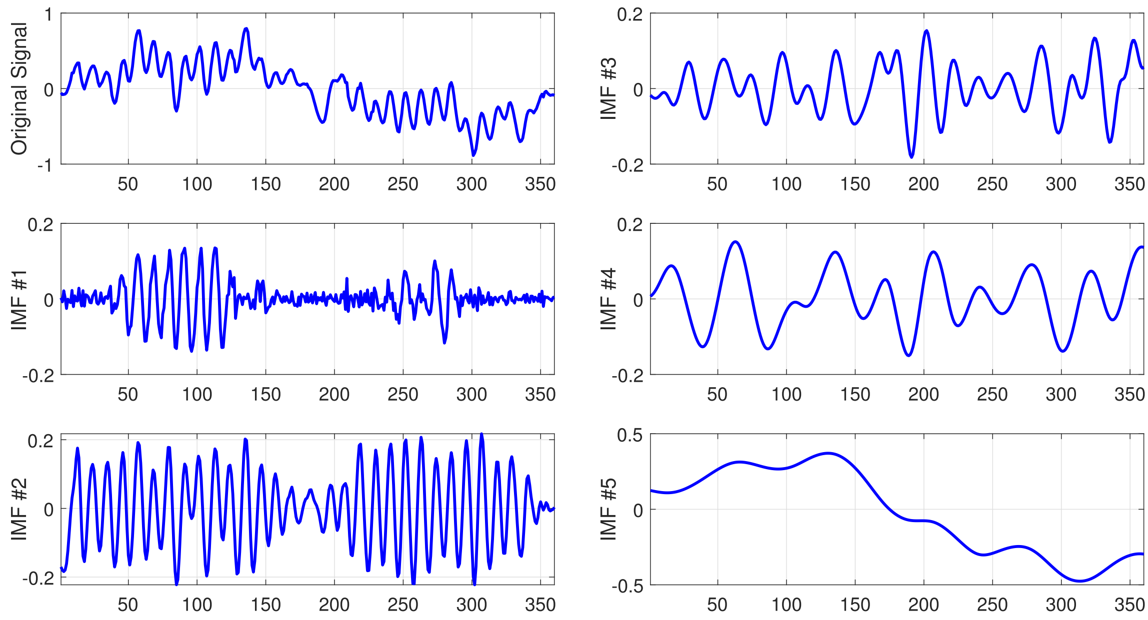


FIGURE 1 Applying complementary EEMD (CEEMD) to a frame of faulty stator current signal.

2.2 | Long Short-Term Memory

Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are two main classes of deep neural networks that have received a wealth of interest in many applications. Unlike more conventional feed-forward neural networks, including CNNs, RNN has feedback connections that allow the network to feed its activation at the current time step back to its input at the next step. In other words, RNN has a hidden state that stores the information from previous steps. This temporal dynamic makes RNNs useful in dealing with signals of the form of time series, such as signals considered in the current study. This mechanism can be better perceived using the diagrams shown in Figure 2 and the equation (1) that describes its operation.

$$h_t = \tanh(W_h x_t + U_h h_{t-1} + b_h), \tag{1}$$

where W_h and U_h matrixes and b_h vector are the parameters of the model that should be learnt during the training phase. W_h and U_h are the weight matrixes that weigh the input vector at the current time step (i.e. x_t) and the hidden state vector at the previous step (i.e. h_t) respectively. The result is then added to the bias vector b_h and is finally passed to a non-linear activation function that is usually a tanh function. The recurrent nature of the RNN model can be more easily perceived by unrolling it into successive time steps, as illustrated in Figure 2b.

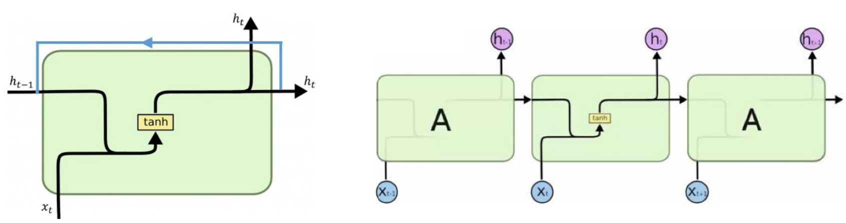
In practice, the standard RNN, also known as the vanilla RNN, suffers from two main problems of gradient vanishing and gradient exploding. In particular, when an RNN is trained using back-propagation, due to round-off errors, the long-term gradient values may tend to zero or infinity. This phenomenon restricts the memory of the vanilla RNN and prevents it from remembering long-term dependencies. Long Short-Term Memory network, first introduced in ref. [37], is a special version of the standard RNN that solves the gradient vanishing problem by controlling the information that should be kept and/or discarded [38, 39]. The structure of an LSTM in its compressed and unrolled forms are depicted in Figure 3 and the governing equations are:

$$\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
\hat{C}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
C_t &= f_t \odot C_{t-1} + i_t \odot \hat{C}_t \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
h_t &= o_t \odot \tanh(C_t)
\end{aligned} \tag{2}$$

where $\odot$ denotes the Hadamard (element-wise) product and again W and U matrixes and b vectors are the trainable the parameters of the model.

