

A Secure Method for Authenticity Verification of Handwritten Signatures through Digital Image Processing and Artificial Neural Networks

Deivison Pinheiro Franco¹, Felipe Dantas Barboza² and Nágila Magalhães Cardoso³

¹Master Program in Technological Innovation, University of Amazônia, Brazil

²Specialization Course in Forensic Sciences, University Center of State of Pará, Brazil

³Specialization Course in Computer Security, Higher Studies Institute of Amazônia, Brazil
deivison.pfranco@gmail.com, felipebarboza@yahoo.com, nagilamagalhaes@gmail.com

Abstract: This paper aims to propose a method to do authenticity verification of handwritten signatures based on the use of digital image processing and artificial neural networks techniques through the backpropagation learning algorithm with 500 and 901 approaches, in order to optimize this verification process and act as a decision support tool, in an automated way. The results showed an average percentage error of 20% in the first and of 5.83% in the second, while the performance of a trained professional for that has an average error of 6.67%. Thus, we could observe the efficiency of the proposed method, as well as the difference and evolution of approaches through the relevance of the results.

Keywords: Authenticity Verification, Handwritten Signatures, Graphoscopy, Digital Image Processing, Neural Networks, Back-propagation Algorithm.

1. Introduction

According to [1], manuscripts signatures still figures as one of the ways to validate documents authenticity due to its intense individual characteristic coupled with its low cost and practicality, despite of the emergence of various technologies related to this field like, for example, digital certificates and biometrics. Thereat, the fraud by signatures falsification is a crime much practiced in Brazil, reaching generate millionaires losses to people and institutions.

In 2009, KPMG Corporation conducted a research in order to investigate and evaluate the general scenario of organizational frauds in the country and showed that, at the time, 68% of companies interviewed suffered fraud. Of these organizations, 77% had losses of up to R\$ 1 million, and 5% of these losses exceeded 10 million. And the type of fraud with higher incidence (29%) was the checks and documents falsification, in which is present the signatures falsification [2]. However, the number of fraud organization cases in Brazil is much greater than the published, because the victims companies are afraid of negative public exposure, which would cause damage to its reputation and image, and even greater financial losses [3].

Graphoscopy is the discipline that certifies a professional to perform the verification of signatures authenticity through concepts and techniques that are the basis for safely conferences with effective results [4]. Thus, the performance of a graphoscopist covers the areas of criminal forensics and litigation, as well as banks, insurance companies, notaries and other financial institutions. And is unquestionable the

relevance of its work, once it is directly linked to the security of various institutions where it operates, as well as your users/clients, can perform as decisive evidence in solving crimes and misdemeanors. However, coupled with the intense workload, this professional is subject to many external factors in the exercise of its functions, such as fatigue, stress and personal problems, which could compromise its results.

The physical fatigue might result in misleading observations and mental fatigue favors forgetfulness, unnecessary repetition or omission of any exam. Such failures can bring losses and constraints for both the professional and the organization where he works, and for his customers, or acquit guilty or even incriminate innocent in court [4].

In order to automate the process of analyzing the authenticity of handwritten signatures and assist the professional in graphoscopy with an instrument to support decision making, this paper proposes the creation of a method that can do this verification automated with using techniques of digital image processing and artificial neural networks through backpropagation learning algorithm, which is capable of extracting "signature model" standards for comparison with one or more test signatures and definition of its degree of authenticity. Were also used graphoscopy concepts and studies proposed by [8], [9], [10], [11], [12] and [13], for signatures classification, analysis, interpretation and presentation of the results.

2. The Method

Artificial Neural Networks (ANNs) are a branch of Artificial Intelligence (AI) that aims at processing information in a similar way to the human brain [5]. Whereas backpropagation, according to [6], is a supervised algorithm by error correction for training multilayer artificial neural networks that minimize the error by running the decreasing gradient in the surface errors space weights, where the height for any point in the space corresponds to the measured weights of the error. Thus, the weights begin to be set in units of output, where the error measure is known and continues with the retro propagation of this error between the layers by adjusting the weights to reach the input layer units.

