STATOR WINDING INTERTURNS SHORT CIRCUIT FAULT DETECTION IN A THREE PHASE INDUCTION MOTOR DRIVEN BY FREQUENCY CONVERTER USING NEURAL NETWORKS

Átila Girão de Oliveira

Universidade Federal do Ceará

Cláudio M. de Sá Medeiros

Instituto Federal de Educação, Ciência e Tecnologia do Ceará

Ricardo Silva Thé Pontes

Universidade Federal do Ceará

Abstract

This work is the application of a Multilayer Perceptron Artificial Neural Network (MLP ANN) to detect early inter turn short-circuit faults in a three phase converter fed induction motor. The quantity used to analyze the problem is the stator current or, more specifically, the harmonic contend of its frequency spectrum, also called current signature. The analysis through the current signature is a non-invasive method and can be embedded in the frequency converter, what is a great advantage. The dataset used to training and to validate the ANN is obtained using a test bench that allows to apply different levels of inter turn short-circuits in the machine. It is observed that the fault motor dataset and healthy motor dataset are difficult to separate due the nonlinear character, which demands a large computational effort to choose an appropriate MLP topology. The MLP is trained by two different algorithms (the classical error Back-Propagation - BP - and an adaptation of the newer Extreme Learning Machine - ELM) and the results are thoroughly explored. Then it is slightly compared with the results of a Self-Organized Map ANN [1] obtained by using the same dataset.

1. Introduction

The induction machine is consolidated as main motor force in industries. According Thomson and Fenger [2], in an industrialized nation the induction motor might consume, typically, among 40% and 50% of all capacity generated.

Despite the recognized robustness and reliability of this machine, it is subject to fault occurrence, many times due to installation environment conditions, inadequate applications and lack of preventive maintenance. The more common occurrences are bearing faults, stator or rotor isolation faults, open bars or crack of the rings and eccentricity fault [3].

Machine faults produce symptoms like unbalanced line voltage and current, increasing in torque pulsation, decreasing in the mean torque, increasing of losses, efficiency decreasing, and excessive heating [4]. Therefore several methods of detection and diagnostic has been developed past years and new solutions keep appearing with the objective to increase the accuracy but also to simplify techniques and to decrease costs [5].

In industries the cost of an unscheduled production downtime is too high; hence they invest increasingly to improve their maintenance programs. Faults like open rotor bars, eccentricity and bearing faults take time to evolve and put the motor out of operation. In this context, the constant online monitoring is important to early fault detection in a way to have time to program a maintenance order and save the machine.

The stator winding inter-turn short-circuit (SWITSC) takes a short time to evolve and to condemn the motor. Thomson and Fenger [2] tested low voltage three-phase induction machines from early SWITSC until complete failure and found that exist a time of a few minutes to occur the fault evolution. This time is probably not enough to avoid unscheduled production downtime. But the early detection makes possible to repair the machine by rewinding it or, in large machines, removing short-circuited coils and the early operation stopping avoids electrical arcs due short-circuit and then offers an

additional protection to areas where there are explosion risks. Moreover, after the severe fault, the ferromagnetic core is damaged and the machine becomes probably irreparable.

It has become more and more common the use of frequency converter to drive induction motors. That gives more versatility to the drive of machines because it allows application with varying rotation speed [6]. The high current in the motor due to fault also affects the frequency converter integrity and an early detection means an additional protection. Moreover, once this electronics is already being in use, could be advantageous have detection systems previously embedded in it. Past years the developing of several computational intelligence techniques added new possibilities to fault detection and diagnosis systems, most of them potentially suitable to be embedded in electronic converters. Examples include ANN, fuzzy systems, genetic algorithms, among others.

Despite the spread of fed-converter machines, most of researches uses line-fed drive machine. Using converter-fed motor, Kowalski and Wolkiewicz [7] analyze the spectrum of instantaneous Park power and torque signals to diagnosis early SWITSC and broken rotor bars. The same authors in [8] use neural networks to diagnosis SWITSC with 80% of accuracy. Coelho and Medeiros in [1] try to map the SWITSC fault using self-organized map and classify the motor.

According Nandi et al [3], isolation fault represents from 30% to 40% of all kinds of faults reported in induction motors. This significant amount justifies the detailed investigation of the problem. In this work is investigated the motor current signature analysis (MCSA) together with the potential of a single hidden-layer feed-forward neural network (SLFN) classifier as a tool of in SWITSC detection to converter-fed three-phase induction machine. Two algorithms were used to training the neural network: the error back-propagation (BP), which is the classical one, and the ELM, which is a more recent algorithm. The datasets were obtained by an experimental test bench where different level of SWITSC can be applied.

The section 2 shows the problem of the SWITSC and how it is been treated past years. Section 3 contains the basic about MLPs and the ELM and BP algorithms. In section 4 the laboratory tests and data acquisition are explained as well as the ANN attributes selection. The classification results are in section 5 and the section 6 concludes this work.

2. Stator Winding Inter-Turn Short-Circuit Overview

Isolation systems are submitted to several kinds of efforts that might cause a failure. Due the use of inverters to drive electrical motors with a typically frequency switching of about 10 kHz, occur voltage peaks such that increases considerably the machine isolation stress. As result, the fed converter motor stress might be until ten times higher than line fed machines [9].

