

# Classification of Recurrence Plots of Voice Signals Using Convolutional Neural Networks

Luana Rodrigues Barros, Gabriel Gutierrez Pereira Soares, Suzete Élide Nóbrega Correia, Gabriel Daltro Duarte and Silvana Luciene do Nascimento Cunha Costa

**Abstract**—In the last decade, texture analysis has been widely applied to image classification. This method has great relevance on recognition of patterns in medical images surfaces. Alternatively, many studies have investigated the usage of Convolutional Neural Networks (CNNs) as a technique for classifying texture images. In this study, a low-complexity CNN was applied to recurrence plots of voice signals to distinguish the presence of laryngeal pathology. A data augmentation technique was employed to increase the number of samples. A study using the same dataset and another classification approach was considered for results comparison. The CNN proposed was proven to be more robust with a 12% increase on accuracy compared to the previous work.

**Keywords**—Convolutional neural networks, texture, recurrence plots, voice signals, laryngeal pathology.

## I. INTRODUCTION

Recurrence plots, developed by Eckmann et al. [1], are based on the idea of recurrence proposed by the French physicist and philosopher J. H. Poincaré, and constitute a method that allows visualizing the dynamics of recurring systems [2]. Such plots consist of two-dimensional matrices that correspond to digital images with two levels of gray (black and white), in which pixels can form typical structures such as isolated points, diagonal, vertical and horizontal lines. These structures characterize the systems and allow the recurrence plots to be treated from a small scale point of view as texture images [3].

In this context, texture analysis is an important tool for statistical and mathematical characterization of a surface appearance. In image processing, a texture represents a set of variations in pixels intensity repeating on a regular or random basis over an area, thus creating a pattern. Because of the effectiveness of this attribute in image features analysis, texture has been used in most studies focused on identifying the composition of an area. Research such as those by Scalco and Rizzo [4] and by Sghaier et al. [5] employ this type of technique.

Several methods have been proposed on the pursuit of an efficient texture description. Three commonly used approaches are statistical, structural and spectral. In the statistical approach, texture analysis is performed using statistical measures

based on the distribution of gray levels. The structural approach is based on the idea that textures are composed of primitive structures, i.e., basic geometric structures forming the texture. In this case, regularly spaced parallel lines are an example. The spectral approach, in turn, represents the image in a space of coordinates whose interpretation is related to the characteristics of the texture (frequency or size) obtained through transforms such as Fourier, Gabor or Wavelet [6, 7].

Another relevant aspect for texture analysis is the scale. According to Parker [8], texture is the property inherent of a region large enough to demonstrate its recurring nature. Therefore, if an image represents a smaller area than the equivalent of a texton (smallest segment of the image in which it is possible to perceive the orientations of the pattern), it can not represent a texture.

In the last decade, the usage of texture analysis gained notoriety in the health field, mainly focused on identifying the composition of surfaces in biomedical images. The research by Beckers et al. [9], for instance, proposes the analysis of texture in CT (computed tomography) of the entire liver to predict the development of colorectal liver metastases. Another application presented by Souza et al. [10] analyzes recurrence plots of voice signals as texture images in order to identify the presence of pathology in vocal folds. The feature extraction was based on the wavelet transform of the original images. Data obtained were later classified by a Multilayer Perceptron (MLP) neural network.

Despite the advantages promoted by texture analysis, according to Andrearczyk [11] classic methods of data processing do not generalize satisfactorily to textures that are complex, numerous or with high intra-class variation, as seen in several problems. In his study, the author also suggests that classical analysis approaches do not keep up with the variation of problems imposed by numerous applications. As a consequence, complex data classification using classical image processing algorithms becomes a costly activity, requiring manual work to implement a unique solution for each problem. The limitations imposed by the aforementioned approaches can be transcended through the usage of artificial intelligence algorithms such as Convolutional Neural Networks (CNNs), commonly applied to classify data in multidimensional matrices format.