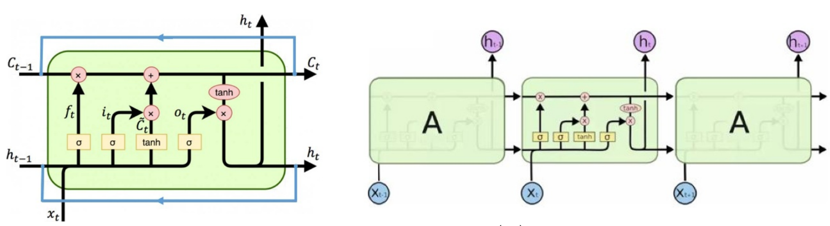
The main feature of this structure is the cell state C_t that like a conveyor belt carries contextual information throughout the time. This context is controlled by the following four gates:

- **Forget gate** decides the information from the previous step that should be forgotten and keeps what remains. This is done by creating f_t vector whose elements are numbers between 0 and 1 that then weight the elements of the previous context C_{t-1}.
- **Candidate gate** creates the new information $\hat{C}_t$ to be added to the context.
- **Input gate** controls the amount of the new information that should be added to the context. This is done by creating i_t vector whose elements are numbers between 0 and 1 that then weight the elements of the newly created context $\hat{C}_t$.
- **Output gate** determines what the output or the next hidden state should be. The result of this gate, i.e. o_t, is then weighted with numbers between 0 and 1 that are created by passing the updated cell state C_t through a tanh layer.



(a) Compressed graph (b) Unrolled graph

FIGURE 2 Structure of the vanilla Recurrent Neural Network (RNN) [36]



(a) Compressed graph (b) Unrolled graph

FIGURE 3 Structure of the Long Short-Term Memory (LSTM) [36]

2.3 | inter-turn short-circuit fault detection using deep learning

1-D CNN was used in ref. [27] to identify stator ITF in squirrel-cage induction motors, in which some well-known wavelets are used to initialise the convolution kernels. Various ITF scenarios, including the simultaneous occurrence of multiple faults, are considered in this study. The reported accuracy is 96.13%. In another study reported in ref. [28], 2-D CNNs composed of multiple layers are proposed for stator ITF detection in induction motors, where 16 different fault categories (including 1 ~ 5 shorted turns in each phase, as well as the healthy state) is considered. The impacts of the structure of the CNN and learning parameters on the performance are investigated and the best accuracy value is reported to be 99.3%. However, no valid comparison with existing methods is reported in these papers.

Three other researches that deserve more attention are reported in refs. [30–32]. In ref. [30], LSTM network is employed in the encoder-decoder structure equipped with the attention mechanism to estimate an indicator of stator ITF in permanent magnet synchronous machines. In the proposed method, a bidirectional LSTM is used in the encoder. A bidirectional LSTM consists of two LSTMs operating in the opposite directions that can extract both forward and backward dependencies. Moreover, raw values of the negative-sequence current and the positive-sequence current, both calculated based on the stator three phase currents, and the rotational speed are used as the inputs to the network. Although the fault indicator indicates the severity of the fault and the reported results show high accuracy of the estimation, it was not utilised for automatic fault detection and classification.

In ref. [32], a method is proposed to detect gearbox bearing and generator winding faults in wind turbines. In this method, LSTM is used to predict some output variables based on some other input variables and then the fault condition is diagnosed by comparing the probability distributions of the true output variables with those of the predicted variables. Discrimination between the true and predicted probabilities is performed based on the Kullback-Leibler Divergence. 128 monitoring variables, including several sensors mounted on the sub-systems of the turbine, from a Supervisory Control And Data Acquisition system is used, where wind speed and active power variables are selected as the model inputs based on which the temperature and the pressure variables are predicted. The best accuracy reported for the generator winding fault is 92% which is higher than those reported for some other state-of-the-art methods. In ref. [31], LSTM and Gated Recurrent Unit (GRU), which is another version of the RNN, are combined with the 1D-CNN for stator ITF detection in induction motors. It is reported that the proposed hybrid methods outperform the individual networks, that is, 1D-CNN, LSTM, and GRU.

3 | PROPOSED METHOD

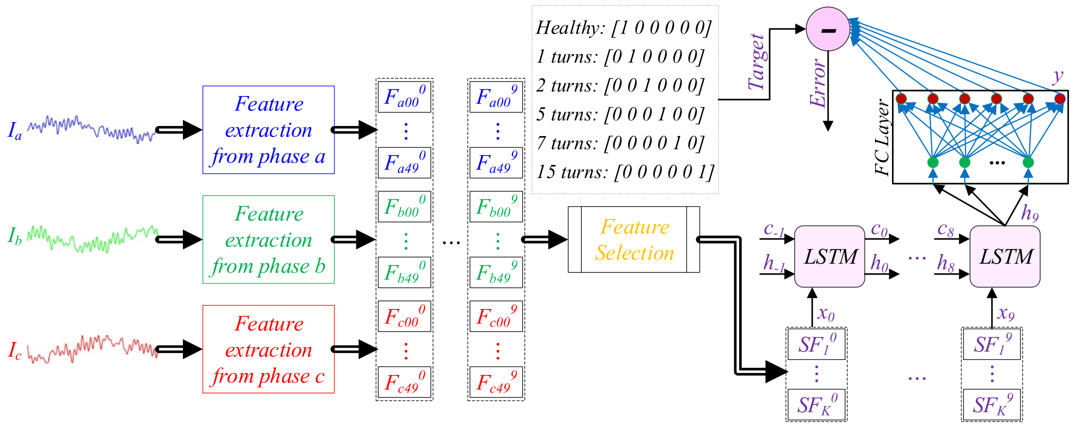
Our proposed method for incipient detection of the stator ITF in DFIGs is pictured in Figure 4. In this paper, three phase signals and corresponding features are denoted by subscripts a, b, and c. Three phase currents, that is, I_a, I_b, and I_c, go through the same procedure of feature extraction to extract 50 features from each current signal.