The failure process is usually initialized as a turn-to-turn high impedance fault (order of kΩ) in the same phase, between phases or phase-to-ground [10]. The fault current can reaches two times the rotor blocked current, which causes high localized heating and makes the fault quickly spreads. If the incipient fault was detected it is possible reutilize the machine after repairs, but if the fault evolves might possibly causes an irreparable damage to the machine core [2].

Different methods have been used in many researches to detect stator inter-turn short-circuit. Ballal et al [11] use the symmetric components theory to do the detection. This technique consists in using an expression to separate the currents in positive, negative and zero sequences. They analyze a graphic that the positive and negative sequences describe a circle with opposite spinning direction. The detection is done by a measure of the deformation caused in the graphics by the fault.

Boqiang et al [12] use as characteristic to detection the negative impedance sequence, which is defined as the negative sequence value of the voltage component divided by the negative sequence current component. In experimental tests they realized that there is an oscillation of the impedance value with time and it is necessary a low-pass filter to guarantee the technic reliability.

Considering the MCSA, Joksimovic and Penman [13] show that there are no novel components in stator motor current frequency spectrum due to the isolation fault. In fact it was observed that just occur an increase of the existents components. Stavrou et al [14] search the current frequency spectrum for the variation in frequencies as function of number of poles, slots and slip, that is, specific constructive features.

Penman et al [15] develop an equation (2.1) to calculate which harmonic components in the axial leakage flux waveform are functions of the SWITSC and propose a method to detect fault by monitoring those components.

$$f_{st} = \{k \pm n(1-s)/p\} f_1 \quad (2.1)$$

The f_{st} is inter-turn short-circuit function components, k =1, 3, 5..., means the temporal harmonics order, n = 1, 2, 3..., means the spatial harmonic order, s is the slip, p is the pair of poles, f_1 is the power supply fundamental frequency. Several kinds of faults affect the current spectrum and some harmonics are affected by more than one abnormal condition, so one must be careful to choose the correct frequencies that will be used to indicate the problem.

According Das et al [16] unbalanced supply voltage might produce current signature which look apparently identical to stator winding inter-turn fault cases. They propose a method to separate these two signatures. Their method is based in the Extend Park's Vector Approach (EPVA) and combined with signal processing tool as Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT) and Power Spectral Density (PSD) to make the discrimination.

Thomson and Fenger [2] expand the concept of the leakage flux to stator currents, once the flux inside the machine also crosses the stator windings. They do an experimental analysis in low voltage motors to verify which frequencies are function just of the short-circuit and no other conditions as unbalanced phases, misalignment of the shaft, broken rotor bars, bearing faults, etc. The components found in [2] using equation (2.1) as function only of short-circuit are f_{st1} when $k = 1$, $n = 3$ and $k = 1$, $n = 5$; to an unload motor ($s \approx 0$) with 2 pair of poles, these harmonics would be $2.5f_1$ and $3.5f_1$.

Also using MCSA, Gazzana et al [17] create a system to early detection and diagnosis rotor broken bars, air-gap eccentricities and SWITSC in induction motors. To SWITSC the equation (2.1) is used with $k = 1$, $n = 7$ and the Welch's method is used to obtain the frequency spectrum. The choice of a high order spatial component in the spectrum is because the low order eccentricity components are coincident with short-circuit components.

Based on stator currents Hyun et al [18] create neural models to simulate the state of one induction motor without any faults, one with isolation fault and other with bearing fault. The models are put in parallel with the system and the real output are constantly compared with the neural models outputs. A Bayesian network evaluates the model residues and detects the isolation or bearing fault.

Bouzid et al [19] use a neural network to locate the phase where the short-circuit is. It is chosen as fault feature the phase shift between the three phase voltages and currents. The detection is made by a MLP NN with 3 outputs, each one referent to one phase. If that neuron was active it means a fault in that phase. They validate the method using two induction motors and conclude it is an efficient method and once a NN was trained to one motor it can be used to other identical machines.

Das et al [20] process the line current signal recorded from motor terminals through a Park's transformation followed by Continuous Wavelet Transformation and uses a Support Vector Machine (SVM) to classify the extracted features. From the 18 test cases used for prediction, a total of 16 fault cases were correctly identified by the proper configured SVM.

Among all possible methods to fault detection, current signature has a great potential because it is non-invasive; does not require installation of sensor in the machine; does not require be adapted to classified areas (because it can be installed at the panel, far from potential explosive mixtures); presents high capacity of remote monitoring reducing the maintenance men exposition to risks; can be applied to any machine, with no power restriction; presents sensitivity to mechanical machine faults, stator electric faults and feed problems; among others [21]. To these advantages it can be added, to converter-fed motors, the possibility of embedded the detection system in the own converter, especially if a computational intelligence technique was used.

This work presents a MLP ANN to classification of SWITSC. The equation (2.1) developed by the Penman's theory and expanded by Thompson is initially used for feature extraction to the fault detection.

3. MLP Artificial Neural Network in a Nutshell

The ANNs appeared as mathematics tool based on the human brain biological model. Simplified models based are usually designed to specific problems solutions like classification, pattern matching, pattern completion, optimization, control, function approximation and data mining [22].

It was proved that an MLP ANN with one hidden layer can approximate any continuous function with a determined precision since it has neurons enough [23]. In classification they are recommended for applications in which there is an unknown non-linear relation between input and output dataset, even for complex non-linear multivariable problems. They are capable to learn this relationship by data presentation and then generalize the knowledge and classify new data.