CNNs can be highlighted among the algorithms based on Artificial Neural Networks (ANNs). These algorithms have a structure similar to biological systems and were developed for data applications such as images and videos. CNNs eliminate the need for the feature extraction process, hence images can

Luana Rodrigues Barros, Gabriel Gutierrez Pereira Soares, Suzete Élide Nóbrega Correia and Silvana Luciene do Nascimento Cunha Costa. Academic Unit 3, Federal Institute of Paraíba (IFPB), João Pessoa-PB, Brazil, E-mails: luana.barros@academico.ifpb.edu.br, gabriel.soares@ifpb.edu.br, suzete@ifpb.edu.br, silvana@ifpb.edu.br. Gabriel Daltro Duarte. Electrical Engineering Department, Federal University of Campina Grande (UFCG), Campina Grande-PB, Brazil, E-mail: gabriel.duarte@ee.ufcg.edu.br.

be used directly as inputs of the network.

In this scenario, this study proposes the use of a CNN to classify recurrence plots of healthy voices and voices affected by pathology. A CNN of low computational cost was developed to achieve this goal. A technique to artificially increase the number of samples in the database was also developed. The results were also compared with Souza et al. [10] which uses the same recurrence plots and a classification method based on feature extraction.

## II. RELATED WORKS

Andrearczyk [11] proposes the use of CNNs for dynamic texture (DT) classification. The use of CNNs, explored as trainable deep filters, are conducted as an alternative for classical filters (implemented manually for feature extraction). One of the experiments shows a low complexity CNN for classification of biomedical images of liver tissues. As the images present repeated patterns all over the surface, each image was split into sub-images and then used as independent inputs to feed the CNN. Splitting the images from the original database increased the number of samples available for training. The classification of a full-size image was given by the most assigned class among the sub-images. Results showed that this approach had a better accuracy than other studies using the same database.

Texture analysis has also aided in classifying data with different image structures. Recurrence plots can provide a way to visualize the periodicity of a trajectory through the phase of a time series, making possible to analyze the signal in a two-dimensional representation. Due to the periodic nature of recurrence plots, they can be analyzed as texture images. A study by Hatami et al. [12] investigates the performance of the recurrence plots applied to a CNN with two convolutional layers as a solution for time series classification. To validate this approach an experiment was conducted using the UCR time series classification file. According to the authors, other approaches have shown inferior results when compared to the use of the CNN in the classification of time series.

In regard of the benefit promoted by CNNs in texture classification, another research can be highlighted. Raw data from an accelerometer was modeled as recurrence plots by Garcia-Ceja et al. [13]. Then, a CNN with two convolutional layers was used to perform physical activity recognition. For comparison purposes, a Deep Belief Network (DBN) was also trained with the raw data from the three axis accelerometer and its magnitudes. An average accuracy of 83% was obtained in the classification performed with DBN, while the proposed approach using the CNN had 94.2% accuracy.

A texture analysis was performed by Souza et al. [10] on the wavelet transforms of recurrence plots of voice signals. Haralick descriptors were used for feature extraction. A particle swarm optimization (PSO) algorithm was applied to select the best subset of features. Then, a MLP was used to classify the data. This approach achieved an average accuracy of 86.32% in the distinction between healthy and pathological voices. The same database is used in the current study to validate the efficiency of CNNs in the detection of laryngeal pathology.

## III. DATABASE

The voice signals used in this research are part of the database recorded by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab [14]. The samples correspond to the sustained vowel /a/, recorded in a controlled environment, and sampled at the rate of 25,000 samples per second, with a resolution of 16 bits per sample. The database consists of 53 healthy voice files and 112 voice signals affected by pathology in the larynx, of which 51 have paralysis, 43 Reinke's edema and 18 nodules in the vocal folds. The recurrence plots used in this research were generated in the Visual Recurrence Analysis (VRA) [15] software, obtained from Souza et al. [10].