As in the output units the desired and obtained values are known, the adjustment of the synaptic weights is relatively

simple. However, for units of hidden layers, the process is not so simple. The weights for a particular neuron in the hidden units, should be adjusted proportionally to the processing unit error which it is connected. Thus, and in accordance with [6], two phases are distinguished in the learning process of the backpropagation: the propagation phase (forward), in which the entries are spread between the layers of the network (from input to output), and the retro propagation phase (backward), whose errors are propagated in the opposite direction to the input stream.

For that, were adopted two multilayer neural networks, progressive and connected. According to [5] the number of source nodes in the input layer of the network is determined by the dimensional observation space, which is responsible for generating input signals. The number of neurons in the output layer is determined by the required dimensionality of the desired response. The existence of layers is to allow the extraction of higher order statistics for some unknown underlying random process responsible for the "behavior" of the input data file on which the network is trying to acquire knowledge [5]. Also for [5], this is an arbitrary value and can vary depending on the analysis of the performance of the model. Another important point relates to the determination of the number of neurons in each hidden layer.

Given the above, and to meet the proposed use of the backpropagation algorithm, the method construction followed sequentially the following steps: Acquisition of Signatures, which were scanned on a common high quality scanner device, Preprocessing and Digital Processing of Images, whose results are the features extracted for analysis, Creation, Training and Testing of the Artificial Neural Networks, from where the results arose.

Due to the need to review the steps of digital image processing and architecture definition of the artificial neural networks, the study was divided into two approaches, which are described below, whereas initial approach did not meet the expectations of improvement in success rates.

Thus, the efficiency and the discrimination capacity of the proposed method are shown through a set of computational experiments with traditional problems related to forgery handwritten signatures by slavish imitation.

2.1 Bank of Images

Applying the concepts of digital image processing proposed by [7], [14] and [15] the process was developed from the signature collection of three distinct authors: Eric, Felipe and Rodrigo. Each author signed twenty times its own signature, falsified twenty times the signature of the second author, and also falsified twenty times the signature of the third author. This resulted in a total of sixty signatures each, and an overall total of one hundred and eighty samples as shown in Table 1. Therefore, was used a single bank of images, once its acquisition process was the same for both approaches.

The samples digitalization was performed from a common scanner device. Then, each image was resized generating images within the maximum range of 731 pixels wide by 180 high and a minimum of 650 pixels wide by 117 high, all in the ".png" format.

Table1. Signatures collected by author

Signatures	With the name of author 1 (Eric)	With the name of author 2 (Felipe)	With the name of author 3 (Rodrigo)
Written by author 1 (Eric)	20 (authentic)	20 (fake)	20 (fake)
Written by author 2 (Eric)	20 (fake)	20 (authentic)	20 (fake)
Written by author 3 (Eric)	20 (fake)	20 (fake)	20 (authentic)

2.2 Approaches

It was used two approaches: the 500 Approach and the 901 Approach - both with features of image processing and different network configurations, where the second came as an evolution of the first.

The two approaches used to prepare the method have scripts and functions to automate the process and, moreover, are based on the idea of using, from the pixel array, one-dimensional characteristics to the input layer in the recognition algorithms signatures through vertical projection (sum of pixels in each column) and of horizontal projection (sum of pixels in each row of the matrix).

In both approaches it was used the Matlab, version R2008a, to perform the procedures for preprocessing, segmentation and feature extraction of images, and to create, training and testing the ANNs, in order to permit an assessment of the characteristics described, comparing them with the activation function traditionally used, which through a graphical interface, allowed to obtain highly favorable results, presented later, showing the functions necessary for training and testing.

The difference between the approaches characteristics becomes more evident from this topic, which differs in the stages of preprocessing and feature extraction, network architectures, and examples of signatures submitted to ANNs in the same training set.

To permit evaluation of the function proposed is the method developed training in two separate networks, with the same basic parameters and using different activation functions. Considering that the architectures are very different and that each one defines different amounts of nodes in the ANNs sensory layers, the initial approach was called 500 Approach, by instituting five hundred entries and the subsequent was called 901 Approach by defining nine hundred and one entries. Such approaches are better described below with their respective processes involved to build the method.

