IMPLEMENTING AN ARTIFICIAL NEURAL NETWORK TO IDENTIFY CHARACTERISTIC LEAKAGES IN PIPELINES ANALYZING THE PRESSURE SIGNAL

Fernando Guimarães Aguiar, ferguimaraes@sc.usp.br
Paulo Seleghim Junior, seleghim@sc.usp.br
Univesidade de São Paulo, Escola de Engenharia de São Carlos, BR-13566590, nº 400, São Carlos, SP, Brasil

Simone Rodrigues Lima, simonelima@sc.usp.br
Univesidade de São Paulo, Escola de Engenharia de São Carlos, BR-13566590, nº 400, São Carlos, SP, Brasil

Abstract.
This paper presents the development of a leak detection system able to identify ruptures in pipelines, analysing pressure signals. The identification of the signal leakage is achieved by an Artificial Neural Network (ANN) feed-forward using multi-layer Perceptron (MLP) previously trained. To simulate leakages, fast opening solenoid valves calibrated and located in well-known positions are used. The experimental results were obtained in a 1500 meter-long and 51.2 millimeter-diameter pilot pipeline at the Center of Thermal Engineering and Fluids. The pressure sensors data were acquired from 9 sensors located along the pipeline at a sample rate of 100Hz and the leakage simulations were performed in 10 solenoid valves. Signals from simulated leakage (valves opening), valve closing and normal noise were acquired. The results with single-phase flow were good as the neural network trained was capable of classifying the 2 types of signals into 2 linearly separable regions.

Keywords: *leak detection, artificial neural networks, signal analysis, instrumentation.*

1. INTRODUCTION

The transport of oil and its derivatives in pipelines is a widespread practice in petrochemical industry . In this context, several techniques for both preventive and corrective leak detections have been studied to prevent and reduce environmental damage and economic loss. It must be emphasized that the risks associated with oil products transportation are very high. A leak or rupture along the pipe may cause irreparable damage to the environment, especially if the duct is built in flooded, submerged or environmental reserves.

According to Papadakis (1999), the main causes of accidents in oil and gas pipelines are connected to operational failures, corrosion, ground motion, natural phenomena and unintentional or unauthorized external actions (illegal connections, for example).

It is important to highlight a citation by Stouffs and Michel (1993): "No method of leak detection is universally applicable and operational requirements dictate which method is most effective for the process. In general, the ideal is to use more than one independent system for detecting leaks in very important pipelines". Such a citation was written 17 years ago, but remains valid and current.

Regarding physical phenomenon, a characteristic acoustic transient is propagated through the fluid flow along the pipeline. This disorder can be acquired using pressure sensors to respond rapidly (typically faster than 1kHz) placed along the pipeline. However, as the signal of leakage is attenuated and distorted by the flow, it is a challenge to develop algorithms to detect leaks. This work aims to develop an algorithm to detect abrupt leaks, using an algorithm based on artificial neural network Perceptron feed-forward. To simulate leak, fast opening solenoid valves are used, calibrated and positioned in well-known locations.

2. ALGORITHM FOR PIPELINE LEAK DETECTION

The systems currently deployed in pipelines exhibit poor performance due to the large number of false leak alarms. In this context several techniques for leak detection in pipelines have been developed, and according to Khomairi (2008), these techniques can be classified into three categories: observational methods - inspection of the pipe in search of abnormal operating conditions (smells, sounds, patches of fluid transported), which are not very effective in long pipelines – hardware-based methods – method that use devices, such as electronic sensors (pressure sensors, infra-red gas detection) to detect the leak - and software-based methods – methods based on modules of softwares that monitor and process control variables such as mass balance contained in the duct and mathematical modeling.

Specifically, the acoustic leak detection methods are based on the phenomenon of an acoustic transient is generated by a leak, which is propagated through structures and fluids along the pipeline. This signal can be acquired by acoustic or pressures sensors (usually piezo-electric or piezo-resistive pressure sensors). However the transient noise is attenuated while propagated along the pipeline, and to achieve a good sensitivity in the leak detection the acoustic sensors must be strategically positioned. To extract information features from signal leakage, signal processing techniques are required to classify the signal as signal leakage or normal operation of the pipeline. If it is possible to obtain the exact time of the signal detection in different acoustic sensors along the pipe, this method provides the location of the leak.

2.1. Artificial Neural Network

In a mathematical analysis, artificial neural networks (ANN) can be defined as nonlinear mappings of an input array in an output array, performed through layers of activation functions (or neurons) to which the input coordinates are added. According to neuron weights and bias a single output, activated or not, depending on the activation function is generated.

Guyon (1991) states that the artificial neuron is a logical-mathematical structure that seeks to simulate the shape, behavior and functions of a biological neuron. Thus, the dendrites are replaced by inputs whose connections to the cell body are carried through artificial elements called weight (simulating the neuronal synapses). The stimuli are received and processed by the sum function and bias. The threshold for triggering the biological neuron has as an analogy with the transfer function in artificial neurons (CHUA L. O., YANG L., 1988). Figure 1 shows a representation of the artificial neuron.