Each signal is first decomposed into its five dominant IMFs using the CEEMD algorithm, whose indices range from 0 to 4. To exploit the temporal evolution of the characteristics of the signals, following the behaviour of the LSTMs, each IMF is then chopped into 10 sub-frames of equal lengths, denoted by L. Each sub-frame is attributed to a time step and the corresponding index, ranging from 0 to 9, is indicated by a superscript. Finally, 10 renowned statistical features are extracted from each sub-frame, whose indices are represented by the third subscript. These features and their definitions and indices are summarised in Table 1. Each feature has unique characteristics and reveals distinct information about the status of the machine. Mean, Root Mean Square (RMS), skewness, and kurtosis are the four lowest-order statistics that contain information about the general tendency, energy, symmetry, and outliers. The crest factor, impulse factor, and margin factor are three impulsive metrics that characterise the peakedness of the signals. On the other hand, entropy quantifies the randomness or disorder and zero-crossing involves the notion of the rate of change or frequency. Changes in these features can indicate changes in the health status of the generator.

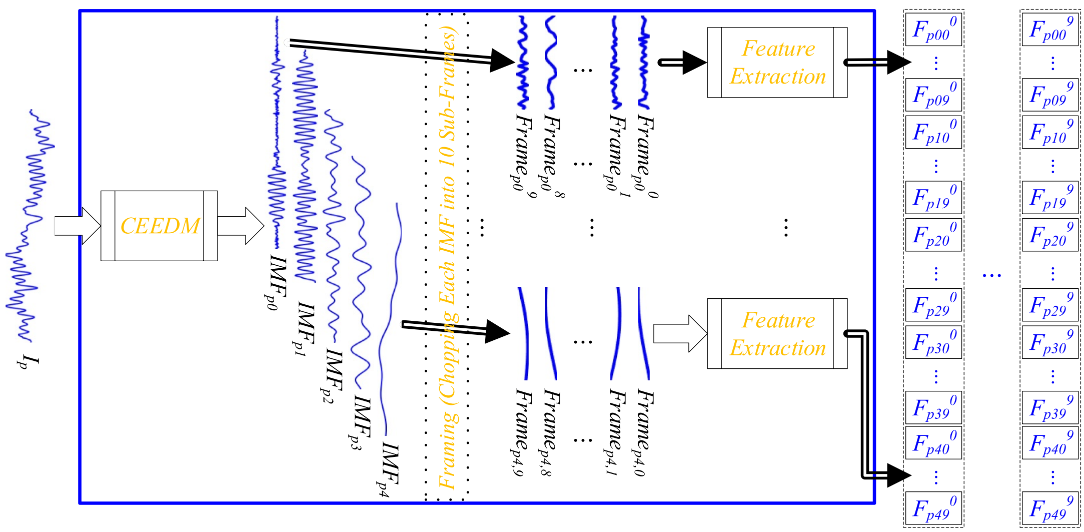
The above-described procedure results in a temporal sequence of 10 feature vectors that each contains 150 statistical features. These features are denoted as F_{pij}^{n} where subscripts p, i, and j represent phase, IMF, and statistical feature indices, respectively and the superscript n shows the time step. The successive 10 feature vectors can serve as the sequential inputs to the LSTM. In this manner, the LSTM processes each vector at a time step and the output at the last time step, that is, h_9 vector, is fed into a fully-connect layer that follows the LSTM and makes the final classification. The target variable is one-hot encoded.

The 150-length feature vectors extracted from 3 current signals can be directly applied to the LSTM. However, not all these features are suitable for classification, since some of them have not enough relevancy with the target variables and there may also be some redundancy among them. In fact, using irrelevant or redundant features may degrade the classification performance or at least increase complexity and overfitting; therefore discarding these features is always desirable. This can be done using feature selection methods whose aim is to reduce the number of features by selecting the most eligible ones [40]. The kth selected feature at time step n is denoted as $SF_k^n \quad k = 1, 2, \ldots K$.

Filter-based and wrapper-based methods are the two main categories of supervised feature selection methods. Filter-based methods sort features based on their statistical measures, independent of the machine learning algorithm. Neighbourhood component analysis (NCA) is a non-parametric feature selection method whose goal is to maximise prediction accuracy of the classification task [41]. On the other hand, wrapper-type methods evaluate the classification algorithm for different subsets of the features and follow a greedy search approach to select the features that



(a) Overall structure



(b) Feature extraction from each phase current; p can be either a, b, or c.

FIGURE 4 Pictorial description of the proposed method

maximise the performance, according to a performance index.

Filter-based feature selection methods are of limited performance; but wrapper-type methods suffer from huge complexity [42], in particular when a deep model is used for classification. Hence, a new hybrid feature selection method is proposed in this section that copes well with the proposed fault detection approach. In this method, the NCA feature selection technique is first employed as a pre-processing step to sort the features according to their importance. This step can be regarded as a coarse search in which features are scored according to their statistical characteristics. In the second step, the wrapper approach is adopted for fine search, where instead of searching among all features, the search area is limited to a pre-defined number of features, say M. In particular, the selected subset is sequentially augmented where at each iteration, M highest-scored entities among the remaining features are individually included into the selected subset and one that brings about the highest improvement in the classification performance is selected. This strategy can attain a near-optimal solution, with affordable complexity.

In summary, in the proposed ITF detection method, the following procedure should be repeated over each stator current signal to extract all statistical features:

- Decompose each signal into its 5 dominant IMFs, using the CEEMD algorithm.
- Chop each IMF into 10 sub-frames of equal lengths, denoted by L.
- Calculate 10 statistical features, described in Table 2, from each sub-frame.