## IV. APPROACH

This paper aims to improve the accuracy achieved in earlier works in distinguish the presence of laryngeal pathology. The major innovation of this approach is the use of a Convolutional Neural Network for classifying recurrence plots of voice signals from the MEEI Voice and Speech Lab [14]. Another point of enhancement is a data augmentation technique developed specifically for recurrence plots, considering the nature of the image and the structures forming the textures. An overview of the steps adopted in this approach are shown in Figure 1.

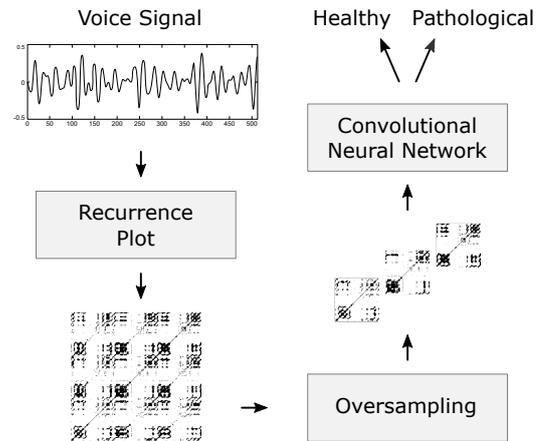


Fig. 1. Block diagram of signals classification.

First, recurrence plots were generated from the voice signals. Then, a data augmentation technique was applied to increase the number of samples. The sub-images were then used as independent inputs to the CNN. The network output were two classes: healthy and pathological. The classification of a full-size recurrence plot from the test data-set was given by the most assigned class among the sub-images.

### A. Recurrence Plots

Recurrence plots are two-dimensional matrices of order  $N$  that allow the visualization of the recurring behavior of dynamic systems, regardless of their dimensionality. In this matrices, black or white points are marked. A black point indicates the recurrence of a specific state at time  $i$  at another time  $j$  [1].

According to Marwan [3], a recurrence plot can be defined mathematically by Equation 1:

$$R_{i,j}^{m,\epsilon} = \theta(\epsilon - \|\vec{\xi}_i - \vec{\xi}_j\|), \quad \vec{x}_i \in \mathbb{R}_m, \quad i, j = 1 \dots N \quad (1)$$

In which  $N$  is the number of states  $\vec{\xi}_i$  formed by the system, obtained through the method of time delays [16]. This method immerses the time series in a  $m$ -dimensional space with time lag  $\tau$ . The neighborhood radius  $\epsilon$  is a fundamental parameter in the construction of the recurrence plot. When calculating the distance between the states of the system, the neighborhood radius defines which points are inside or outside that neighborhood (i.e. recurring or non-recurring points).

The step function  $\theta(x)$  is responsible for placing the recurring points within the recurrence plot. If the operation contained by the step function obtains a value less than or equal to 0,  $R_{i,j}^{m,\epsilon}$  assumes a value of 0 indicating that there is no recurrence between the states. If the operation results in a value greater than 0,  $R_{i,j}^{m,\epsilon}$  assumes a value of 1, indicating that there is a recurrence between states [17].

The recurrence plots used in this research were the same from Souza et al. [10] in order to guarantee that the same values of  $\epsilon$  and  $\tau$ , for comparison purposes. These values were obtained using the VRA software [15].

### B. Data augmentation

Biomedical signal databases generally have a limited amount of material available due to privacy policies to which the data is submitted. It is known that CNN algorithms require a large volume of images in order to reach a generalization that understands the possible variations in the dataset. In this context, the recurrence plots obtained from the voice signals [14] count a total of 165 samples. Despite the small number, due to the recurring nature of the graphs, they can be treated as texture images, allowing the number of samples to be increased by subdividing the original image.