2.2.1 500 Approach

- Images preprocessing routines:

1. Capture the recognized sample as three-dimensional matrix representing the RGB colors scale;
2. Transform in grayscale, which matrix format becomes two-dimensional, facilitating manipulation, because it involves fewer variables in the required calculations, both in processing and in training;
3. Contrast adjust, where the image pixels are highlighted and is highlighted the intensity difference between the darker and the lighter shades;

4. Histogram equalization adjust, producing an increase in brightness and also in contrast;
5. Resizing, which reduces the image to the default size of 400 pixels wide by 100 pixels high;
6. Binarization through logic operation, with the goal of making the background and lighter shades removal, ie, segmentation, such that the signature region becomes black and the remaining regions becomes white.

Given the above, Figure 1 below shows images of the same signature in relation to the histogram equalization.

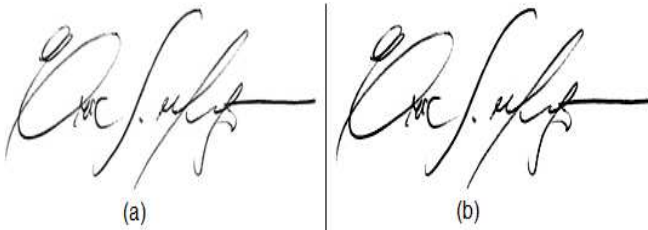


Figure 1. Histogram equalization: (a) Signature captured without equalization, (b) Signature captured with equalization

- Features extraction:

This step consists in the generation of a concatenated vector of 500 positions for each signature, where the 400 first are related to the sum of the columns (vertical projection) and the others 100 corresponds to the sum of the lines (horizontal projection) of the image, as shown in Figure 2. This vector represents the features extracted from the image of the signature that matches, and acts as a sample entry in the training set or test of the ANN in the 500 Approach.

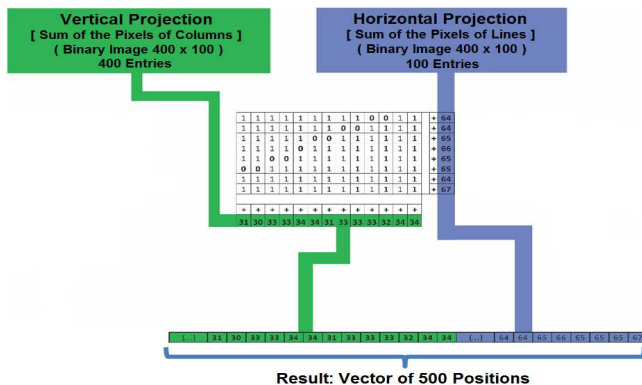


Figure 2. Features extraction of signatures in 500 Approach

- Artificial Neural Network:

Overall, there were several trainings, however, was highlighted in the 500 Approach only one ANN, called “Eric45”, because it was the network that showed better results in this approach. The name assigned to this network is composed by the first analyzed author's name followed by the number of examples presented to this ANN in its training set. Thus, the architecture of the “Eric45” network was:

1. 1 direct network, multilayer, fully connected;
2. 500 entries, corresponding to the vector of 500 positions;
3. 2 intermediate layers with 200 neurons in each;
4. 2 neurons in the output layer, where one is activated in case of authenticity, and the second in case of falsity.

A better view of the architecture of “Eric45” network in 500 Approach can be seen in the Figure 3 below.