The successive 10 feature vectors, each containing 150 statistical features, can serve as the sequential inputs to the

TABLE 1 Statistical features extracted from IMFs

Index	Feature	Definition
0	Mean	$\bar{x} = \frac{1}{L}\sum_{l=1}^{L} x(l)$
1	RMS	$x_{rms} = \sqrt{\frac{1}{L}\sum_{l=1}^{L} x^2(l)}$
2	Skewness	$s = \frac{\frac{1}{L}\sum_{l=1}^{L}(x(l)-\bar{x})^3}{\left(\sqrt{\frac{1}{L}\sum_{l=1}^{L}(x(l)-\bar{x})^2}\right)^3}$
3	Kurtosis	$k = \frac{\frac{1}{L}\sum_{l=1}^{L}(x(l)-\bar{x})^4}{\left(\frac{1}{L}\sum_{l=1}^{L}(x(l)-\bar{x})^2\right)^2}$
4	Zero-crossing	Number of times x crosses the zero value
5	Shape factor	$sf = \frac{x_{rms}}{mean(\lvert x\rvert)}$
6	Crest factor	$cf = \frac{\max(x)}{x_{rms}}$
7	Impulse factor	$if = \frac{\max(\lvert x\rvert)}{mean(\lvert x\rvert)}$
8	Margin factor	$mf = \frac{\max(\lvert x\rvert)}{mean(\sqrt{\lvert x\rvert})^2}$
9	Entropy	$en = -\sum_{l=1}^{L} p(x(l))\log_2(p(x(l)))$, where the probability $p(\cdot)$ is represented by the normalised histogram counts.

TABLE 2 Data description

Quantity	Value
Fault severity (number of shorted turns)	0, 1, 2, 5, 7 and 15
Generator speed (rpm)	1350, 1500 and 1650
Active power injected to the grid (W)	0, 1000 and 2000
Used signal	I_{as}, I_{bs} and I_{cs}

LSTM. However, it is proposed to extract and use only a subset of these 150 features, over each sub-frame. These features are selected based on the proposed feature selection method that can be summarised as follows:

- **Coarse Search:** Employing the 150-feature vectors extracted from all training sub-frames, sort these 150 features using the NCA feature selection technique.
- **Fine Search:** Starting from an empty subset, apply the wrapper approach to sequentially augment the selected subset; but instead of searching among all features, the search area is limited to a pre-defined number of features.

Note that the feature selection step is applied once in advance, to specify the most informative features that should be extracted.

4 | EXPERIMENTAL TESTS

4.1 | Dataset

For effective evaluation of the proposed detection method, a DFIG test bed has been utilised, whose experimental configuration is depicted in Figure 5. As is usual in DFIGs, the stator of the wound rotor induction machine is directly connected to the grid by means of a three-phase transformer, while the rotor windings are fed by a rotor-side converter. The DFIG is mechanically coupled to an induction motor supplied by a variable speed drive. In this way, the shaft speed is appropriately set by the drive. The DFIG is controlled by vector control strategy which has been discussed in ref. [14]. Based on this control system, the value of injected active and reactive power to the grid could be adjusted. The whole framework of the control

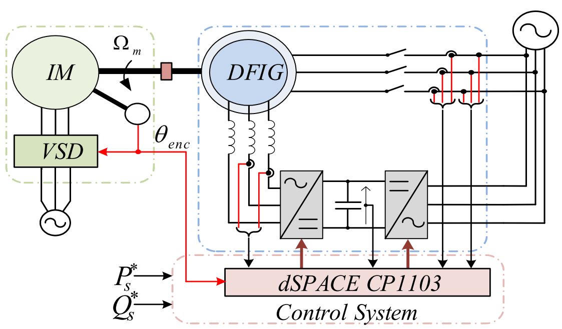
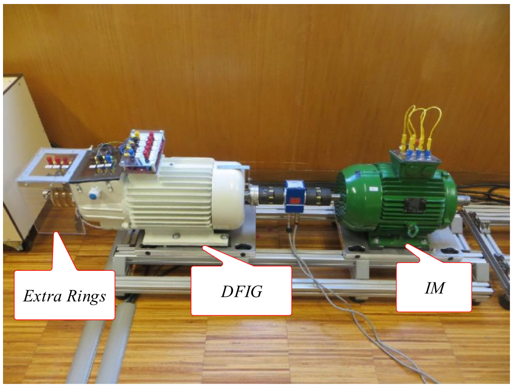


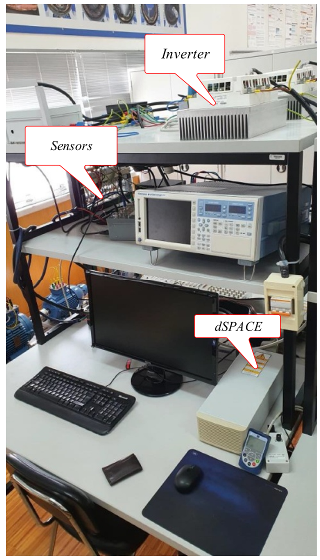
FIGURE 5 Schematic diagram of the experimental Doubly-Fed Induction Generator (DFIG) system

TABLE 3 Parameters of the Doubly-Fed Induction Generator (DFIG) used in the experimental tests

Quantity	Value	Quantity	Value
Rated power (kW)	4	Rated frequency (Hz)	50
Rated stator voltage (V)	400	Rated rotor voltage (V)	230
Rated stator current (A)	9.4	Rated rotor current (A)	11.5
Rated speed (rpm)	1420	Number of poles	40
Inertia (kgm^2)	0.15	Friction coefficient (Nms/rad)	0.03



(a) DFIG and IM



(b) Control setup and inverter

FIGURE 6 Overall control platform and measurement system

system and gating signals computation was implemented in the dSPACE 1103 platform.

Hall effect current sensors measure the stator and rotor currents, while voltage sensors measure the stator voltages and the dc-link voltage of the rotor-side converter (based on a 3-phase Semikron inverter bridge $SKiiP132GD120-3DUL$). An incremental encoder of $2048ppr$ measures the rotor position. The main parameters of the DFIG used in the experimental tests are listed in Table 3. The test bed accompanied by the overall control platform and the rotor side converter are illustrated in Figure 6. The stator inter-turn fault is introduced by providing some taps connected to different turns of one of the stator coils. By selecting two taps and their interconnection, various fault severities are achieved.