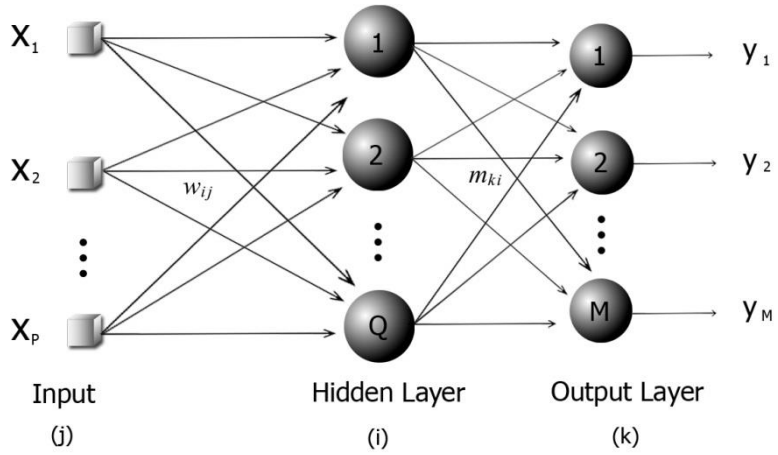

Figure 1. Generic Single-Hidden Layer Feedforward Neural Network.

In Fig. 1 is shown a generic architecture of a SLFN. First there is the input vector, that's fully connected with the hidden layer by the weights w_{ij} just like the synapses that connect biological neurons. The hidden layer uses non-linear function to make a transformation in data space and produces a linear-separable data space. The hidden-layer output is the input to the output layer, where the classification is done.

The MLP ANN trained by BP algorithm (MLP/BP) is probably the most studied and classical neural model, especially in classification applications, but even nowadays, a user soon becomes aware of the difficulties in finding an optimal architecture for real-world applications. An architecture that is too small will not be able to learn from data properly. Otherwise, an architecture having too many hidden neurons are prone to fit too much of the noise on the training data.

The many parameters that need to be adjust by heuristic or more commonly by attempt and error demands much time and effort to design a proper MLP, also the time wasted with the BP training is usually elevated. That led many researchers to look for a new algorithm that overcome the classical.

A new learning algorithm for SLFNs named ELM was presented in [24] and has become the aim of many studies. The two great advantages of the ELM are easy neural network design, which has practically no parameters to adjust and the training algorithm that computes extremely fast.

3.1 Backpropagation

This section describes briefly the most common algorithm used to training the MLP ANN, a detailed version can be find in Engelbrecht's book [22]. The learning algorithm requires two passes of computation: a forward pass and a backward pass. During the forward pass the synaptic weights remain unaltered, while the activations and outputs are computed on a neuron-by-neuron basis. At iteration t, the activation of a hidden neuron (one-hidden-layered MLP) is computed as

$$u_i^{(h)}(t) = \sum_{j=0}^{P} w_{ij}(t)x_j(t), \qquad i = 1, \dots, Q \tag{3.1}$$

where w_{ij} is the weight connecting the input j to the hidden neuron i, Q ($2 \leq Q < \infty$) is the number of hidden neurons and P is the dimension of the input vector (excluding the threshold). For the purpose of simplifying notation, it was set $x_0(t) = -1$ and $w_{i0} = \theta_i^{(h)}(t)$, where $\theta_i^{(h)}(t)$ is the threshold of the hidden neuron i. The output of neuron i is then defined by

$$y_i^{(h)}(t) = \varphi_i\left[u_i^{(h)}(t)\right] = \varphi_i\left[\sum_{j=0}^{P} w_{ij}(t)x_j(t)\right] \tag{3.2}$$

Where $\varphi_i(\cdot)$ is usually a sigmoidal function. Similarly, the output values of the output neurons are given by

$$y_k^{(o)}(t) = \varphi_k\left[u_k^{(o)}(t)\right] = \varphi_k\left[\sum_{i=0}^{Q} w_{ij}(t)x_j(t)\right] \tag{3.3}$$

where m_{ki} is the weight connecting the hidden neuron i to the output neuron k ($k = 1, \dots, M$), and $M \geq 1$ is the number of output neurons. For the same purpose, it was set $y_0(t) = -1$ and $m_{k0} = \theta_k^{(o)}(t)$, where $\theta_k^{(o)}(t)$ is the threshold of the output neuron k. The backward pass starts at the output layer by propagating the error signals from it towards the hidden layer. For that, first is computed the error value $e_k^{(o)}(t)$ generated by each output neuron at time step t

$$e_k^{(o)}(t) = d_k(t) - y_k^{(o)}(t), k = 1, \dots, M \tag{3.4}$$

where $d_k(t)$ is the target output value for the output neuron k. Due to the sigmoidal nonlinearity, the error signal $e_k(t)$ should be multiplied by the derivative $\phi_k'\left[u_k^{(o}(t)\right] = \partial\phi_k/\partial u_k^{(o)}$ before being back-propagated, thus generating the so called *local gradient* of the output neuron k

$$\delta_k^{(o)}(t) = \phi_k'\left[u_k^{(o)}(t)\right]e_k^{(o)}(t) \tag{3.5}$$

Similarly, the local gradient $\delta_i^{(h)}(t)$ of the hidden neuron i is then computed as

$$\begin{aligned} \delta_i^{(h)}(t) &= \phi_i'\left[u_i^{(h)}(t)\right]\sum_{k=1}^{M} m_{ki}(t)\delta_k^{(o)}(t) \\ &= \phi_i'\left[u_i^{(h)}(t)\right]e_i^{(h)}(t), \quad i = 0, \dots, Q \end{aligned} \tag{3.6}$$