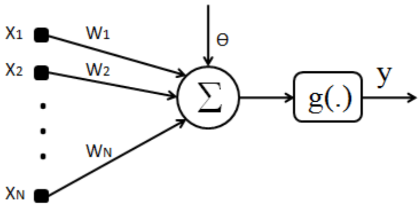


Figure 1. Artificial neuron model.

The research on neural networks dates back to 1943 (PICCININI G., 2004), when, in a pioneering work, McCulloch (physiologist) and Pitts (mathematician) studied the behavior of biological neurons towards creating the corresponding mathematical model for the interpretation of the neuron operation as a binary circuit (MCCULLOCH W. S.; PITTS W. H., 1943).

It must be emphasized that the 80's was the culmination of the research into neural networks with the development of the backpropagation training algorithm, formalized by Rumelhart, Hinton and Williams (RUMELHART D. E., *et al.*, 1986). The authors provided training for multilayer Perceptron networks, resulting in a network with high generalization power, allowing the implementation of various applications.

In the backpropagation training method a training set is presented to the neural network with pairs of known input and output. The input array is propagated layer by layer until the output layer. The array of the network output is then compared to the desired output array, and the difference between the two outputs (calculated and desired) is the output error of the network. This error is then propagated backwards to the network in order to adjust the synaptic weights, so that in the next iteration the output error is reduced. This process is repeated for all pairs of the network training set until the output error is acceptable, thereby reaching the stop condition of the training method.

Multilayer Perceptron neural networks (MLP) are characterized by the presence of multiple layers which can contain several basic processing units (neurons). These networks have supervised training by the backpropagation algorithm, which adjusts the weights and bias of neurons in the network, having a training set of a known set of inputs and a known set of desired outputs. Figure 2 shows the topology of RNA PMC implemented.

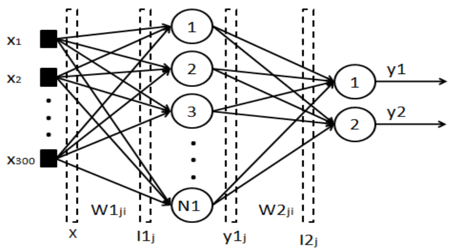


Figure 2. Topology of MLP ANN implemented.

Cross-validation is a methodology commonly used to test the generalization power of an artificial neural network. In this context, the data set is partitioned into two subsets, of which the first about 70 to 90% of the data used for training, while the other comprises the test set to verify if the neural network is generalizing satisfactorily.

2.2. Experimental Data Acquisition

All experimental tests were performed in a 1500 meter-long and 51.2 millimeter-diameter pilot pipeline at the Center of Thermal Engineering and Fluids. A tank system installed in the downstream sections of the test is responsible for the primary air-liquid separation and, subsequently, the liquid-liquid separation.

Centrifugal pump was controlled by a frequency inverter of 15kW (Yaskawa CIMR-P5U2011), which recirculated the liquid phases. This architecture was adopted to allow larger-scale variations in flow control strategies. In relation to the pipeline instrumentation, electromagnetic flow meter transmitters were installed both upstream and downstream, (EMERSON ROSEMOUNT 570TM) with full scale of 23.76 m³ / h (396 l / min). The instrumentation also included pressure transducers (WIKA A10) to measure the head loss and acquisition of acoustic signal, with full scale of 16 bar.

The set of signals to be acquired by the data acquisition system is composed of 10 pressure transducers, 2 flow meter transducers and 22 solenoid valves distributed along the pipe to simulate leakage in certain positions. Figure 3 shows a pipeline scheme of the general instrumentation in pilot pipeline.

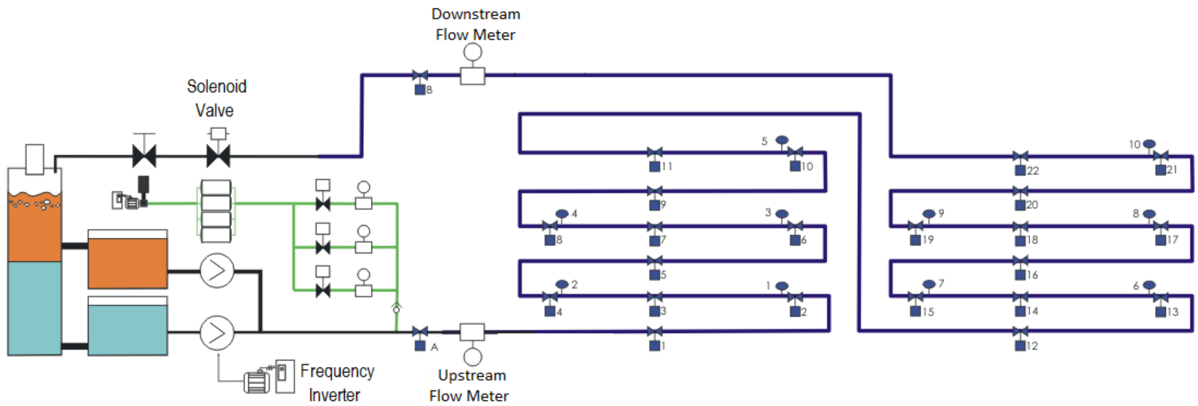


Figure 3. Pipeline scheme of the general instrumentation.