The described test bench was then utilised for our investigation. In the current study, different fault severities (0, 1, 2, 5, 7, and 15 shorted turns) at several values of injected powers and shaft speeds was considered, in steady-state conditions. The shaft speed was determined by the induction motor drive adjustment and three speeds were selected, each belonging to a special mode of DFIG operation; 1350 rpm for sub-synchronous mode (the rotor circuit absorbs active power from the power converter), 1500 rpm for synchronous mode (the rotor circuit has almost no active power exchange with the power converter) and 1650 rpm for super-synchronous mode (the rotor circuit delivers active power to the power converter). Furthermore, 3 active power values were selected for injection into the grid (0, 1000, and 2000 W). The amount of exchanged reactive power is chosen to be zero, as is usually the case in the normal condition of the grid. Different signals have been measured, but the stator three-phase currents was regarded for the current study. The mentioned data have been acquired at a sampling frequency of 18 kHz. These specifications are summarised in Table 2.

4.2 | Test settings

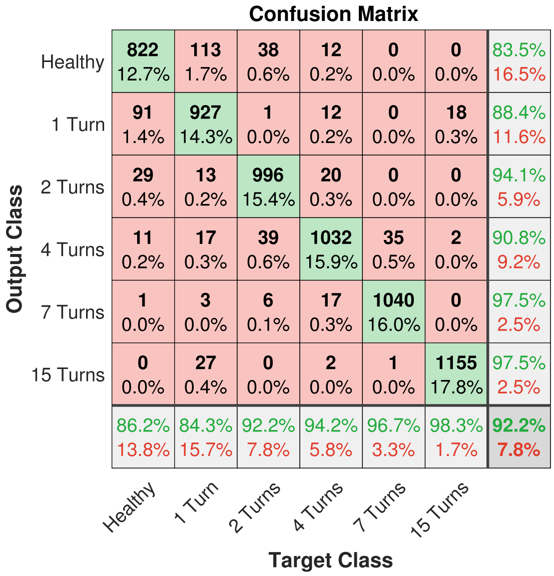
All tests were performed using MATLAB 2019*b* in $64-$bit Windows OS with 6 GB memory on either a 2.2 GHz Intel CPU while only one core is enabled or an 810 MHz NVIDIA GPU with 384 cores.[1]

In all tests that are reported in this section, each signal is split into segments of length 360 points, corresponding to a cycle of the periodic current signals, and each segment is regarded independently as a data sample. Furthermore, a one-layer LSTM with 150 hidden units is used that is trained using the Adam optimisation algorithm in mini-batches of 144 training samples.

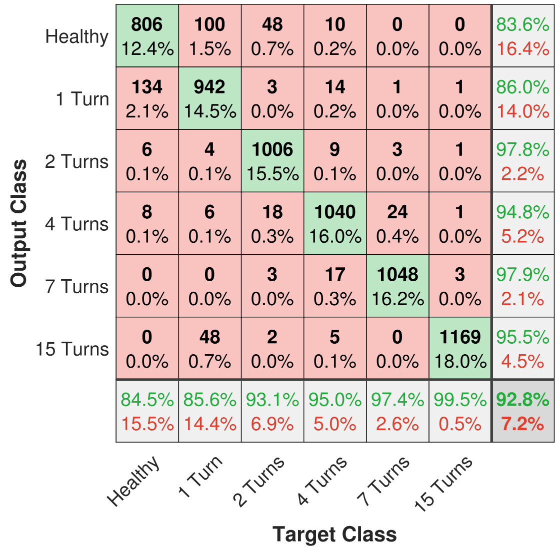
4.3 | Results

At first, feature selection is not involved, that is, all extracted features are applied to the classifier. The confusion matrix obtained by training the LSTM model over all extracted features using 300 epochs is depicted in Figure 7. This test was conducted using hold-out cross-validation, where 70% of data samples were used for training and the remaining samples were held out for the test. As a comparison, the accuracy obtained by the LSTM model is compared, in Table 4, with those of the most successful traditional shallow classifiers; including Linear Discriminant Analysis, K-Nearest Neighbours (KNN), Naive Bayes, SVM, Decision Tree and Multi-Layer Perceptron. This test was conducted using 10-fold cross-validation which is a standard choice in machine learning. Furthermore, the shallow

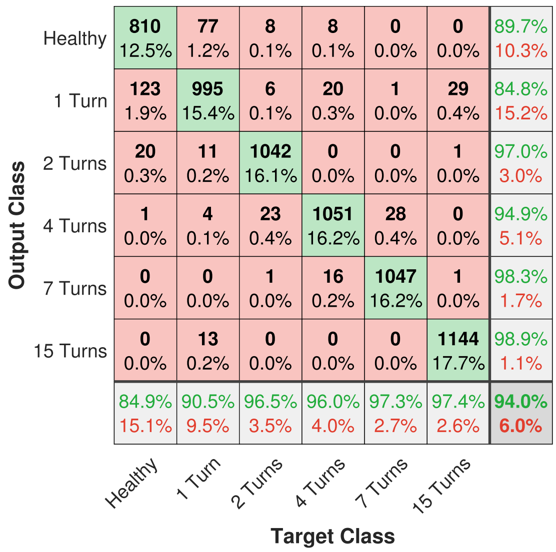
[1] Simulation codes are available at https://github.com/G-Alipoor/DFIG_ITF.