In this study, a data augmentation technique was developed for recurrence plots based on splitting the images of the original dataset. The orientation of the pattern structures along the main diagonal and a minimum scale was considered to preserve the concept of *texton* [8]. Originally, recurrence plots have a size of  $384 \times 384$  pixels. A mask of size  $128 \times 128$  pixels was used to obtain the sub-images. Figure 2 illustrates how the original images were split. Three sub-images were generated without superposition and used as independent inputs to the CNN training.

One great advantage of the data augmentation technique developed in this work is the fact that the original features of the image are not modified. The proposed method maintains the characteristic of the recurrence plots to be symmetrical about the main diagonal. In this case, some types of operations could include new features in the database, which could be erroneously attributed to a non-existent anomaly.

### C. Convolutional Neural Network

1) *Architecture*: CNNs are a type of ANN specialized in applications involving data whose format is known as grid

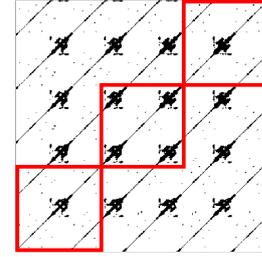


Fig. 2. Sub-images delimited by red squares.

topology, such as images and videos. In these networks, convolution can be highlighted as the most important operation. A CNN model was developed to classify the recurrence plots of voice signals into two classes: healthy and pathological. The CNN architecture was based on the ones used by Andrearczyk [11]. More specifically, the number of convolutional layers was based on the network that obtained the best performance in the classification of texture images from different databases, which had a total of 3 convolutional layers. According to Andrearczyk [11], the experiments conducted demonstrate that simple networks with reduced number of neurons and weights are able to obtain competitive results on texture recognition datasets.

In these circumstances, the architecture consisted of three convolutional layers followed by pooling operators, a flatten layer and two fully connected layers. The convolutional layers were composed of 32 filters of size  $3 \times 3$ . These filters convolve inputs of size  $128 \times 128$  pixels producing feature maps. The ReLU activation function was applied with the main purpose of avoiding a problem known as leak gradient, which can be circumvented due to the value assigned to the gradient being always constant. The pooling layers had the discretization process based on the max value of the receptive field. This operation is also known as max pooling, which was applied with a  $2 \times 2$  kernel and a stride of 2. This layer was used in order to reduce the input image, also reducing computational cost, memory usage and the number of parameters [18].

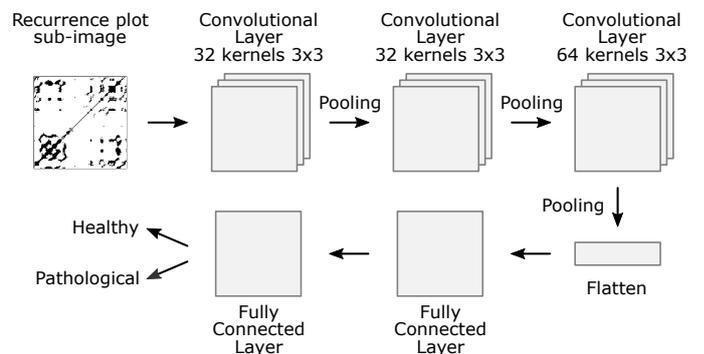


Fig. 3. CNN architecture.

2) *Training*: Because of the small number of healthy samples, the number of pathological ones was reduced. Thus, 30 samples of edema, 14 of nodule and 30 of paralysis were randomly selected, also reducing problems associated with the high degree of heterogeneity of the pathological class.

Consequently, the final set used for testing and training the CNN was made of 53 healthy and 74 pathological samples.

After composing the dataset, 80% of the samples were reserved for training and 20% for testing, considering each class proportion. The training group was split according to the technique of data augmentation exposed in this study. The images were then used as independent inputs to train the CNN.