The system control and data acquisition are managed by a Programmable Automation Controller (PAC), which is an electronic hardware from National Instruments® (NI), named CompactRIO. It is a low-cost, high-performance reliable system. CompactRIO applications use three separate processors: a PC with Windows operating system, a controller with real-time operating system (RTOS) and an array of field programmable gate array (FPGA).

For the acquisition and storage of experimental data a software user interface, where the parameters are configured for the tests, was developed. The software on the real time controller filters the data using a Butterworth low-pass filter with cutoff frequency of 55Hz and stores the acquired data in high-capacity storage devices. Finally, the FPGA performs the acquisition and control of data at 100 Hz rate.

For the experimental tests a single-phase flow of water was used, and the software automatically performed cyclical tests according to the following operations:

i. Set the frequency of the water pump frequency inverter;
ii. Define the solenoid valve to simulate the leakage;
iii. Wait a certain time to stabilize the system (10 seconds);
iv. Start storing acquired data in high-capacity sotorage devices (ASCII files);
v. Wait a certain time to open solenoid valve (60 seconds);
vi. Open pre-defined solenoid valve;
vii. Wait a certain time to close the solenoid valve (60 seconds);
viii. Close the solenoid valve;
ix. Wait a certain time to stabilize the system (60 seconds);
x. Stop acquiring data.

Four frequencies (30, 40, 50 and 60 Hz) were arbitrated to control the pump. Due to maintenance in the pipeline pilot, only 10 of the 22 valves for leak simulation were working (V2, V4, V6, V8, V10, V13, V15, V17, V19 and V21), and the pressure sensor P1 was also inoperating. Combining the four frequencies and the 10 valves for leak simulation, there were a total of 40 cycles of experimental tests, resulting in 360 experimental signals from the 9 pressure sensors. The average flows of each experiment for the pump frequencies of 30, 40, 50 and 60 Hz were 73.7, 100.4, 125.4 and 146.6 l / min, respectively. The openings of the valves used to simulate leakage were calibrated to provide a flow rate of approximately 7 l / min with pump frequency inverter drive to 60Hz.

2.3. Using RNA for Leak Detection in Pipeline

2.3.1. Training ANN with Experimental Leak Signals

The artificial neural networks implemented were trained to classify leak and no leak signals (valve closed and normal operation of the pipeline) into separable areas and generalize the leak signal that is softened and deformed along the pipeline. To achieve this goal the definition of the number of the ANN layers was guided by Cybenko's work, which in 1989 proved that arbitrary decisions can be well approximated by a feed-forward ANN with a single hidden neural layer with sigmoid activation function, and a neural output layer with linear activation function. The input data are normalized on the network ($[-1; 1] \in \mathbf{R}$). Hyperbolic tangent sigmoid function was defined for the activation function of the hidden neural layer.

Experimental tests simulating leak were performed in 10 different valves and in four different rotations of the pump, and were recorded in nine pressure sensors. Each experimental test consisted initially of an instant of pipeline background noise, pursued by opening a solenoid valve, waiting the system to restabilize and then closing the solenoid valve.

The experimental signals were classified into three classes:

i. Leak (solenoid valve opening);
ii. Closing Leak (solenoid valve closing);
iii. Pipeline Noise (without triggering solenoid valve).

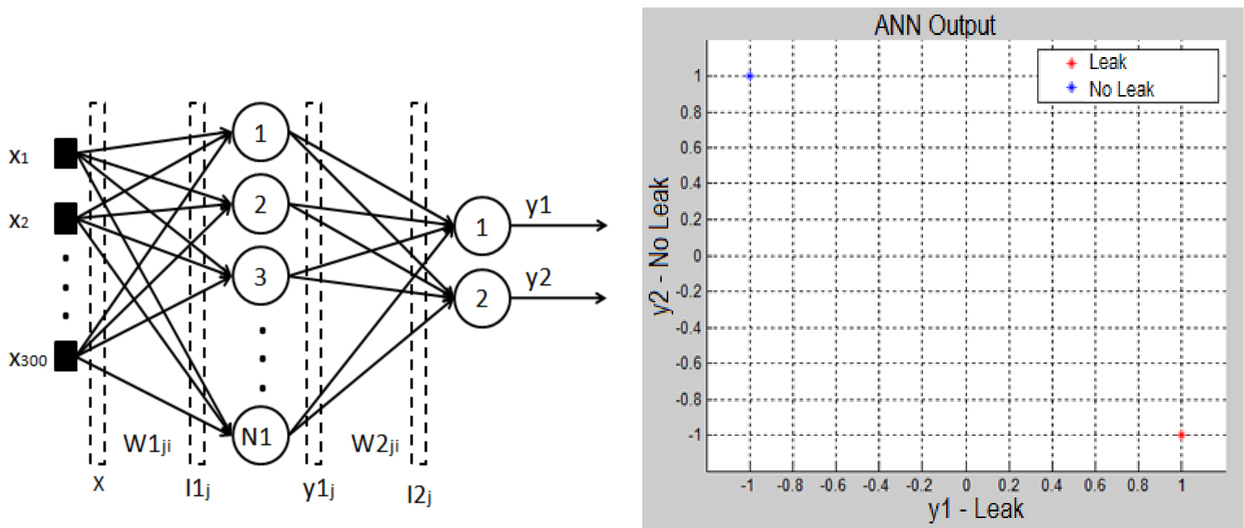
The signals of closing the leak should be classified as a pipeline operation. To simplify the ANN training the classification was standardized into only two classes:

i. Leak;
ii. No Leak (closing Leak and pipeline noise).