(a) Using all 150 features



(b) Using 6 most informative features



(c) Using 21 most informative features

FIGURE 7 Confusion matrix of the classification task; effect of selecting a limited number of features based on the proposed hybrid feature selection method

TABLE 4 Averaged accuracy employing all features

Method	LSTM	LDA	KNN	Bayes	SVM	DT	MLP
Accuracy	92.16%	62.62%	67.02%	33.75%	18.52%	80.63%	71.92%

Abbreviations: DT, Decision Tree (DT); KNN, K-Nearest Neighbours; LDA, Linear Discriminant Analysis; LSTM, Long Short-Term Memory; MLP, Multi-Layer Perceptron; SVM, Support Vector Machine.

binary classification models were extended to the multi-class problem using the error-correcting output code technique [43].

These results demonstrate the higher accuracy obtained by the LSTM model. Moreover, one can note that distinguishing between two classes becomes more difficult as the difference in the number of shorted turns in these classes decreases. For instance, discriminating between the healthy and the 1-turn cases is more difficult than discriminating between the 15−turns case and any of the other cases. It is in accordance with the expectation since the higher the number of shorted turns is, the more the measured signals are affected by the fault. This in turn causes the sensitivity for the 1−turn class, calculated in the one-against-all-approach (that is the ratio of all cases that were truly predicted as the 1−turn case to all cases that were actually in the 1−turn case) to be relatively small, because this

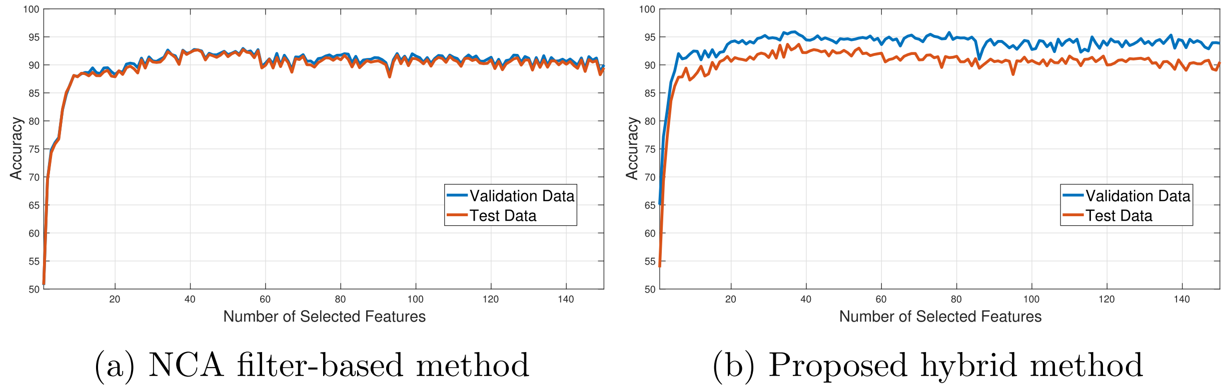


FIGURE 8 Classification accuracy as a function of the number of selected features

TABLE 5 The most informative features selected by the proposed feature selection method

Order	Features	Order	Features	Order	Features
1	F_{c52}	8	F_{b54}	15	F_{b31}
2	F_{b59}	9	F_{c56}	16	F_{a56}
3	F_{c59}	10	F_{b40}	17	F_{b49}
4	F_{a59}	11	F_{b41}	18	F_{c34}
5	F_{b50}	12	F_{c40}	19	F_{a52}
6	F_{c51}	13	F_{b52}	20	F_{c11}
7	F_{b51}	14	F_{b11}	21	F_{b20}

case is more prone to be misclassified as one of the two close cases of healthy and 2−turns. In addition, from a physical point of view, the impedance decrease of the winding caused by one-turn inter-turn is nearly compensated by added impedance value owing to the extra terminal and cabling. So, the stator current is less affected than what would be expected.

To investigate the effect of utilising the feature selection step on the performance of the proposed method, a test was carried out in which the classification accuracy is compared against the number of selected features. In this test, data samples were first split into training and test sub-sets that contain 80% and 20% of available data samples respectively. Then 4 − fold cross-validation was performed on the training set, that is, in each fold 60% of data samples were used for training and the other 20% samples were used for validation. This specific cross-validation makes test and validation sub-sets of equal size.

Results achieved using NCA and proposed selection algorithms are shown in Figure 8, in which the accuracy achieved based on validation and test sub-sets are reported. This test reveals the outperformance of the proposed feature selection method. Particularly, one can see that 6 features carry the most information required for classification and the best accuracy is achieved when 21 features are used. The averaged classification accuracies, over the test data set, with 6 and 21 most informative features selected by the proposed hybrid feature selection method are 87.80% and 91.30%, respectively; while these results are 82.10% and 88.94% for the NCA method. The 21 most informative features selected by the proposed feature selection method along with their descriptions are tabulated in Table 5. Confusion matrixes of the proposed method, when the first 6 and 21 features selected by the proposed methods are utilised, are also shown in Figure 7. These results show that discarding the non-informative features, despite decreasing the complexity, can improve the accuracy of the proposed fault detection method. Moreover, better results can be achieved using the proposed hybrid feature selection method.

Finally, the averaged accuracies of the proposed method in comparison with the conventional shallow machine learning models, when only the 6 or the 21 selected features are used, are summarised in Table 6. These tests are also obtained by 10−fold cross-validation, similar to the results reported in Table 4. It can be seen that our proposed method results in substantially higher accuracy, while its complexity is considerably lowered by decreasing the data dimensionality using the proposed feature selection method. More importantly, the proposed method can detect the fault at its initial stages where only one turn is shorted. This outcome confirms the high performance of the proposed method for incipient detection of stator winding ITF in the DFIG.

5 | CONCLUSION

A new machine learning-based method for incipient detection of stator ITF in DFIGs using the stator current signals was proposed in this paper. To this end, the above problem was considered as a multi-class classification problem with which it is possible to detect the fault incipiently with only one turn is shorted. The classification was carried out by a state-of-the-art deep model, known as the LSTM network, that can better cope with the temporal evolution of signals characteristics. This model is further improved by equipping it with the EMD signal processing tool as well as extracting the most representative features for which a new feature selection method was also presented.