A back-propagation algorithm was used to estimate the network parameters, which is a method based on the gradient descent optimization. More specifically, the Adam optimizer algorithm was used, with a learning rate  $\eta = 10^{-4}$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and  $\epsilon = 10^{-7}$ . During training, the stop conditions were if the number of epochs reached 100, or, to avoid overfitting, if the validation loss increased five consecutive times. In order to ensure a good generalization of the network regardless of the dataset division, a 5-fold cross-validation technique was applied, which also helps on avoiding network overfitting.

The CNN was performed with TensorFlow on a NVIDIA GeForce GTX 1050 GPU, with 768 CUDA cores and 3 GB of dedicated memory. Each training epoch took 0.26 seconds on average, while the prediction of each recurrence plot took 7.3 milliseconds.

3) *Classification*: After the training finished, the classification of the test dataset was performed. In this process, each image was sliced according to the split applied in training dataset. An example of the class assignment method is shown in Figure 4. According to that, the class of the original image is assigned by summing the classification of different parts of the image.

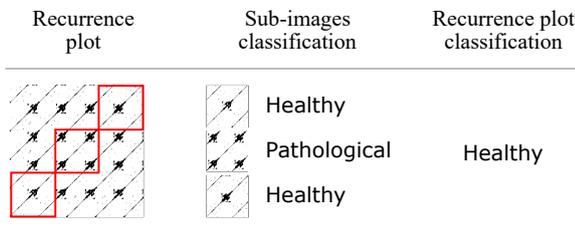


Fig. 4. Method for sample class assignment.

## V. RESULTS

The performance evaluation of a neural network consists on verifying if the class predicted corresponds, in fact, to the class of the input sample. It can be considered that a pathological sample corresponds to a positive class, while a healthy sample corresponds to a negative class. Given that, four outcomes can be considered: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Some measures, such as accuracy, sensitivity and specificity can be obtained from the parameters TP, FP, TN and FN, contributing to the evaluation of the network's performance.

- **Accuracy** – Indicates the network's ability to classify samples correctly. It is mathematically expressed by Equation 2.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

- **Sensitivity** – Measures the network's ability to correctly identify pathological samples. It is mathematically expressed by Equation 3.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

- **Specificity** – It measures the network's ability to correctly determine healthy samples. It is mathematically expressed by Equation 4.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

Results obtained from the CNN trained with an augmented database are given in Table I. The averages and the respective standard deviations for accuracy, sensitivity and specificity are shown for individual classification of sub-images and for the original recurrence plots.

TABLE I  
CLASSIFICATION RESULTS.

	Sub-Images	Recurrence Plots
Accuracy (%)	96.4 ± 2.3	98.3 ± 2.3
Sensitivity (%)	94.3 ± 3.6	97.1 ± 3.9
Specificity (%)	99.3 ± 1.5	100.0

The accuracy was satisfactory, with an average 98.3% and standard deviation of 2.3% for the original recurrence plots. In addition, the network showed better efficiency in classifying healthy images, as indicated by the specificity of 100%.

Table II gives a comparison between Souza et al. [10] and the approach proposed in this work. Our method obtained the most satisfactory results for all parameters with a smaller standard deviation. In particular, the network obtained an even greater gain in relation to the classification of pathological samples, as indicated by sensitivity. This proves that the approach proposed was able to perform a better generalization of the characteristics of the pathology when compared to the state of the art. In these circumstances, it is important to highlight that the intraclass variation is greater in the set of pathological samples, which is composed of three types of pathology. In this case, the features extraction using classical methods has a greater disadvantage because it is unable to satisfactorily relate the attributes of all pathology types.

TABLE II  
COMPARISON BETWEEN PROPOSED CNN AND SOUZA ET AL. [10]

	CNN	Souza et al. [10]
Accuracy (%)	98.3 ± 2.3	86.3 ± 2.7
Sensitivity (%)	97.1 ± 3.9	66.5 ± 6.7
Specificity (%)	100.0	89.8 ± 2.9

## VI. CONCLUSIONS

A CNN was proposed as an alternative method for the classification of recurrence plots of voice signals. A types of data augmentation made specifically for recurrence plots were applied in the network's training. In addition, the best

performance obtained in this work were compared to Souza et al. [10], which classified the same dataset with a method based on feature extraction.