Thus, the experimental tests provided a total of 324 leak signals and 648 no leak signals, and 75% of the signals were used to train the ANN. The other signals were used to perform ANN cross-validation. The output of ANN has two neurons, which produce an output type (y1, y2). The ideal outputs for each class (leak and no leak) are:

- (1,-1) – Leak;
- (-1,1) – No Leak.

For a graphic display one has two cartesian axis (y1 and y2), where "1" indicates leak in y1, and "-1" in y1 indicates no leak, therefore, "1" in y2 indicates no leak, while "-1" in y2 indicates leak, as exemplified in Figure 4b. Figure 4a shows the ANN architecture used in this work.



(a) (b)

Figure 4. (a) ANN architeture used in this work. (b) Graphic visualization of the ANN outputs.

2.3.2. ANN Topology

As the input data in the network were normalized ([-1; 1] ϵ **R)**. The activation function of the hidden neural layer was defined as hyperbolic tangent sigmoid function. Equation (1) expresses mathematically the hyperbolic tangent sigmoid activation function.

$$\mathbf{y(x)} = \left(\frac{2}{1+e^{-2x}}\right) - \mathbf{1} \tag{1}$$

In this work the ANNs were defined as feedforward Perceptron with one hidden layer with hyperbolic tangent activation function and identity linear activation function for the output layer, which has 2 neurons.

2.3.3. ANN Cross-validation

Cross-validation is a methodology commonly used to test the generalization power of an artificial neural network. In this context, the data set should be partitioned into two subsets, of which the first contains about 70 to 90% of the data used only for training, while the other comprises the test set to verify if the neural network is generalizing satisfactorily.

In this work, the data obtained with the pump frequency inverter at frequencies of 30, 50 and 60 Hertz were used for training, while the data from 40Hz were used to validate the ANN. Thus, 75% of data were used for training and 25% for cross-validation.

3. RESULTS

In this study four different topologies of ANN were analyzed to verify the applicability of the trained artificial neural networks. All topologies are constituted by one hidden neural layer with N1neurons and an output neural layer with two neurons.

There is no formula to determine the optimal number of the amount of neurons in the hidden layers (N1). The values of N1 should be chosen so that there exist a good relationship between the computational cost and the minimization of the error, which is calculated by the difference between the ANN output and the desired output. After analyzing the results of a preliminary empirical study with different values of N1, the topologies for the values of N1 equal to 3, 5, 10 and 20 were studied detail.

Figure 5 shows the learning curves (mean square error versus training epochs) for different values of N1 analyzed. The criterion used for the training was a maximum of 10000 times or mean square error below 0.0005. The training was concluded when one criterion had been reached.

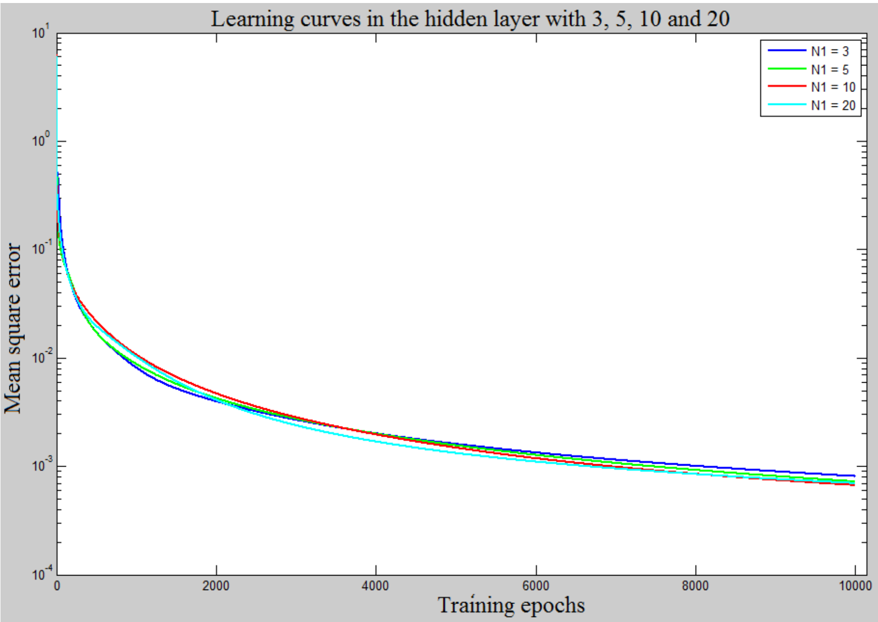


Figure 5. Learning curves in the hidden layer with 3, 5, 10 and 20.

