DOI: 10.61782/fa.2025.0853

FORUM ACUSTICUM EURONOISE 2025

TEXTURE ANALYSIS IN RECURRENCE PLOTS FOR PATHOLOGICAL VOICE RECOGNITION

H. Fandiño-Toro[1]* **D. Torricelli**[2] **J.A. Gómez-García**[2]
[1] Grupo de Máquinas Inteligentes y Reconocimiento de Patrones, Instituto Tecnológico Metropolitano, Colombia
[2] Spanish National Research Council, Arganda del Rey, Spain

ABSTRACT

In this paper, we explore the integration of recurrence quantification analysis (RQA) features and texture features, both extracted from unthresholded RPs. The results obtained show a balanced performance for the two classifiers considered, and that texture features could be a useful approach as a tool to extract meaningful information from recurrent plots for the task of automated pathological voice recognition.

Keywords: Pathological voice, recurrence plots, texture features

1. INTRODUCTION

The introduction of the term Recurrence Plot (RP) is attributed to Eckmann, who used it to refer to a method of visualizing the recurrence of dynamical systems [1]. An RP is a representation obtained from a recurrence matrix $\mathbf{R}_{i,j}$ described by the equation 1, with $i, j = 1, ..., N$, where N is the number of states of the system and $\approx$ denotes equality up to an error or distance ε [2].

$$\mathbf{R}_{i,j} = \begin{cases} 1: & \vec{x}_i \approx \vec{x}_j, \\ 0: & \vec{x}_i \not\approx \vec{x}_j, \end{cases} \tag{1}$$

Early applications of recurrence plots can be found in trajectory exploration in physics and chaos analysis, but their use has been extended to geophysics, industrial applications, and economics, among others.

In medicine, RPs generated from physiological signals have been analyzed to detect various pathologies. For example, RPs extracted from ECG signals have been used to analyze heart rate variability, and to detect arrhythmias. On the other hand, RPs generated from EEG signals have been used to classify motor movement/imagery signals, and to detect attention deficit hyperactivity disorder. One area that has received increasing attention in the analysis of RPs is voice analysis. This growing interest stems from the fact that speech production is considered a nonlinear dynamic system, and it is well known that the presence of various voice disorders can introduce aperiodicity and instability into the voice signal - features that can be revealed through the visual patterns of an RP [3]. Representative work on pathological voice analysis using RPs includes recurrence quantization analysis, where a reduced set of descriptors specifically designed for RPs is used to characterize them. Such an approach has been used for pathological voice detection from multiband analysis of voice signals [4], and from glottal signals [5]. Other studies have integrated features extracted from the RPs, such as dynamic invariants and others extracted from the wavelet transform [6]; and more recently, the use of RPs as input to convolutional neural networks for Parkinsonian voice detection has been reported [7]. In this paper, we explore the integration of recurrence quantification analysis (RQA) features and texture features, both extracted from unthresholded RPs. The goal of this integration is not only to identify a feature set capable of detecting pathological voice, but also to improve the interpretability of the visual patterns observed in RPs in the context of voice disorders.

**Corresponding author*: hermesfandino@itm.edu.co.

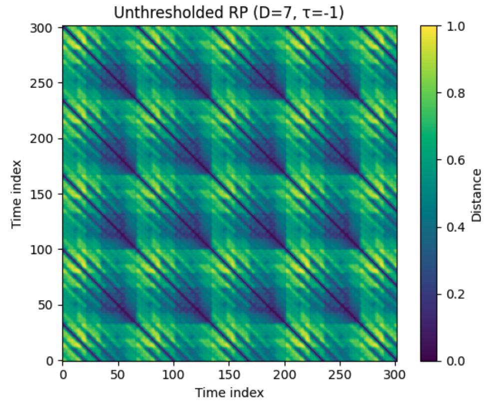
2. MATERIALS AND METHODS

2.1 Dataset

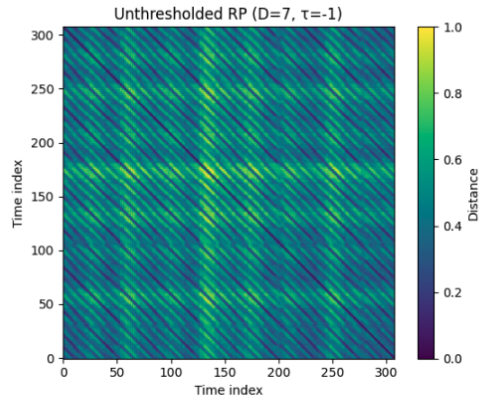
The data set considered in this paper is the HUPA data set recorded at the Príncipe de Asturias Hospital in Alcalá de Henares, Madrid, Spain. This data set contains audio signals of the sustained vowel /a/ from 366 adult Spanish speakers, including 169 with voice pathologies and 197 with normal voice quality. The recordings were made using the Kay Computerized Speech Lab Analysis Station 4300B, sampled at 50 kHz and with 16-bit resolution. The pathological samples represent a range of voice conditions such as nodules, polyps, edema, and carcinoma [8], [9–11].