where the term $e_i^{(h)}(t)$ plays the role of a *back-propagated* error signal for the hidden neuron i. Finally, the synaptic weights of the output neurons are updated according to the following rule

$$m_{ki}(t+1) = m_{ki}(t) + \eta\delta_k^{(o)}(t)y_i^{(h)}(t), \quad i = 0, \dots, Q, \tag{3.7}$$

where $0 < \eta \ll 1$ is the learning rate. The weights of the hidden neurons are adjusted through a similar learning rule

$$w_{ij}(t+1) = w_{ij}(t) + \eta\delta_i^{(h)}(t)x_j(t), \quad j = 0, \dots, P. \tag{3.8}$$

One complete presentation of the entire training set during the learning process is called an *epoch*. Many epochs may be required until the convergence of the BP algorithm is verified. Thus, it is good practice to randomize the order of presentation of training examples from one epoch to the next, in order to make the search in the weight space stochastic over the learning cycles.

A simple (and naive) way of evaluating convergence is through the average squared error

$$\varepsilon_{train} = \frac{1}{2N}\sum_{t=1}^{N}\sum_{k=1}^{M}\left[d_k(t) - y_k^{(o)}(t)\right]^2, \tag{3.9}$$

computed at the end of training run using the training data vectors. If it falls below a pre-specified value then convergence is achieved. The generalization performance of the MLP should be evaluated on a testing set, which contains examples not seen before by the network.

3.2 ELM

The ELM algorithm was proposed in 2004 by Guang-Bin et al [24] as an attractive option that can be used to training SLFNs instead classical gradient descent-based methods like the BP. The authors prove that the proposed algorithm can typically training any dataset thousands times faster than the error BP. It is possible because it was proved in [25] that SLFNs with arbitrarily assigned input weights and hidden layer biases and with almost any nonzero activation function can universally approximate any continuous functions on any compact input sets.

These researches results imply that in the applications of feed-forward neural networks input weights may not be necessarily adjusted at all. They just assume that the hidden-layer weights are chosen arbitrarily and can make a non-linear transformation in data space that makes the hidden-layer output as a linear system and therefore can be solved by a simple generalized inverse operation of the hidden-layer output matrices [24]. It follows a possibility to the ELM algorithm:

The equations (3.1) and (3.2) can be expressed in a matrix-vector notation respectively as (3.10) and (3.11),

$$u(t) = Wx(t) \tag{3.10}$$

$$y^{(h)}(t) = \varphi_i(u_i(t)) = \varphi_i(Wx(t)) \tag{3.11}$$

where the function $\varphi_i(\cdot)$ is applied to each one of the Q components of vector $u(t)$. The vector $y^{(h)}(t)$ is calculated to each dataset sample and organized in a matrix $Y^{(h)}$ with Q (number of hidden neurons) lines and N (number of samples) rows. This matrix is used to calculate the output weights.

To each input vector $x(t), t = 1, \dots, N,$ there exists a target output vector $d(t)$. The N target output vectors can be organized in a matrix with M (number of output neurons) lines and N rows.

$$D = [d(1) \mid d(2) \mid \cdots \mid d(N)]_{MxN} \tag{3.12}$$

The calculation of the output weights can be considered as the calculation of a linear mapping between the output layer and the hidden layer. That means it searches the matrix M that better represents the transformation of the input vectors $x(t)$ to its correspondent target vector $d(t)$

$$d(t) = My^{(h)}(t) \tag{3.13}$$

which can be done by the least-square error method, also known as pseudo-inverse method. The expression is given by

$$M = DY^{(h)^T}(Y^{(h)}Y^T)^{-1} \tag{3.14}$$

The same way back-propagation, the average squared error can be used to evaluating convergence.

4. Experimental Data Acquisition

To collect the dataset needed to training the ANNs, an induction motor was rewound by a specialized company. The motor is a WEG with rated values: $0{,}75\ kW\ (1.0\ CV), 60\ Hz, 220/380\ V, 3.02/1.75\ A, n = 1720\ rpm, efficiency: 79.5\%, F.P. = 0.82$. Originally each phase is composed by 2 groups of 3 concentric windings, each one with 58 coils. After rewinding, coil derivations of one group of each one of the three phases were let outside the motor frame the way to allow apply SWITSC. In Figure 2 (b) it is shown the derivations details. A frequency converter WEG CFW-09 was used to drive the motor and a Foucault's brake - Figure 2 (a) - was used to apply load. The data was acquired by a data

acquisition system Agilent U2352 with 16 bits of resolution, passing through a 1 kHz analog filter and a signal amplifier; that implies the band used to work is limited to 500 Hz by the Nyquist's[1].