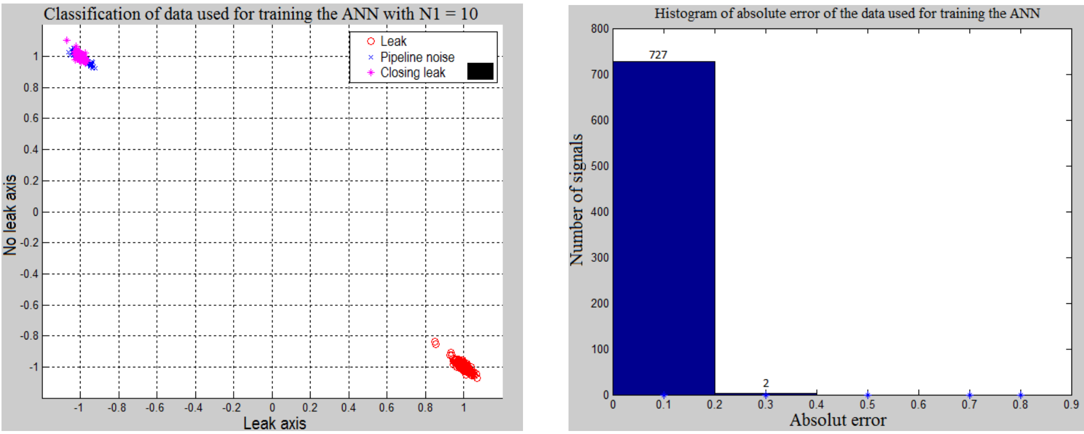
Analyzing Figure 5, all the learning curves show satisfactory results, with mean square error close to the desired error limit after 10000 training epochs. For a more detailed analysis of the results a neural network with 10 neurons in the hidden layer, which had the lowest mean square error in the training, was used. After training the network with N1 equal 10, the training data were classified.

To quantify the absolute error of each datum classified by the neural network, the geometric distance between the network output – (y1, y2) – and the ideal centroid leakage (y1d, y2d) was used. For a leak signal, the ideal centroid data is (1, -1), while for a no leak signal, the ideal centroid data is (-1, 1). Equation (2) expresses mathematically the absolute error.

$$\mathbf{erro} = \sqrt{(\mathbf{y_1} - \mathbf{y_{1d}})^2 + (\mathbf{y_2} - \mathbf{y_{2d}})^2} \tag{2}$$

Figure 6a shows the classification of the experimental data used for training the ANN. It is possible to observe that the leak and no leak data (noise and closing leak) are linearly separable, which allows the use of an identity function to separate the spaces.

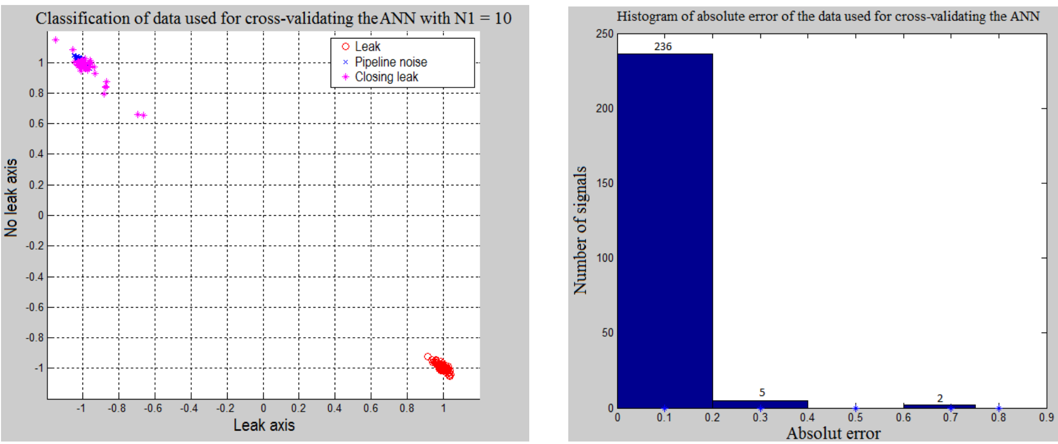
Figure 6b shows a histogram of the absolute mean square error of the output of the ANN, of which 99.7% of the data were classified as experimental absolute error between 0 and 0.2.



(a) (b)

Figure 6. (a) Classification of data used for training the ANN with N1 = 10. (b) Histogram of absolute error of the data used for training the ANN with N1 = 10.

The cross-validation method was applied to check the ANN generalization. In this method, the data which were not used for training the neural network are presented for classification. Figure 7a shows the classification of the test data used for cross validation. As the result of the data used for ANN training, we can observe that leak and no leak data (noise and closing leak) are linearly separable and may use an identity function to separate the spaces.