TABLE 6 Averaged accuracy employing the 6 or the 21 most informative features

Method		Proposed	LDA	KNN	Bayes	SVM	DT	MLP
Accuracy	6 features	93.88%	37.05%	60.15%	35.33%	60.68%	65.60%	54.61%
	21 features	94.59%	43.96%	51.18%	35.16%	61.75%	73.07%	60.48%

Abbreviations: DT, Decision Tree (DT); KNN, K-Nearest Neighbours; LDA, Linear Discriminant Analysis; MLP, Multi-Layer Perceptron; SVM, Support Vector Machine.

Tests conducted on experimental data showed the effectiveness of the presented feature selection technique and the high performance of the proposed method for incipient detection of steady-state faults. In particular, the proposed method can detect various stator short-circuit faults with an accuracy of about 95%, even in the case of only one shorted turn. The proposed method is very general and can be applied to any similar problem, for example, detecting other electrical faults and even detecting the faulted phase. The only necessity is to apply the proposed feature extraction and feature selection methods to data samples collected for the new problem and then train the LSTM network on the training data samples. It is worth noting that the selected features may be different for other problems.

AUTHOR CONTRIBUTION

Ghasem Alipoor: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualisation, Writing – original draft, Writing – review & editing. **Seyed Jafar Mirbagheri:** Investigation, Methodology, Software. **Seyed Mohammad Mahdi Moosavi:** Data curation, Validation, Writing – review & editing. **Sérgio M.A. Cruz:** Data curation, Validation, Writing – review & editing.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICTS OF INTEREST

None of the authors have a conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

Data is available on request from the corresponding author.

ORCID

Ghasem Alipoor https://orcid.org/0000-0003-1971-2828
Sérgio M. A. Cruz https://orcid.org/0000-0002-9651-8925

REFERENCES

1. Irena, A.: Renewable capacity highlights. In: Proc. Int. Renew. Energy Agency (IRENA), pp. 1–3 (2021)
2. Agency, I.E.: Global Energy Review 2019. OECD Publishing (2020)
3. Hahn, B., Durstewitz, M., Rohrig, K.: Reliability of wind turbines. In: Wind Energy, pp. 329–332. Springer (2007)
4. Zhang, Z., et al.: High-power generators for offshore wind turbines. Energy Proc. 35, 52–61 (2013). https://doi.org/10.1016/j.egypro.2013.07.158
5. Bell, R.N., et al.: Report of large motor reliability survey of industrial and commercial installations, part I. IEEE Trans. Ind. Appl. IA-21(4), 853–864 (1985)
6. Albrecht, P., et al.: Assessment of the reliability of motors in utility applications - updated. IEEE Trans. Energy Convers. EC-1(1), 39–46 (1986). https://doi.org/10.1109/tec.1986.4765668
7. Djurovic, S., Williamson, S., Renfrew, A.: Dynamic model for doubly-fed induction generators with unbalanced excitation, both with and without winding faults. IET Electr. Power Appl. 3(6), 171–177 (2009). https://doi.org/10.1049/iet-epa.2008.0054
8. Faiz, J., et al.: Magnetic equivalent circuit modelling of doubly-fed induction generator with assessment of rotor inter-turn short-circuit fault indices. IET Renew. Power Gener. 10(9), 1431–1440 (2016). https://doi.org/10.1049/iet-rpg.2016.0189
9. Hichem, M., Tahar, B.: Fuzzy monitoring of stator and rotor winding faults for DFIG used in wind energy conversion system. Int. J. Model. Ident. Control. 27(1), 49–57 (2017). https://doi.org/10.1504/ijmic.2017.082485
10. Mojallal, A., Lotfifard, S.: DFIG wind generators fault diagnosis considering parameter and measurement uncertainties. IEEE Trans. Sustain. Energy. 9(2), 792–804 (2018). https://doi.org/10.1109/tste.2017.2761842
11. He, S., Shen, X., Jiang, Z.: Detection and location of stator winding interturn fault at different slots of DFIG. IEEE Access. 7, 89342–89353 (2019). https://doi.org/10.1109/access.2019.2926538
12. Fu, Y., et al.: Using flux linkage difference vector in early inter-turn short circuit detection for the windings of offshore wind DFIGs. IEEE Trans. Energy Convers. 36(4), 3007–3015 (2021). https://doi.org/10.1109/tec.2021.3073007
13. Cheng, J., et al.: Stator inter-turn fault analysis in doubly-fed induction generators using rotor current based on finite element analysis. In: 2018 IEEE International Conference on Progress in Informatics and Computing (PIC), pp. 414–419 (2018)
14. Moosavi, S., et al.: Comparison of rotor electrical fault indices owing to inter-turn short circuit and unbalanced resistance in doubly-fed induction generator. IET Electr. Power Appl. 13(2), 235–242 (2019). https://doi.org/10.1049/iet-epa.2018.5528
15. Shah, D., Nandi, S., Neti, P.: Stator-interturn-fault detection of doubly-fed induction generators using rotor-current and search-coil-voltage signature analysis. IEEE Trans. Ind. Appl. 45(5), 1831–1842 (2009). https://doi.org/10.1109/tia.2009.2027406
16. Bilal, H., Heraud, N., Sambatra, E.J.R.: An experimental approach for detection and quantification of short-circuit on a doubly-fed induction machine (DFIM) windings. J. Control Automation Electrical Syst. 32(4), 1–8 (2021). https://doi.org/10.1007/s40313-021-00733-w
17. Thomson, W., Fenger, M.: Current signature analysis to detect induction motor faults. IEEE Ind. Appl. Mag. 7(4), 26–34 (2001). https://doi.org/10.1109/2943.930988
18. Rebouças Filho, P.P., et al.: A reliable approach for detection of incipient faults of short-circuits in induction generators using machine learning. Comput. Electr. Eng. 71, 440–451 (2018). https://doi.org/10.1016/j.compeleceng.2018.07.046
19. Swana, E.F., Doorsamy, W.: Investigation of combined electrical modalities for fault diagnosis on a wound-rotor induction generator. IEEE Access. 7, 32333–32342 (2019). https://doi.org/10.1109/access.2019.2904238
20. Vinayak, B.A., Anand, K.A., Jagadanand, G.: Wavelet-based real-time stator fault detection of inverter-fed induction motor. IET Electr. Power Appl. 14(8), 82–90 (2020). https://doi.org/10.1049/iet-epa.2019.0273
21. Ahouee, R.A., Mola, M.: Inter-turn fault detection in PM synchronous motor by neuro-fuzzy technique. Int. J. Syst. Assurance Eng. Manag. 11(5), 923–934 (2020). https://doi.org/10.1007/s13198-020-01019-1