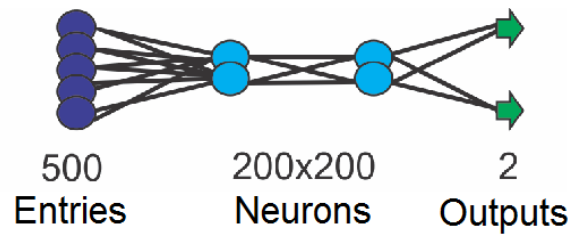


Figure 3. Architecture form of the “Eric45” network in 500 Approach

Therefore, is possible to observe and assign the following configurations and features of the “Eric45” network:

1. It has up to 1000 times to do the learning;
2. The mean square error to be achieved is 10^{-3} ;
3. Uses the backpropagation algorithm cause has great capacity for generalization and enable supervised learning;
4. The training is done with supervised learning, once the classes that the network must be distinguished are known;
5. The learning rate is adaptive;
6. Uses the momentum;
7. Uses the logistic activation function, where the response values are provided in the open interval between 0 and 1.

- Training Set:

To the 500 Approach, the training set was created with 45 signatures in order to recognize the author's signature Eric. Each of the three authors were used 15 signatures with the name Eric, being the first 15 authentic, activating the 1st neuron of the output layer, and the other 30 falsifications by slavish imitation (spelled by Felipe and Rodrigo), activating the 2nd neuron in the output layer.

2.2.2 901 Approach

The 901 Approach arose from the need to correct the errors of the 500 Approach, besides trying to improve performance. The propose presented in 901 Approach, while optimizing the efficiency and convergence rate of the backpropagation algorithm, provides the use of an activation function with features that allow greater flexibility in the representation of the phenomena modeled, in addition to using traditional parameters used for this purpose, which implies in a better solution for the saturation problem, and enables the treatment to avoid minima local. Another advantage, empirically observed, is to be computationally fast [6]. Also for [6], this further improves the performance of the backpropagation algorithm acting directly on the network topology. Many parameters were changed as observed in the description of the following steps.

- Images preprocessing routines:

1. Capture image in RGB format;
2. Transform in grayscale;
3. The contrast adjustment was modified to intensity adjustment. The function used was the same, but the

- intensity was changed manually, instead of automatically method of the previous approach (500);
- The adjust in the histogram equalization was not used, cause this practice emphasized the presence of noise in the image, which could interfere in the network learning;
 - Due to the detection of some noise points in the samples submitted to the processing algorithm, it was did a scan in order to turn in white all the pixels in the edges of the image;
 - After the manual removal of unwanted pixels, it has become possible to cut the sample by reducing the image area of the rectangle that delimits the exact size of the signature;
 - It was created a copy of the image, which has been reduced to a size of 40 pixels wide and 10 pixels high, and then binarized, so that each pixel could be used as input to the ANN;
 - With the original image, it was calculated the ratio between width and height (width/height), so that the result was used as an input of the network;
 - And finally, it was did the procedures of binarization and resizing of the sample to the size of 400 pixels wide by 100 pixels high, withdrawing the sum of each row and columns.

The Figure 4 below shows the main steps of the images preprocessing.



Figure 4. Images preprocessing procedures

• Features Extraction:

Unlike previous approach, the resulting vector in the 901 Approach, which is used in the subsequent training and testing process, consists of 901 positions, of which:

- The 400 first positions corresponding to each pixel of binary value (0 or 1) of the resized image in proportion 40x10, thus representing the positioning feature of the pixels in that image;
- The next 500 positions of the vector, ie, the 401th to 900th position, corresponding to the sums of rows and columns pixels of the resized image in proportion 400x100, equally to 500 Approach;
- And the last position corresponding to the result in pixels of the ratio between width and height of the image in 400x100 ratio, calculated in the preprocessing.

The Figure 5 below illustrates the features extraction of the signatures in the 901 Approach.

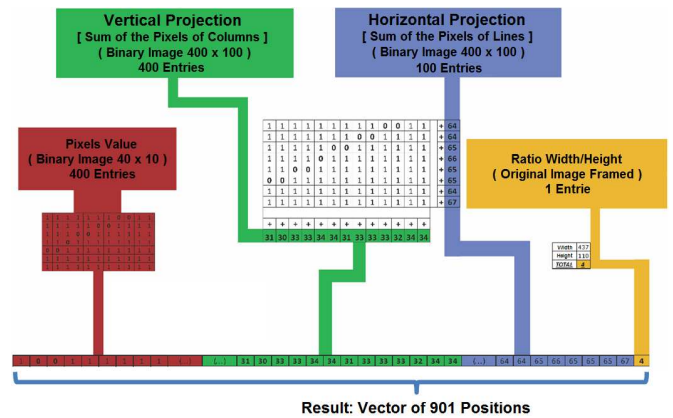


Figure 5. Features extraction of signatures in 901 Approach

• Artificial Neural Network:

In this approach were highlighted four networks, called "Eric35", "Felipe40", "Rodrigo40" and "Eric40", which have the same characteristics of architecture and configuration, but distinctions among themselves as to its training sets. Each name assigned to the network also consists of the first author's name followed by the number of analyzed samples presented to the ANN in its training set. Thus, the architecture of the "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks were composed of:

- 2 intermediate layers with 500 neurons each;
- 1 neuron in the output layer, which is activated in case of authenticity of the submitted sample;
- 901 entries, corresponding to the vector of 901 positions, of which: From 1 to 400, the minimum value is 0 and the maximum is 1, since each entry corresponds to a pixel matrix of binary image pixels 40x10; From 401 to 800, the minimum value is 0 and the maximum is 100, because it correspond to the vertical projection of the signature in 400x100 format; from 801 to 900, a minimum of 0 and a maximum of 400, considering the horizontal projection matrix 400x100; At position 901, the minimum value is 0 and the maximum is 10, according to the original size of the acquired images.

A better view of the architecture of "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks in 901 Approach can be seen in the Figure 6 below.

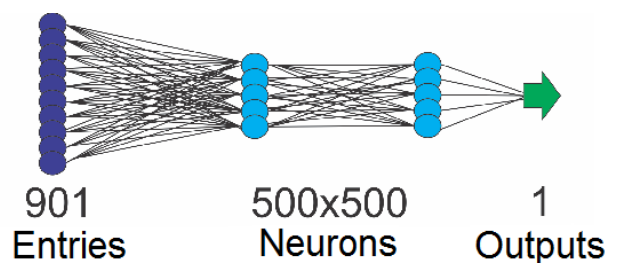


Figure 6. Architecture form of "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks in 901 Approach

Therefore, is possible to observe and assign the following configurations and features of "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks:

- Comprised 15.000 times for training;
- Used the backpropagation algorithm with adaptive learning rate and momentum;

3. The outputs in the range between 0 and 1, used logistic activation function;
4. Minimum squared error reached of 10^{-3} .

• Training Set:

To the 901 Approach, the training set used had distinctions as to the amount of samples used, as shown in Table 2 below:

Table2. Training sets in the 901 Approach

Network	Analyzed author	Originals samples	Fake samples	Samples with the names of other authors	Total of examples
Eric35	Eric	5	10	20	35
Felipe40	Felipe	10	10	20	40
Rodrigo40	Rodrigo	10	10	20	40
Eric40	Eric	10	10	20	40

Given the above, it was observed that in the 901 Approach the training set of each network is formed by the original samples plus the fakes and plus the names of the others authors. And the first was formed by ten original signatures of the author examined, except “Eric35” network - composed of five signatures.

While the false sample sets are composed of ten falsification by signature slavish imitation of the analyzed author, and being five from each one of the two other authors, the sample sets with the names of other authors consist of: Five authentic signatures of a second author; Five authentic signatures of a third author; Five falsification by slavish imitation of the signature of the second author; and Five falsifications, also by slavish imitation signature, of the third author. So, all behave like falsification without imitation of the analyzed author.

3. Experiments and Results

This section will present and discuss the experiments and results of the tests on the networks that make up the proposed method, displaying the error rates and the hit rates for each. Also described are the percentages of the two types of error found, such as: acceptance of false signatures (false positive) and rejection of original signatures (false negative).

It is noteworthy that external factors (such as color, material (pen or pencil) and its ink width, paper form and psychological changes of humor or disposition of the authors) were not considered in a first moment. These features will be addressed in future work for (and with) the evolution of the proposed method.