2.2 Recurrence plot generation

The audio signals in the HUPA dataset were subjected to short-time analysis using Hamming windows, with window lengths of 40ms and strides of 20ms. For each frame produced, the time delay is calculated using the Rosenstein method [12], with the embedding dimension set to 7, which proved to be optimal across all signals in our internal tests. Recurrence plots are then generated for each frame using the PyRQA toolbox [13]. Finally, unthreshed recurrence plots are generated and normalized from 0 to 255 to derive the images (textures) shown in figure 1.



(a) Recurrence plot of normophonic signal



(b) Recurrence plot of pathological signal

Figure 1: Example recurrence plots.

2.3 Feature extraction

The resulting recurrence plots, contain visual information that can be characterized by texture descriptors. In this work, each recurrence plot is transformed into a feature vector by computing three sets of descriptors.

A first set contains four of the fourteen descriptors originally proposed by Haralick, which are obtained from the co-occurrence matrix of each recurrence plot. In general, a co-occurrence matrix is one in which each element denotes the number of times that a pair of pixel intensities appears at a distance d, and at a certain angle θ, in the image being evaluated. In this work, co-occurrence matrices were computed for values of $d = 1$ and $\theta = 0$, 90, 180, and 270°, from which a final average co-occurrence matrix was obtained in which the following descriptors were calculated: dissimilarity, homogeneity, AMS, energy, and correlation [14]. These descriptors have been used extensively for texture analysis in digital images, since of the original set of 14, these 4 tend to show comparatively little correlation between them. A second set includes five descriptors proposed by Tamura: Coarseness, which aims to quantify the intrinsic size of the texture elements in an image; Contrast, which measures the property of an image to simultaneously exhibit low and high pixel intensities; Directionality, which refers to the probability of finding pixel intensity variations preferentially in one direction; Line-likeliness, which measures the probability that the texture elements in an image are line-shaped; and Roughness, which refers to the overall variability or uniformity of the image's pixel intensities [15, 16]. The third set of features comprises the quantitative measures included in the PyRQA

2.4 Classification

A 5-fold group cross-validation was used to classify between pathological and normophonic signals. Care was taken to ensure that each frame belonging to a given subject was kept in the same training or test fold to avoid data leakage. The resulting features extracted from the feature set described above were then used to train two different classifiers: XGBoost and Random Forest. These classifiers were optimized using a random hyperparameter search with the parameters described in Table 1.

Table 1: Hyperparameter random grid search space for XGBoost and Random Forest classifiers.

	Classifier	
Hyperparameter	XGBoost	Random Forest
n_estimators	[50, 100, 200]	[50, 100, 200]
max_depth	[3, 6, 9]	[None, 10, 20, 30]
learning_rate	[0.01, 0.1, 0.3]	–
colsample_bytree	[0.6, 0.8, 1.0]	–
gamma	[0, 0.1, 0.5]	–
min_samples_split	–	[2, 5, 10]
min_samples_leaf	–	[1, 2, 4]

3. RESULTS

Table 2 summarizes the performance of the evaluated classifiers in terms of accuracy, precision, recall, and F1-score, expressed as the mean $\pm$ standard deviation over the five cross-validation folds. The results obtained show a relatively similar performance between the Random Forest and XGBoost classifiers in detecting pathological voice signals. This is evidenced by the average values of the performance metrics, where both classifiers achieve an overall accuracy of around 70%, and similar standard deviation values. While the XGBoost classifier shows slightly higher values in terms of precision, indicating that it generates fewer false positives when identifying pathological signals, the Random Forest classifier achieves a better recall, indicating a greater ability to correctly detect pathological signals.

Table 2: Performance of the evaluated classifiers.

	Classifier	
	Random Forest	Xgboost
Accuracy	0.7042 ± 0.0236	0.7043 ± 0.0518
Precision	0.6849 ± 0.0229	0.6972 ± 0.0493
Recall	0.6946 ± 0.0632	0.6726 ± 0.0686
F1	0.6873 ± 0.0227	0.6819 ± 0.0419

To complement the results reported in Table 2, the corresponding confusion matrices are shown in Figure 2. These matrices show that the XGBoost classifier correctly classifies a greater number of healthy signals (145 vs. 141), while the Random Forest classifier correctly classifies a greater number of pathological signals (121 vs. 117). This pattern is consistent with the trade-off between precision and recall observed above.

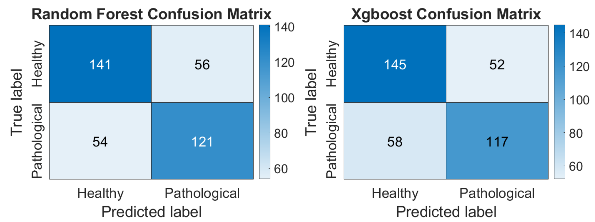


Figure 2: Confusion matrices for the two classifiers considered.