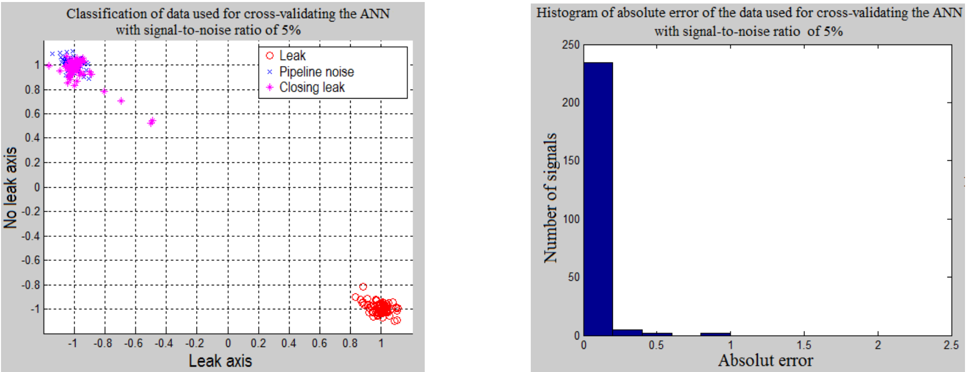


(a) (b)

Figure 7. (a) Classification of data used for cross-validating the ANN with N1 = 10. (b) Histogram of absolute error of the data used for cross-validating the ANN with N1 = 10.

3.1. Robustness Related to Noise in Input Signal

Since the ANN with 10 neurons in the hidden layer have outputs satisfactorily to the results obtained in cross-validation, were also analyzed the robustness to noise interference on the ANN input signal. To obtain this information, pseudo-random noise uniformly distributed was added to the signals used in the cross-validation, testing some different noise amplitude. For the analysis, signals were evaluated with a signal-to-noise ratio of 5, 20, 50 and 100%. The results are graphically showed in Figures 8a, 8b, 9, 10a, 10b, 11, 12a, 12b, 13, 14a, 14b and 15.



(a) (b)

Figure 8. (a) Classification of data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 5%. (b) Histogram of absolute error of the data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 5%.

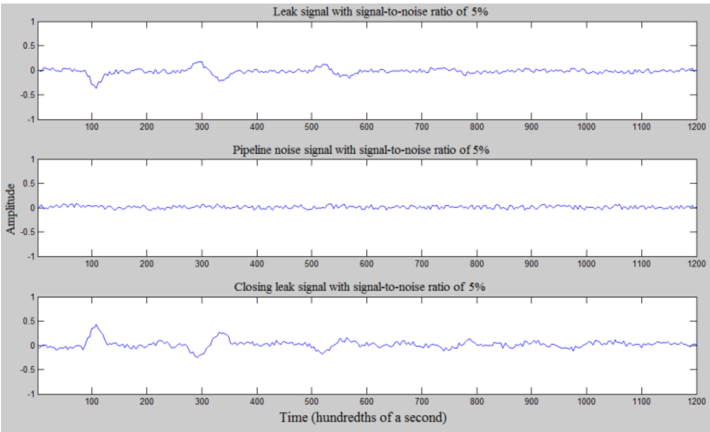
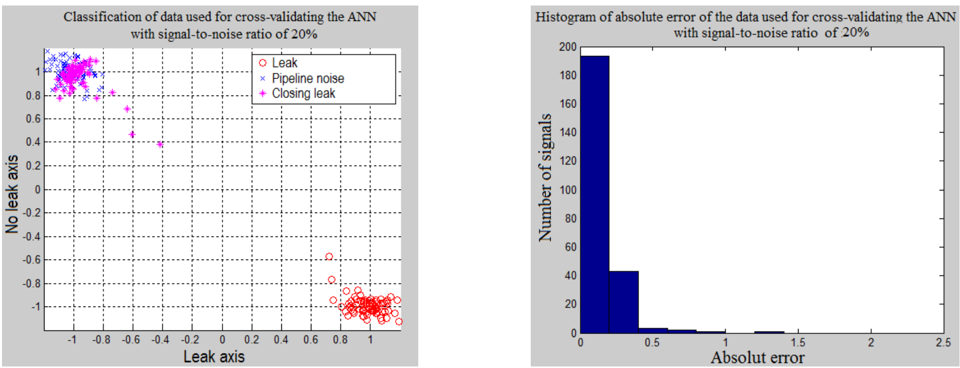


Figure 9. Example of leak, noise and closing valve signal with signal-to-noise ratio of 5% (Inverter frequency of 40Hz).



(a) (b)

Figure10. (a) Classification of data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 20%. (b) Histogram of absolute error of the data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 20%.