3.1 In the 500 Approach

The results of the “Eric45” network were acquired from the creation of a set test with 25 signatures (different from those used in the training set), defined as follows:

1. 5 original signatures of the author Eric;
2. 5 falsification by slavish imitation spelled by the author Felipe;
3. 5 falsification by slavish imitation produced by the author Rodrigo;

4. 5 original signatures of the author Felipe, functioning as falsifications with no imitation;
5. 5 original signatures of the author Rodrigo, also functioning as falsifications with no imitation.

The learning of the "Eric45" network resulted in 80% hit rate in the tests, i.e., 20 signatures of the sample space presented. The graphs in Figure 7, below, summarizes the information about the results obtained from this network. The first shows the hits and errors for the network in question while the second shows the separation of learning errors types observed.

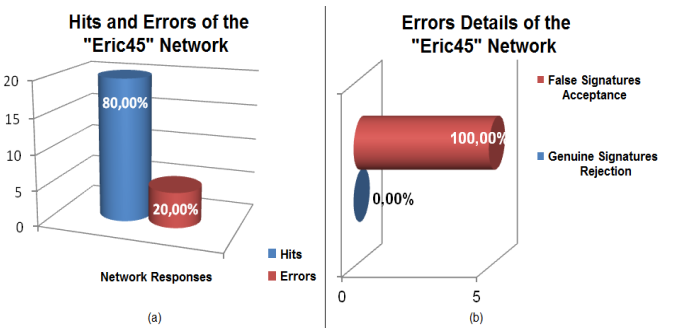


Figure 7. “Eric45” network results: (a) Hits and erros; (b) Erros types

3.2 In the 901 Approach

The Table 3, below, summarizes the results of the tests performed on all networks trained in this approach (“Eric45”, “Felipe40”, “Rodrigo40” and “Eric40”), with its successes and failures, as well as the values for the acceptance of inauthentic samples and rejection of original signatures.

Table3. Results of the tests applied to the 901 Approach networks

Samples in the test sets	Network	Hits	Hits %	Errors	Errors %	Original Signatures Rejection	False Signatures Acceptance
60	Eric35	55	91,67	5	8,33	5	0
60	Felipe40	52	86,67	8	13,33	2	6
60	Rodrigo40	59	98,33	1	1,67	1	0
60	Eric40	60	100	0	0	0	0

3.3 Tests With a Graphoscopist

Tests were conducted with a graphoscopist of the Bank of Brazil, with 15 years of experience as a lecturer of signatures, in order to compare his performance with the network "Eric35" because this is the only network implemented at the time of this professional availability. Therefore, in this procedure was used only samples of the training set (query patterns) and testing (questioned signatures) of this ANN. The results of the tests with the professional are detailed below in Table 4.

Table4. Results of the tests applied to the graphoscopist

Analyzed Samples	Hits	Hits %	Errors	Errors %	Original Signatures Rejection	False Signatures Acceptance
60	56	93,33	4	6,67	3	1

3.4 Results Comparison

Based on the tests results of all networks of the 500 and 901 Approaches, as well as the graphoscopist, it was traced the hit rates comparison, as shown in Figure 8 below.

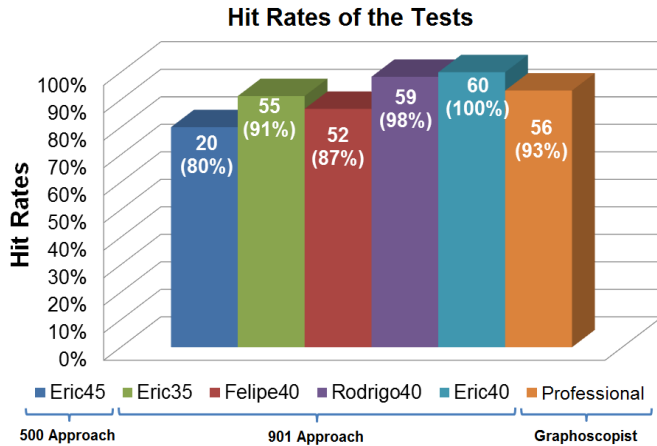


Figure 8. Hit rates graphic of the networks and of the graphoscopist

It was also observed the evolution of the errors types in the tests with the networks and with the graphoscopist, as shown in the graph of the Figure 9, which shows the percentage of false responses acceptance and rejection of true ones.