In general, the approach proposed in this study was efficient. When compared to Souza et al. [10], the method proposed also had a superior performance, demonstrating average accuracy 12% higher.

#### REFERENCES

- [1] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, “Recurrence plots of dynamical systems,” *Europhysics Letters*, vol. 5, pp. 973–977, 1987.
- [2] E. G. Souza, “Caracterização de sistemas dinâmicos através de gráficos de recorrência,” Master’s thesis, Federal University of Paraná, Curitiba, 2008.
- [3] N. Marwan, *Encounters with neighbours: current developments of concepts based on recurrence plots and their applications*. Norbert Marwan, 2003.
- [4] E. Scalco and G. Rizzo, “Texture analysis of medical images for radiotherapy applications,” *The British Journal of Radiology*, vol. 90, no. 1070, p. 20160642, 2017.
- [5] M. O. Sghaier, I. Hammami, S. Foucher, and R. Lepage, “Flood hazard mapping from sar images using texture analysis and fuzzy logic,” in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2017, pp. 1900–1903.
- [6] M. L. Alves, “Reconhecimento da rugosidade em imagens monocromáticas por meio de análise de textura,” Ph.D. dissertation, Universidade Federal Fluminense, Outubro 2011.
- [7] A. Materka and M. Strzelecki, “Texture analysis methods - a review,” *Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels*, 1998.
- [8] J. R. Parker, *Algorithms for Image Processing and Computer Vision*, 2nd ed. Wiley Publishing, 2010.
- [9] R. C. Beckers, D. M. Lambregts, R. S. Schnerr, M. Maas, S.-X. Rao, A. G. Kessels, T. Thywissen, G. L. Beets, S. Trebeschi, J. B. Houwers *et al.*, “Whole liver ct texture analysis to predict the development of colorectal liver metastases—a multicentre study,” *European journal of radiology*, vol. 92, pp. 64–71, 2017.
- [10] T. A. Souza, M. A. Souza, C. d. A. Washington, S. C. Costa, S. E. Correia, and V. J. Vieira, “Feature selection based on binary particle swarm optimization and neural networks for pathological voice detection,” in *2015 Latin America Congress on Computational Intelligence (LACCI)*. IEEE, 2015, pp. 1–6.
- [11] V. Andrearczyk, “Deep learning for texture and dynamic texture analysis,” Ph.D. dissertation, Dublin City University, 2017.
- [12] N. Hatami, Y. Gavet, and J. Debayle, “Classification of time-series images using deep convolutional neural networks,” in *Tenth International Conference on Machine Vision (ICMV 2017)*, vol. 10696. International Society for Optics and Photonics, 2018, p. 106960Y.
- [13] E. Garcia-Ceja, M. Z. Uddin, and J. Torresen, “Classification of recurrence plots’ distance matrices with a convolutional neural network for activity recognition,” *Procedia computer science*, vol. 130, pp. 157–163, 2018.
- [14] Kay Elemetrics, “Disordered voice database, model 4337, 3a ed,” 1 CD, 1994.
- [15] E. Kononov, “Visual recurrence analysis (vra), version 4.9,” mar 2006. [Online]. Available: <http://nonlinear.110mb.com/vra/>
- [16] F. Takens, “Detecting strange attractors in turbulence,” in *Dynamical systems and turbulence, Warwick 1980*. Springer, 1981, pp. 366–381.
- [17] S. L. do N. C. Costa, S. E. N. Correia, W. C. de A. Costa, and V. J. D. Vieira, *Análise de Sinais por meio de Gráficos de Recorrência*. Curitiba: CRV, 2015, ch. 6, pp. 83–97.
- [18] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems*, 2nd ed. "O’Reilly Media, Inc.", 2019.