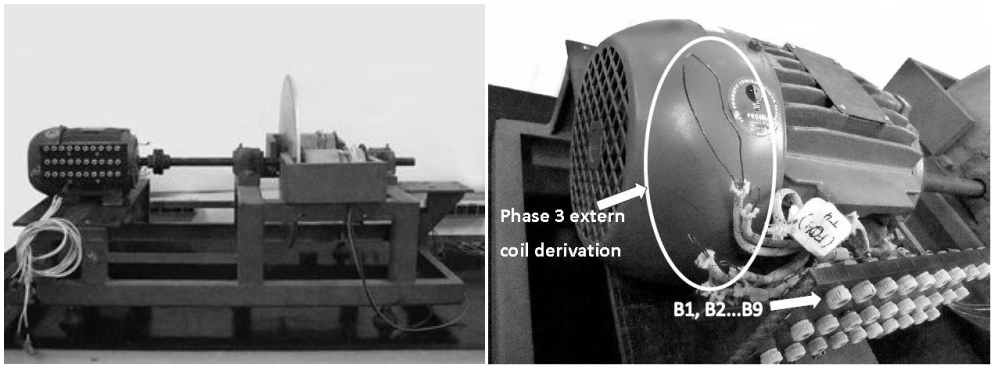

Figure 2. (a) Induction motor extern derivations details and bornes and (b) Coupling Motor-Load (Foulcault's brake).

The three phase current signals are measured by hall sensors and collected with a frequency of sample of 10 kHz during 10 seconds. That creates a dataset of 100,000 samples to each phase.

The motor is delta connected. To the acquisition the frequency converter is set to seven different values: 30 Hz, 35 Hz, …, 60 Hz. Furthermore, three load conditions are considered: unload motor, 50% rated load motor and 100% rated load motor.

To cover a considerable range of the SWITSC fault it was defined two kinds of short circuit: high impedance (HI) and low impedance (LI). The first one emulates the incipient SWITSC and the second one emulates a more severe fault. In the Figure 3 it is shown the simplified scheme of each kind of fault. In Figure 3 (a) the resistor creates a parallel way with high impedance to the current flows, similar to the first stage of short-circuit. In Figure 3 (b) the resistor limits at nominal the current that flows by electromagnetic induction in group 2, but the impedance between groups 1 and 3 is low.

One of the three phases is chosen to be short-circuited to the data acquisition tests. Three crescent levels of inter-turn short- circuits are applied both in HI and LI. Represented as the percentage of the total number of stator windings in that phase, the levels are approximately 1.41%, 4.81% and 9.26%. In the following text the HI fault might be referred as HI1, HI2 and HI3 that means HI with 1.41%, with 4.81% and with 9.26% of short circuited windings respectively. The same is valid to LI fault.

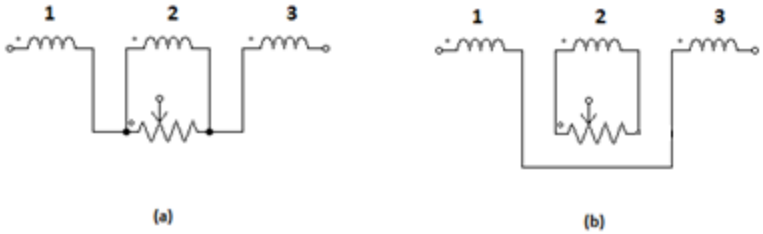

Figure 3. Emulation scheme of: a) high impedance and b) low impedance.

4.1 Datasets

Two mainly datasets created are the healthy and the fault datasets, but is important to emphasize that the fault condition can be subdivided in HI fault and LI fault. The HI and LI can once more be subdivided according the level of inter-turn short circuit (HI1, HI2, HI3, LI1, LI2, LI3).

[1] Sample Nyquist's Theorem: considering $x_c(t)$ a limited signal in band with its frequency domain version given by $X_c(j\Omega) = 0$ to $|j\Omega| \geq \Omega_N$, then $x_c(t)$ is only determined by its samples $x(n) = x_c(nT), n = 0, \pm 1, \pm 2, \ldots$ if $\Omega_s = 2\pi/T \geq 2\Omega_N$. The $\Omega_N$is usually defined as the Nyquist frequency ['OPPENHEIM et al., 1999']

The normal condition and each one of the subdivided fault (HI1, HI2, HI3, LI1, LI2, LI3) contains 100,000 time domain samples to each phase. The spectrum in frequency domain is obtained by the Fast Fourrier Transform (FFT). From the each spectrum some multiples are chosen as neural networks input attributes.

Due there is redundant information given by the two phases directly connect with the short circuit, one of them is unnecessary and can be excluded. Then the phase B current is not used in the datasets which gives a total amount of samples equal to 378 (2 classes x 2 currents x 7 frequencies x 3 loads x 2 kinds of fault x 3 levels of fault). The normal conditions dataset has 42 samples (2 currents x 7 frequencies x 3 loads), and the fault condition has 336 samples, which 168 are HI and 168 LI.

To a practical implementation of the classifier, only one sensor in anyone of the three phases would be necessary to detection. That is possible because the ANN is trained with the characteristics of the phase directly connected to the short-circuit as well as with the phase not directly connected to it.

4.2 Attributes Selection

The equation (2.1) is used to make a pre-selection of the harmonics used as ANN attributes. Considering the $s = 0$ and known that $p = 2$, the spectrums when $k = 1$ and $n = 1, 2, 3, 4, 5, ...$ obtained from equation (2.1) are: $0.5f_1$, $1f_1$, $1.5f_1$, $2f_1$, $2.5f_1$, $3f_1$, …. But the slip will never be zero; it depends on the load and also of the drive frequency. So an algorithm is used to get the amplitude of approximated harmonics: those harmonics are used as a central point and a search for the maximum amplitude value around each one of them is done. The range to the search is $\pm s$. Due the band limit of 500 Hz, it is done until 8th harmonic, which gives 16 pre-selected parameters.