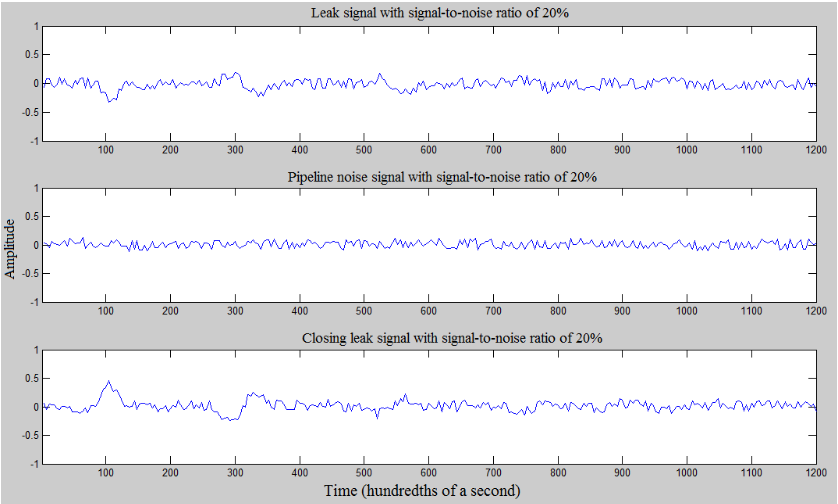
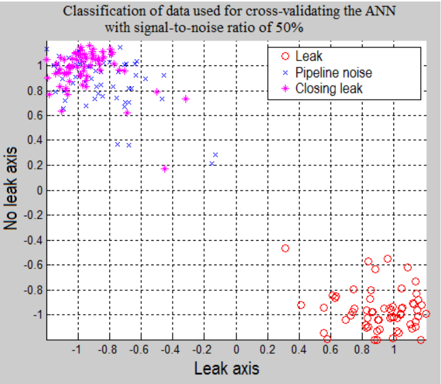
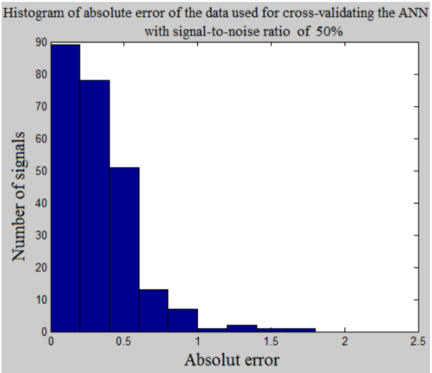


Figure 11. Example of leak, noise and closing valve signal with signal-to-noise ratio of 20% (Inverter frequency of 40Hz).



(a)



(b)

Figure 12. (a) Classification of data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 50%. (b) Histogram of absolute error of the data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 50%.

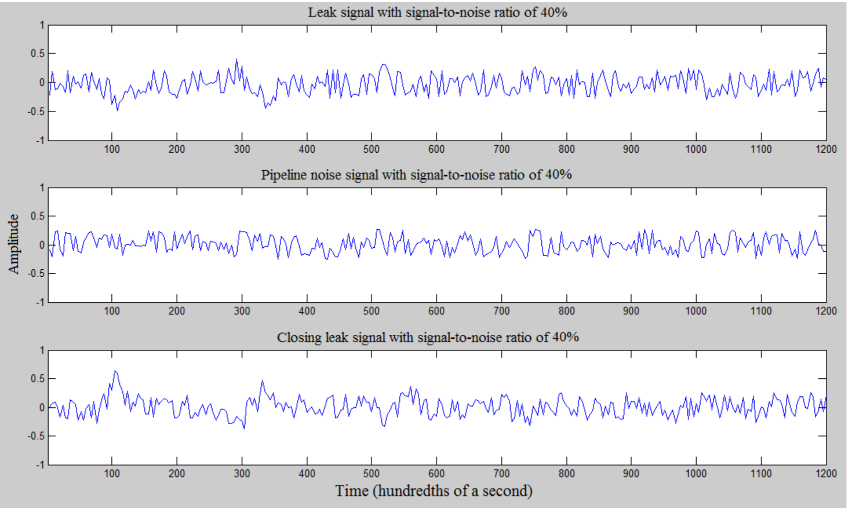
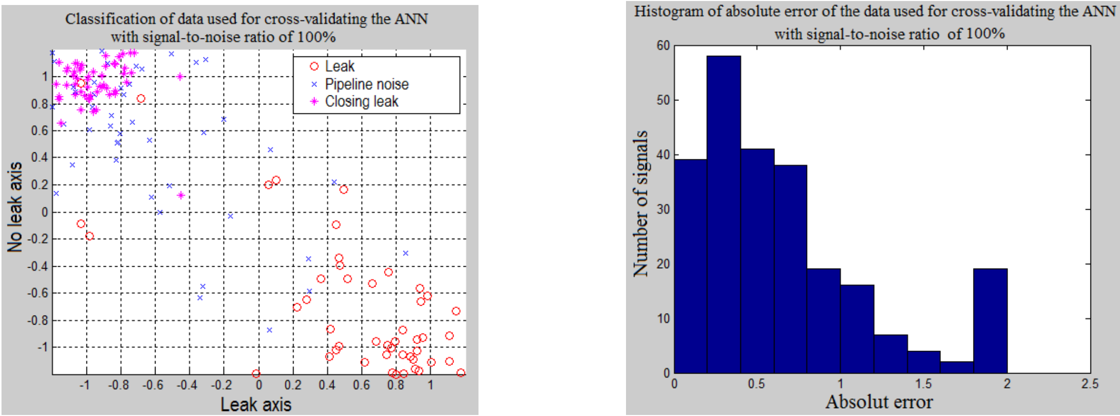


Figure 13. Example of leak, noise and closing valve signal with signal-to-noise ratio of 50% (Inverter frequency of 40Hz).