22. Karatzinis, G.D., Boutalis, Y.S., Karnavas, Y.L.: An accurate multiple cognitive classifier system for incipient short-circuit fault detection in induction generators. Electr. Eng. 104(3), 1–16 (2021). https://doi.org/10.1007/s00202-021-01445-9
23. Ehya, H., Skreien, T., Nysveen, A.: Intelligent data-driven diagnosis of incipient inter-turn short circuit fault in field winding of salient pole synchronous generators. IEEE Trans. Ind. Inf. 18(5), 3286–3294 (2021). https://doi.org/10.1109/tii.2021.3054674
24. Pietrzak, P., Wolkiewicz, M.: On-line detection and classification of PMSM stator winding faults based on stator current symmetrical components analysis and the KNN algorithm. Electronics. 10(15), 1786 (2021). https://doi.org/10.3390/electronics10151786
25. Xu, Y., et al.: Multi-sensor edge computing architecture for identification of failures short-circuits in wind turbine generators. Appl. Soft Comput. 101, 107053 (2021). https://doi.org/10.1016/j.asoc.2020.107053
26. Liu, W., et al.: A survey of deep neural network architectures and their applications. Neurocomputing. 234, 11–26 (2017). https://doi.org/10.1016/j.neucom.2016.12.038
27. Ray, S., Ganguly, B., Dey, D.: Identification and classification of stator inter-turn faults in induction motor using wavelet kernel based convolutional neural network. Elec. Power Compon. Syst. 48(12-13), 1421–1432 (2020). https://doi.org/10.1080/15325008.2020.1854384
28. Skowron, M., et al.: Convolutional neural network-based stator current data-driven incipient stator fault diagnosis of inverter-fed induction motor. Energies. 13(6), 1475 (2020). https://doi.org/10.3390/en13061475
29. Li, Y., et al.: Diagnosis of inter-turn short circuit of permanent magnet synchronous motor based on deep learning and small fault samples. Neurocomputing. 442, 348–358 (2021). https://doi.org/10.1016/j.neucom.2020.04.160
30. Lee, H., et al.: Attention recurrent neural network-based severity estimation method for interturn short-circuit fault in permanent magnet synchronous machines. IEEE Trans. Ind. Electron. 68(4), 3445–3453 (2021). https://doi.org/10.1109/tie.2020.2978690
31. Husari, F., Seshadrinath, J.: Early stator fault detection and condition identification in induction motor using novel deep network. In: IEEE Transactions on Artificial Intelligence, pp. 1–1 (2021)
32. Wu, Y., Ma, X.: A hybrid LSTM-KLD approach to condition monitoring of operational wind turbines. Renew. Energy. 181, 554–566 (2022). https://doi.org/10.1016/j.renene.2021.09.067
33. Huang, N.E., et al.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceed. Royal Soc. London Series A: Math. Phys. Eng. Sci. 454(1971), 903–995 (1998). https://doi.org/10.1098/rspa.1998.0193
34. Wu, Z., Huang, N.E.: Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adapt. Data Anal. 1(01), 1–41 (2009). https://doi.org/10.1142/s1793536909000047
35. Torres, M.E., et al.: A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4144–4147, IEEE (2011)
36. Olah, C.: Understanding LSTM Networks (2015). https://colah.github.io/posts/2015-08-Understanding-LSTMs
37. Hochreiter, S., Schmidhuber, J.: LSTM can solve hard long time lag problems. Adv. Neural Inf. Process. Syst., 473–479 (1997)
38. Staudemeyer, R.C., Morris, E.R.: Understanding LSTM-A Tutorial into Long Short-Term Memory Recurrent Neural Networks. arXiv preprint arXiv:1909.09586 (2019)
39. Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. Nonlinear Phenom. 404, 132306 (2020). https://doi.org/10.1016/j.physd.2019.132306
40. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. 40(1), 16–28 (2014). https://doi.org/10.1016/j.compeleceng.2013.11.024
41. Yang, W., Wang, K., Zuo, W.: Neighborhood component feature selection for high-dimensional data. J. Comput. 7(1), 161–168 (2012). https://doi.org/10.4304/jcp.7.1.161-168
42. Zhang, X., et al.: A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method. Neurocomputing. 275, 2426–2439 (2018). https://doi.org/10.1016/j.neucom.2017.11.016
43. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes. J. Artif. Intell. Res. 2, 263–286 (1994). https://doi.org/10.1613/jair.105

How to cite this article: Alipoor, G., et al.: Incipient detection of stator inter-turn short-circuit faults in a Doubly-Fed Induction Generator using deep learning. IET Electr. Power Appl. 17(2), 256–267 (2023). https://doi.org/10.1049/elp2.12262