Summarizing, the pre-selected attribute vector contain 16 values that are the approximated spectrum given by equation (2.1). To reduce this number, it was done a statistic analysis (**Erro! Fonte de referência não encontrada.**) of the variance of the pre-selected attributes considering all the conditions used (load, frequencies, kind of faults, level of faults). The attributes are reduced to 0.5f; 1f; 1.5f; 2f; 3f; 5f; 7f, but it was not the final ones. It was included also 2.5f, 3.5f because these are the approximated harmonics to a 4 pole motor according [2] that gives information about short-circuit, but after testing each component relevancy of these harmonics to the ANN, the final attributes chosen were 0,5f; 1,5f; 2,5f; 3f; 5f; 7f.

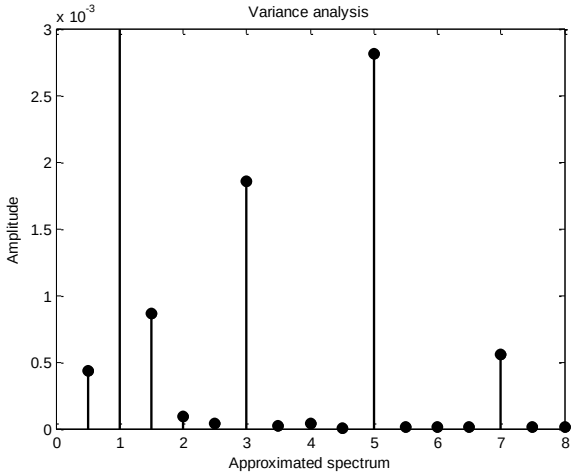

Figure 4. Variance analysis to attributes selection.

5. Results

During the training of the ANNs 90% of data samples were used. The 10% remained composed the validation dataset. However, the training dataset was equally divided to each class in a way to avoid tendentious classification. Consequently, due the healthy dataset has 42 samples and the faulty condition has 336 samples, great part of the faulty dataset are not used during the training, so these data are added to validation dataset.

There are no formulas to select the BP parameters, so all the parameters used were selected after several attempts. At the end the hidden-layer was set to 5 neurons. The learning rate used decreases exponentially with the number of epochs until a final value. Also it was used a momentum term of 0.8. It was implemented the early stop, that uses a test dataset to stop the training when the generalizing error is growing. Thus the datasets has to be divided to training, testing and validation. Respectively, they were set to 70%, 20% and 10% of the total. To the ELM, the only parameter to be adjusted is the number of hidden-layers that was set to 20 after many experiments.

In Both ANNs the activation functions are hyperbolic tangents and the datasets were normalized by two ways: i) removing the average and dividing by variance and ii) by adjusting the values between -1 and +1.

To evaluate a classifier performance it is commonly used the classification rate, that is given by the numbers of vectors correctly classified divided by the total numbers of all classes vectors. Many researchers usually evaluate the validation dataset only, i.e. the generalization capability with new data. But many samples are used during the training and then the classification rate found to training dataset is considered important.

Moreover, using simple global classification rate do not give an evaluation about individual performance of each class classification, then is common to complement the performance evaluation by the use of a confusion matrix CM. The diagonal of the CM has the classification rate to each class.

The ANNs and its algorithms are programed using the software MATLAB version 7.11. In Table 1 is shown the average results found after 50 trainings and its standard deviation. Some abbreviations are used: CR for classification rate; the subscripts TR, TS and VAL refer to training, test and validation datasets; respectively; σ for standard deviation; Nh for the number of hidden-neurons; N_W for the total number of weights.

To attest the non-linear separability of the datasets it was included a test with the simple perceptron which is a linear neural classifier. In Table 2 is shown the Confusion Matrix Average CM.

Table 1. Average results of the ANNs.

ANN	Nh	N_W	CR_{TR}	σ_{TR}	CR_{TS}	σ_{TS}	CR_{VAL}	σ_{VAL}
Perceptron	1	7	60.07%	20.56	-	-	50.46%	19.20
MLP/BP	5	41	77.95%	8.82	74.91%	11.03	64.92%	11.26
MLP/ELM	20	161	82.48 %	3.67	-	-	65.17 %	4.80

The linear classifier Perceptron hits on average about 60% of the training dataset. It means the Perceptron has no capability of mapping the dataset. It also presents poor generalization capability. Looking to MLP results with BP and ELM it is realizable that the classification to this problem is a hard task even using non-linear classifiers: on average about 65% of the new data are correctly classified both to MLP/BP as to MLP/ELM. The number of weights in MLP/ELM is almost 4 times greater than the MLP/BP, which is an important feature when one thinks in to embed the ANN in micro-processors with limited memory capacity.

Table 2. Confusion Matrices Average.

ANN	CM_{TR} (%)			CM_{TS} (%)			CM_{VAL} (%)		
		Healthy	Faulty		Healthy	Faulty		Healthy	Faulty
MLP/BP	Healthy	80.21	19.79	Healthy	77.45	22.55	Healthy	69.33	30.67
	Faulty	24.31	75.69	Faulty	27.64	72.36	Faulty	35.17	64.83
		Healthy	Faulty					Healthy	Faulty
MLP/ELM	Healthy	87.84	12.16	-			Healthy	75.60	24.40
	Faulty	23.00	77.00				Faulty	35.01	64.99

The MLP/BP present more equally classification, it is possible to see in CM_{TR} that about 75% of faulty data were correctly classified on average, and about 80% of the healthy were correctly classified. That is a discrepancy of about 5% whereas in the MLP/ELM this discrepancy is about 10%.