(a) (b)

Figure 14. (a) Classification of data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 100%. (b) Histogram of absolute error of the data used for cross-validating the ANN with N1 = 10 with signal-to-noise ratio of 100%.

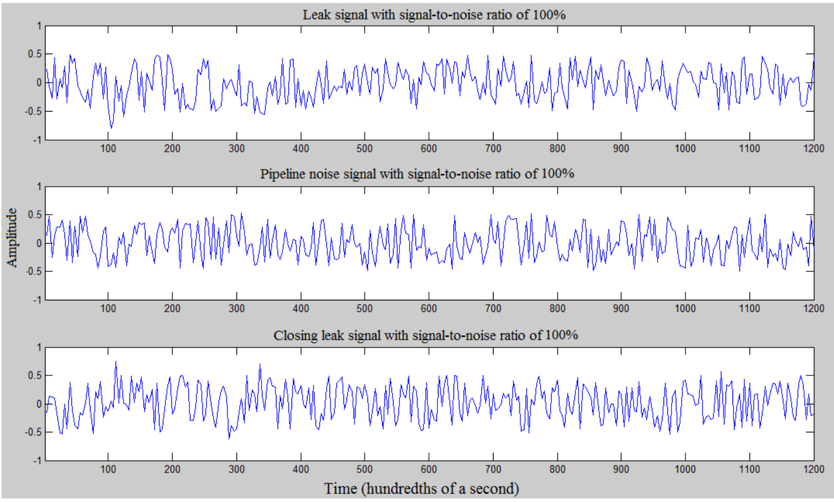


Figure 15. Example of leak, noise and closing valve signal with signal-to-noise ratio of 100% (Inverter frequency of 40Hz).

3.2. Analysis of results

Figures 8a, 8b, 10a, 10b, 12a, 12b, 14a and 14b demonstrate the robustness of the trained ANN, as the ANN produces a satisfactory output classifying leaks and no leaks into two distinct regions, which are linearly separable, for a signal-to-noise ratio of up to 50%.

For a signal-to-noise ratio equal to or greater than 50% the large number of leak data classified in the region of no leak and also the number of no leak data classified as leaks demonstrate that ANN did not produce a satisfactory response to signal-to-noise ratio over 50%.

Also, the results obtained from cross-validation demonstrate the power of ANN to generalize data which had never been applied as input, and classify them satisfactorily.

This work studied the operating fundamentals of artificial neural networks and presented a technique to detect a pattern of leak using a neural network feed-forward Perceptron with two neural layers trained as a classifier distinguishing between leak and no leak signals, using backpropagation training method.

Experiments were performed in the NETeF pipeline pilot to extract patterns of leak and no leak (background noises of pipes and signals of closing valve). The patterns of leak (rupture) were obtained by performing the activation (opening) of a solenoid valve quickly.

Due to the low computational cost the neural network under study can be easily implemented in embedded hardware with a floating point processor, as the Programmable Automation Controller (PAC) used to acquire experimental data.

4. ACKNOWLEDGEMENTS

The authors would like to acknowledge all the technicians, engineers and staff of the Núcleo de Engenharia Térmica e Fluidos from Universidade de São Paulo for supporting this research.

5. REFERENCES

CHUA L.O.; YANG L. (1988). *Cellular neural networks: theory*. IEEE Transactions on Circuits and Systems, v. 35, n°. 10, p. 1257-1272.

CYBENKO, G.V. (1989). *Approximation by Superpositions of Sigmoidal Function*. Mathematics of Control, Signals and Systems, v. 2, n°. 4, p. 303-314.

GUYON I. (1991). *Neural networks and applications tutorial*. Physics Reports, v. 17 n°. 3, p. 215-259.

KHOMAIRI, A. A. (2008). *Leak detection in long pipelines using the least square method*. Journal of Hydraulic Research, v. 46, n°. 3, p. 392-401.

MCCULLOCH W. S.; PITTS W. H. (1943). *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, v. 5, p. 115-133.

PAPADAKIS G. A.; PORTER S.; WETTING J. (1999). *EU initiative on the control of major accident hazards arising from pipelines*. Journal of Loss Prevention in The Process Industries, v. 12, n°. 1, p. 85-90.

PICCININI G. (2004). *The first computational theory of mind and brain: a close look at McCulloch and Pitts's "Logical calculus of ideas immanent in nervous activity"* Kluwer Academic Publisher, v. 141, p. 175-215.

RUMELHART D. E.; HINTON G. E.; WILLIANS R. J. (1986). *Learning representations by back propagation errors*. Nature, v. 323, n°. 9, p. 533-536.

STOUFFS J.; MICHEL G. (1993). *Pipeline leak detection based on mass balance: Importance of the packing term*. Journal of Loss Prevention in the Process Industries. v. 6, n°. 5, p. 307-312.

6. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.