In general, Table 1 and Figure 2 indicate that both classifiers show a balanced behavior, although with slightly limited performance. A possible explanation for this behavior is that the evaluated pathological signals may introduce subtle changes in the RPs that are not easily distinguishable from healthy signals based only on the small set of texture and recurrence-based features considered in this study. Nevertheless, it should be noted that this relatively small set of features allowed us to achieve a fairly acceptable classification performance, which encourages us to further explore texture features as a possible strategy to improve the classification performance of healthy and pathological voice signals based on recurrence plot analysis.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we present the results of transforming healthy and pathological voice signals into recurrence plots, which were characterized by texture and recurrence-based features. XGBoost and Random Forest classifiers were used for classification. The results indicate that texture features could be a useful approach as a tool to extract meaningful information from recurrence plots for the task of automated pathological voice recognition. Future work should aim at exploring new texture features, as well as additional two-dimensional representation spaces of voice signals.

5. ACKNOWLEDGMENTS

This work is partially funded by the grant AI4HealthyAging with reference MIA.2021.M02.0007 funded by the ”Spanish ministry of Economy and Digital transformation” and by the “European Union NextGenerationEU/PRTR” and by the NEUROMARK project (PID2020-120491RA-I00) funded by MCIN/AEI/10.13039/501100011033.

6. REFERENCES

[1] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, “Recurrence plots of dynamical systems,” in *Turbulence, Strange Attractors and Chaos*, pp. 441–445, World Scientific, 1995.

[2] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Physics reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[3] V. J. Vieira, S. C. Costa, S. L. Correia, L. W. Lopes, W. C. d. A. Costa, and F. M. de Assis, “Exploiting nonlinearity of the speech production system for voice disorder assessment by recurrence quantification analysis,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 8, 2018.

[4] X.-C. Zhu, D.-H. Zhao, Y.-H. Zhang, X.-J. Zhang, and Z. Tao, “Multi-scale recurrence quantification measurements for voice disorder detection,” *Applied Sciences*, vol. 12, no. 18, p. 9196, 2022.

[5] M. Dahmani and M. Guerti, “Recurrence quantification analysis of glottal signal as non linear tool for pathological voice assessment and classification.,” *Int. Arab J. Inf. Technol.*, vol. 17, no. 6, pp. 857–866, 2020.

[6] S. G. Firooz, F. Almasganj, and Y. Shekofteh, “Improvement of automatic speech recognition systems via nonlinear dynamical features evaluated from the recurrence plot of speech signals,” *Computers & Electrical Engineering*, vol. 58, pp. 215–226, 2017.

[7] V. Skaramagkas, A. Pentari, D. I. Fotiadis, and M. Tsiknakis, “Using the recurrence plots as indicators for the recognition of parkinson’s disease through phonemes assessment,” in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1–4, IEEE, 2023.

[8] J. Gómez-García, L. Moro-Velázquez, and J. Godino-Llorente, “On the design of automatic voice condition analysis systems. part ii: Review of speaker recognition techniques and study on the effects of different variability factors,” *Biomedical Signal Processing and Control*, vol. 48, pp. 128–143, 2019.

[9] L. Moro-Velázquez, J. A. Gómez-García, and J. I. Godino-Llorente, “Voice pathology detection using modulation spectrum-optimized metrics,” *Frontiers in bioengineering and biotechnology*, vol. 4, p. 1, 2016.

[10] L. Moro-Velázquez, J. A. Gómez-García, J. I. Godino-Llorente, and G. Andrade-Miranda, “Modulation spectra morphological parameters: A new method to assess voice pathologies according to the grbas scale,” *BioMed research international*, vol. 2015, no. 1, p. 259239, 2015.

[11] N. Sáenz-Lechón, V. Osma-Ruiz, J. I. Godino-Llorente, M. Blanco-Velasco, F. Cruz-Roldán, and J. D. Arias-Londono, “Effects of audio compression in automatic detection of voice pathologies,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 12, pp. 2831–2835, 2008.

[12] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, “Reconstruction expansion as a geometry-based framework for choosing proper delay times,” *Physica D: Nonlinear Phenomena*, vol. 73, no. 1-2, pp. 82–98, 1994.

[13] T. Rawald, M. Sips, and N. Marwan, “Pyrqa—conducting recurrence quantification analysis on very long time series efficiently,” *Computers & Geosciences*, vol. 104, pp. 101–108, 2017.

[14] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[15] F. Bianconi, A. Álvarez-Larrán, and A. Fernández, “Discrimination between tumour epithelium and stroma via perception-based features,” *Neurocomputing*, vol. 154, pp. 119–126, 2015.

[16] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” *IEEE Transactions on Systems, man, and cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.