Table 3. Percentages of correctly classification to each level of short-circuit.

ANN - Dataset	HI1	HI2	HI3	LI1	LI2	LI3
$MLP/BP - TR$	67%	72%	74%	73%	84%	91%
$MLP/ELM - TR$	70%	75%	77%	75%	80%	92%
$MLP/BP - VAL$	59%	60%	66%	64%	70%	70%
$MLP/ELM - VAL$	60%	61%	66%	64%	67%	72%

The faulty dataset is composed by 6 sub-divisions of faults, as early mentioned, named here as HI1, HI2, HI3, LI1, LI2, LI3. Table 3 shows the average of the correctly classification to each level of fault to training and test dataset. It is possible to see that the more coils are short-circuited, the faults are more correctly classified.

Average results are used to evaluate the general behavior of the designed neural networks, but in the real implementation one ANN must be chosen. Then one ANN trained by back-propagation and other trained by ELM were chosen to be pruning and then used to show the final results. The method used is called CAPE and can be found in [26].

The choice of the specifics ANNs take in consideration the classification rate of validation and training datasets, but it was observed mainly the *CR* to each class. The priority was to choose an ANN that correctly classified all the healthy conditions in the validation datasets. That choice aims to avoid false positives in an online constant monitoring.

Table 4. Results to specifics ANNs.

MLP	N_W	CR_{TR}	CR_{TS}	CR_{VAL}	CM_{TR} (%)	CM_{TS} (%)	CM_{VAL} (%)
BP	41	89.7	81.8	68.5	H: 94.9 5.1 F: 15.3 84.7	H: 81.8 18.2 F: 18.2 81.8	H: 100.0 0.0 F: 32.1 67.9
BP/ CAPE	34	87.1	81.8	70.2	H: 94.9 5.1 F: 20.5 79.5	H: 81.8 18.2 F: 18.2 81.8	H: 100.0 0.0 F: 39.7 69.3
ELM	161	84.1	-	63.8	H: 90.2 9.8 F: 22.0 78.0	-	H: 100.0 0.0 F: 36.1 63.9
ELM/ CAPE	-	-	-	-	-	-	-

In Table 4 it is shown that the specific MLP /BP reaches better classification both to training dataset as to validation dataset. Moreover, after pruning the MLP/BP was able to improve its generalization capacity. The MLP/ELM was not able to be appropriated pruned by the CAPE method.

In [1] the same dataset used in this work was used in a Self-Organized Map ANN. There are differences in attributes used, and in datasets normalization. The final result presented by [1] gives a global *CR* of 87.5%. However, the *CR* to healthy dataset is 52%, whereas the *CR* to faulty dataset is 94.5%. That result means the most data are classified as faulty which is larger than the healthy dataset. In an online constant monitoring it also probably means that there is a high probability of false positives occurrence.

6. Conclusions

The problem of early fault detection in fed-converter induction motor is a subject that is far to be completely solved. The investigation with real datasets reveals difficulties in separating the faulty datasets from the healthy ones, which reinforces the importance of constant on-line monitoring. The use of converter adds the possibility of to embed the system directly in the equipment, which means more protection to the converter besides all the discussed benefits of early fault detection to the

machine and to industries. The great advantages of this non-invasive detection system make its improvement a task with great potential; one possibility involves the choice of relevant spectrums as parameters to training the classificator which is not an ended issue and is directly related with the classifier accuracy.

The two algorithms used to training the classifier showed similar results, but the ELM computes much faster and is much easier designed, although the MLP/ELM needed four times more neurons in the hidden- layer to do so. The pruning method was capable to improve the generalization in MLP/BP through the removing of connections, but the MLP/ELM was not able to be pruning by the method used.

6.1 Acknowledgment

The authors thank the FUNCAP support to the creation of the test bench and also the CAPES for the master's scholarship.

References

[1] Coelho, D.; Medeiros, C., "Short Circuit Incipient Fault Detection and Supervision in a Three-Phase Induction Motor with a SOM-Based Algorithm." *Book of Advances in Self-Organizing Maps*. vol. 198, p. 315-323. Jan. 2013. ISBN 978-3-642-35229-4.

[2] Thomson, W.T.; Fenger, M.; , "Current signature analysis to detect induction motor faults," *Industry Applications Magazine, IEEE* , vol.7, no.4, pp.26-34, Jul/Aug 2001.

[3] Nandi, S.; Toliyat, H.A.; Xiaodong Li; , "Condition monitoring and fault diagnosis of electrical motors-a review," Energy Conversion, IEEE Transactions on , vol.20, no.4, pp. 719- 729, Dec. 2005.

[4] Nandi, S.; Toliyat, H.A.; "Condition monitoring and fault diagnosis of electrical machines-a review," *Industry Applications Conference, 1999. Thirty-Fourth IAS Annual Meeting. Conference Record of the 1999 IEEE* , vol.1, no., pp.197-204 vol.1, 1999.

[5] Baccarini, L. M. R. Detecção e diagnóstico de falhas em motores de indução. 2005. 179 f. Tesis (Electric Engineering PhD) – Universidade Federal de Minas Gerais. Belo Horizonte. 2005.