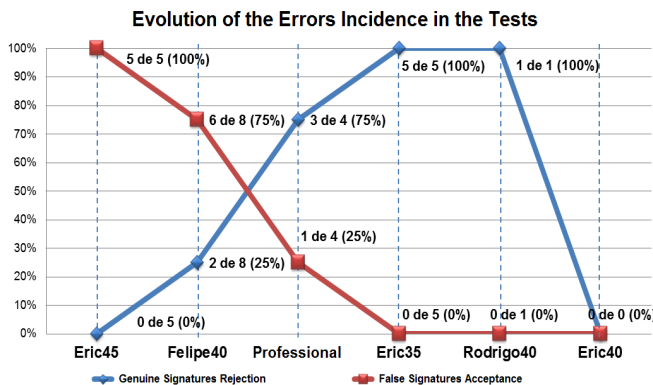


Figure 9. Evolution of the errors types percentage presented by the networks and by the graphoscopist

Given the results above, it was found that the 500 Approach, despite the considerable success rate, 100% of the network errors have been cases of falsifications acceptance, which consist in serious failure in the signatures authenticity verification, and it causes higher damages compared to the original signatures rejection. These factors lead to the conclusion that the techniques used at that time required modifications in relation to the size, quantity and variety of the examples, as well as changes in the quantities of neurons in the three network layers for better solutions. Thus, the efforts were directed to the reformulation of the resolution method, justifying the existence of the 901 Approach.

In the 901 Approach, the training times have grown exponentially, like the number of times required. The outputs became less diffuse, because the analysis was performed based only on the single neuron in the output layer.

When it draws a parallel between the “Eric45” and “Eric35” networks whose training was to recognize the Eric author’s signature, it was noticed considerable improvement, because

the initial RNA had an error rate of 20%, among which, 100% were characterized by the false signatures acceptance. And the “Eric35” network presented 8.67% error, which its set was composed only of original signatures rejection.

Comparing “Eric35” network with the graphoscopist, it is clear that the RNA results are under of the professional results relative to the quantity of hits. However, the implementation quality was better, in consequence of the only error committed by the professional included in the set acceptance of false signatures.

The “Felipe40” network, trained to the author’s signature Felipe, had the highest amount of errors in the 901 Approach, beyond accept fake samples.

The tests on the “Rodrigo40” network returned only 1,67% hits in the 901 Approach, with much original signatures rejections, allowing the conclusion that there is still difficulty in recognition of signatures standards, so that there are differences inside of the authentic sample set that can be clearly perceived. An example of this is the result of the graphoscopist analysis, which also missed credible examples.

The “Eric40” network, whose results were the best, had 100% hits on the test set. However, due to the find of distinctions between the authentic signatures, there is no guarantee that the network will behave the same way in the case of tests for other examples, even if it is in the standards adopted for the image bank formation.

The fact that the “Felipe40” network have lower hit rate between the three networks of the same training set architecture and configuration (“Felipe40”, “Eric40” and “Rodrigo40”), can be attributed to the medium graphical culture of Felipe signatures standards.

The “Rodrigo40” and “Eric40” networks, whose signature standards have high graphical culture, had higher hit rates. However, the “Eric40” network presented higher rate because Eric signatures standards have more facility areas that Rodrigo. Thus, it was assumed that the tests hit rate in the networks is directly proportional to the graphic culture level of the standards signatures by author analyzed.

4. Conclusions

In the method development there was little variety of different authors signatures standards, as well as considerable influence on the samples quality, once the device used to scan and the writing strategy of the signatures adapted the environment to facilitate collection and analysis.