[6] Bezesky, D.M.; Kreitzer, S.; , "Selecting ASD systems," *Industry Applications Magazine, IEEE* , vol.9, no.4, pp. 39- 49, July-Aug. 2003

[7] Kowalski, C.T.; Wolkiewicz, M., "Stator faults diagnosis of the converter-fed induction motor using symmetrical components and neural networks," *Power Electronics and Applications, 2009. EPE '09. 13th European Conference on* , vol., no., pp.1,6, 8-10 Sept. 2009

[8] Kowalski, C.T.; Wolkiewicz, M., "Converter-fed induction motor diagnosis using instantaneous electromagnetic torque and power signals," *EUROCON 2009, EUROCON '09. IEEE* , vol., no., pp.811,816, 18-23 May 2009

[9] Kaufhold, M.; Schäfer, K.; Bauer, K.; Bethge, A. and Risse, J.: "Interface phenomena in stator winding insulation – Challenges in design, diagnosis, and service experience", IEEE Electrical Insulation Magazine, vol. 18, nº 2, pp. 27-36, March/April 2002.

[10] Natarajan, R., "Failure identification of induction motors by sensing unbalanced stator currents," Energy Conversion, IEEE Transactions on , vol.4, no.4, pp.585,590, Dec 1989

[11] Ballal, M.S.; Khan, Z.J.; Suryawanshi, H.M.; Mishra, M.K., "Detection of Inter-turn Short-circuit Fault in Induction Motor Using Theory of Instantaneous Symmetrical Components," *Industrial Technology, 2006. ICIT 2006. IEEE International Conference on* , vol., no., pp.460,464, 15-17 Dec. 2006

[12] Sayed-Ahmed, A.; Chia-Chou Yeh; Demerdash, N.A.O.; Mirafzal, B.; , "Analysis of Stator Winding Inter-Turn Short-Circuit Faults in Induction Machines for Identification of the Faulty

Phase," Industry Applications Conference, 2006. 41st IAS Annual Meeting. Conference Record of the 2006 IEEE , vol.3, no., pp.1519-1524, 8-12 Oct. 2006

[13] Joksimovic, G.M.; Penman, J., "The detection of inter-turn short circuits in the stator windings of operating motors," *Industrial Electronics, IEEE Transactions on* , vol.47, no.5, pp.1078,1084, Oct 2000

[14] Stavrou, A.; Sedding, H.; Penman, J., "Current monitoring for detecting inter-turn short circuits in induction motors," Electric Machines and Drives, 1999. International Conference IEMD '99 , vol., no., pp.345,347, May 1999

[15] Penman, J.; Sedding, H.G.; Lloyd, B.A.; Fink, W. T., "Detection and location of interturn short circuits in the stator windings of operating motors," Energy Conversion, IEEE Transactions on , vol.9, no.4, pp.652,658, Dec 1994.

[16] Das, S.; Purkait, P.; Chakravorti, S., "Separating induction Motor Current Signature for stator winding faults from that due to supply voltage unbalances," *Power and Energy in NERIST (ICPEN), 2012 1st International Conference on* , vol., no., pp.1,6, 28-29 Dec. 2012

[17] da S Gazzana, D.; Pereira, L.A.; Fernandes, D., "A system for incipient fault detection and fault diagnosis based on MCSA," Transmission and Distribution Conference and Exposition, 2010 IEEE PES , vol., no., pp.1,6, 19-22 April 2010

[18] Hyun Cheol Cho; Knowles, J.; Fadali, M.S.; Kwon-Soon Lee, "Fault Detection and Isolation of Induction Motors Using Recurrent Neural Networks and Dynamic Bayesian Modeling," *Control Systems Technology, IEEE Transactions on* , vol.18, no.2, pp.430,437, March 2010

[19] Bouzid, M.; Champenois, G.; Bellaaj, N.M.; Signac, L.; Jelassi, K.; , "An Effective Neural Approach for the Automatic Location of Stator Interturn Faults in Induction Motor," *Industrial Electronics, IEEE Transactions on* , vol.55, no.12, pp.4277-4289, Dec. 2008

[20] Das, S.; Koley, C.; Purkait, P.; Chakravorti, S., "Wavelet aided SVM classifier for stator inter-turn fault monitoring in induction motors," *Power and Energy Society General Meeting, 2010 IEEE* , vol., no., pp.1,6, 25-29 July 2010

[21] Thorsen, O.; Dalva, M.; , "Condition monitoring methods, failure identification and analysis for high voltage motors in petrochemical industry," *Electrical Machines and Drives, 1997 Eighth International Conference on (Conf. Publ. No. 444)* , vol., no., pp.109-113, 1-3 Sep 1997.

[22] Engelbrecht, A. P. (2007), “Introduction to Computational Intelligence, in Computational Intelligence: An Introduction”, Second Edition, John Wiley & Sons, Ltd, Chichester, UK.

[23] Hornik, K.; Stinchcombe, M.; White, H. “Multilayer feedforward networks are universal approximators”. Neural Networks, v. 2, p 359-366, 1989.

[24] Guang-Bin Huang; Qin-Yu Zhu; Chee-Kheong Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on* , vol.2, no., pp.985,990 vol.2, 25-29 July 2004.

[25] G.-B. Huang, L. Chen, and C.-K. Siew, “Universal approximation using incremental feedforward networks with arbitrary . input weights,” in Technical Report ICIS/46/2003, (School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore), Oct. 2003.

[26] Medeiros, Cláudio M.S.; Barreto, Guilherme. “A novel weight pruning method for MLP classifiers based on the MAXCORE principle”. *Journal of Neural Computing and Applications*, vol. 22, no 1. 2013. ISSN 0941-0643.