According to what was explained previously, the method did not consider external factors such as color, material (pen or pencil) and its ink width, paper form and psychological changes of humor or disposition of the authors. Furthermore, the extracted features combination may still be lower than necessary to enable great generalization without losing the recognition reliability of each signature standard. So that the conclusions inferred on the results may not present the same behavior in all environments and existing signature standards, even if the provided responses are considered perfectly applicable.

Finally, the method may be considered perfectly valid and viable, once that achieved significant results in the signatures

authenticity verification. Fitting to point that there is no consolidated model, but a large increase in technical studies related to the subject. Therefore, this proposal aims to collaborate with the maturing of the theme that revolves around the use of Artificial Intelligence as decision support in situations that present a high degree of complexity, such as the case of signatures verification authenticity.

5. Future Work

For future work, it is proposed: collecting examples of other authors, for both the training set and the test set; add to the networks new extracted features from the images; cross-validation of the image bank samples to extract the best training sets and testing sets to optimize results; and consider external factors such as color, material (pen or pencil) and its ink width, paper form and psychological changes of humor or disposition of the authors.

References

- [1] Queiroz, F. e Sousa, A., "Exames Periciais a Documentos Manuscritos", Available from www.queirozportela.com/psicologiadacrita/pericias.pdf; [Visited November, 2012]; [In portuguese];
- [2] KPMG., "Relatório de Pesquisa Sobre Fraudes no Brasil", Available from www.kpmg.com.br/publicacoes/forensic/Fraudes_2009_port.pdf; [Visited May, 2013]; [In portuguese];
- [3] Theodoro, A., "Fraude Manuscrita: Um Risco Iminente", Available from www.peritocontador.com.br/artigos/Adriano_Theodoro/fraude.htm; [Visited May, 2013]; [In portuguese];
- [4] Gomide, F., "Manual de Grafoscopia", Livraria e Editora Universitária de Direito, 2ª edição 2008; [In portuguese];
- [5] Carvalho, A., "Redes Neurais Artificiais – Teoria e Aplicações", LTC, 2ª edição, 2007; [In portuguese];
- [6] Rumelhart, D., and Chauvin, Y., "Backpropagation: Theory, Architectures, and Applications", Wiley, 1st edition, 2000;
- [7] Gonzalez, R. e Woods, R., "Processamento Digital de Imagens", Pearson, 3ª edição, 2010; [In portuguese];
- [8] A. Leuski, "Evaluating document clustering for interactive information retrieval," in Tenth international conference on information and knowledge management. Atlanta, Georgia: ACM Press; 2001;
- [9] Zaroni Dias; Anderson Rocha; Siome Goldstein. "First Steps Toward Image Phylogeny" In: IEEE International Workshop on Information Forensics and Security (WIFS). Seattle, USA, 2010. p. 1-6;
- [10] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez Rodriguez, H. Fronthaler, K. Kollreider e J. Bigun, "A Comparative Study of Fingerprint Image-Quality Estimation Methods". IEEE Transactions on Information Forensics and Security, pp. 734-743, December 2007;
- [11] J. Lee, "Proposal for Efficient Searching and Presentation in Digital Forensics". In Third International Conference on Availability, Reliability and Security, IEEE Computer Society, pp. 1377-1381 (2008);
- [12] M. Schwartz, C. Hash, and L. M. Liebrock, "Term distribution visualizations with Focus+Context". In Proceedings of the 2009 ACM symposium on Applied Computing (SAC '09). ACM, New York, NY, USA, pp. 1792-1799. 2009;
- [13] N. Beebe and G. Dietrich, "A new process model for text string searching". In Research advances in digital forensics III, Sheno S, Craiger P, Eds.. Norwell: Springer; 2007. pp. 73-85;
- [14] João Papa; Anderson Rocha, "Image Categorization Through Optimum-Path Forest and Visual Words". In: IEEE Intl. Conference on Image Processing (ICIP), 2011, Brussels, Belgium;
- [15] Walter Scheirer; Anderson Rocha; Jonathan Parris; Terrance E. Bout, "Learning for Meta-Recognition". In: IEEE Transactions on Information Forensics and Security (T.IFS), v. 7, issue 4, p. 1214-1224, 2012.