

Research Article

Classification System of Pathological Voices Using Correntropy

**Aluisio I. R. Fontes,^{1,2} Pedro T. V. Souza,³ Adrião D. D. Neto,²
Allan de M. Martins,³ and Luiz F. Q. Silveira²**

¹ Postgraduate Program in Electrical and Computer Engineering (PPgEEC), Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil

² Department of Computer Engineering, Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil

³ Department of Electrical Engineering, Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil

Correspondence should be addressed to Aluisio I. R. Fontes; aluisio.igor@gmail.com

Received 16 April 2014; Accepted 22 July 2014; Published 19 August 2014

Academic Editor: Pak-Kin Wong

Copyright © 2014 Aluisio I. R. Fontes et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes the use of a similarity measure based on information theory called correntropy for the automatic classification of pathological voices. By using correntropy, it is possible to obtain descriptors that aggregate distinct spectral characteristics for healthy and pathological voices. Experiments using computational simulation demonstrate that such descriptors are very efficient in the characterization of vocal dysfunctions, leading to a success rate of 97% in the classification. With this new architecture, the classification process of vocal pathologies becomes much more simple and efficient.

1. Introduction

In the past few decades, medicine has received noteworthy contributions, which have allowed for great advancement in medical activity within various contexts, for example, improvement of surgery techniques, description of the human genome, or even assistance in medical diagnosis. In particular regarding medical diagnosis, digital signal processing techniques have been employed recently as an efficient, noninvasive, and low-cost tool to analyze vocal signals, with the aim of detecting and classifying alterations in the production of sounds that may be associated with larynx pathologies [1]. The human voice is an important communication tool and any inadequate behavior may have deep implications in an individual's social and professional life.

The acoustic analysis is a complementary technique to methods based on the direct inspection of vocal folds, which may reduce the frequency of invasive exams, since such measurements are able to reveal important physiological characteristics of the vocal tract [2]. In general, most of the techniques involving the analysis of dysfunctions available in the literature employ several sorts of preprocessing stages in order to extract useful characteristics for the classification of

diverted patterns and to obtain performance improvement of classifiers [2–11].

The existing systems for detection and classification of laryngeal pathologies are phoneme-dependent, where the train step used fixed phonemes, for example, /a/ [2, 3], and diagnosis can be done by taking any vowel, which can be phonated comfortably. Various techniques have been proposed for the extraction of signal features in order to improve the performance of automatic pathological speech detection systems. A method based on MPEG-7 audio low level is proposed in [3] for the extraction of features that can be classified by support vector machine (SVM). Similarly, SVM classifiers have also been used in the features extraction based on wavelet transform, nonlinear analysis of temporal series, and information theory [4–8]. Mel-frequency cepstral coefficients and linear prediction cepstral coefficients are used as acoustic features in [9] associated with a strategy based on combining classifiers with Gaussian mixture model, hidden Markov model (HMM), and SVM. Specifically, feature extraction is achieved in [10] by using eight measures derived from the nonlinear dynamic analysis (correlation dimension, four entropy measures, Hurst exponent, the largest Lyapunov exponent, and the first minimum of mutual

information function) and quadratic discriminant analysis (QDA) by classifier [10]. Nonlinear Teager-Kaiser energy operator has been used for features extraction in [11] with a classifier based on neural system of multilayer perceptron (MLP).

Thus, it is possible to state that all aforementioned methods employ classifiers with complex architectures, for example, MLP, SVM, and HMM, and a complex stage for the features extraction, which can restrict their application or even make the technique unfeasible in real-time systems. In addition, most research works are based on the discrimination between healthy and pathological voices, but not the pathology types.

This paper introduces a new low-complexity approach for the detection and classification of laryngeal pathologies. This system uses a measure of information theory known as correntropy to characterize vocal pathologies by means of spectral characteristics and statistical moments of high order involving the vocal signal. The system is characterized by high success rates and low computational complexity associated with a simple classification stage based on Euclidian distance. The influence of the kernel size in the classification hit rate is analyzed, since it is the only free parameter of the method. Besides the classification between healthy and pathological voices, it is possible to define which pathology is most likely to exist in the voice under analysis.

The paper is organized as follows. Section 2 presents the main concepts related to correntropy. The proposed architecture is described in Section 3. In Section 4, the performance results of the proposed technique are discussed, while Section 5 presents the final considerations.

2. Correntropy

The extraction of information from a database is a frequent and relevant problem in various applications involving signal processing. Within this context, statistical measures of similarity such as correntropy may be successfully used for the extraction of information and data characterization.

Correntropy is a generalization of the correlation measure between random signals because such measure is able to extract both second-order and higher-order statistical information from the analyzed signals [12]. In the past few years, this concept been successfully applied in the solution of various engineering problems, such as the modeling of temporal series [13], nonlinearity tests [14], objects recognition [15], analysis of independent components [16], and automatic modulation classification [17]. Although correntropy is similar to correlation by definition, recent studies have demonstrated that its performance is superior when dealing with nonlinear or non-Gaussian systems without any significant increase of computational cost [12].

According to [12] correntropy is between random signals $\{x_t : t \in T\}$, where t represents time and T is the set of interest indices defined as

$$\begin{aligned} V(t_1, t_2) &= E[k_\sigma(x_{t_1}, x_{t_2})] \\ &= \iint k_\sigma(x_{t_1}, x_{t_2}) p(x_{t_1}, x_{t_2}) dx_{t_1} dx_{t_2}, \end{aligned} \quad (1)$$

where $E[\cdot]$ is the expectation operator and $k_\sigma(\cdot, \cdot)$ corresponds to any symmetric function defined as positive. It is observed that correntropy is defined as the joint expectation of $k_\sigma(x_{t_1}, x_{t_2})$. In this work, $k_\sigma(\cdot, \cdot)$ is a Gaussian kernel given as

$$k_\sigma(x_{t_1}, x_{t_2}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_{t_1} - x_{t_2})^2 / 2\sigma^2}, \quad (2)$$

where σ is the variance defined as the kernel size. The kernel size may be interpreted as the resolution for which correntropy measures similarity in a space with characteristics of high dimensionality [12].

In fact, by applying an extension of the Taylor series to the correntropy measure, (1) can be rewritten as [12]

$$V(t_1, t_2) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n \sigma^{2n} n!} E[\|x_{t_1} - x_{t_2}\|^{2n}]. \quad (3)$$

It is possible to notice that (3) defines the sum of infinite moments of even order, which therefore includes its own conventional covariance. Consequently, correntropy contains information of infinite statistical moments involving its data.

It is interesting to observe that the kernel size in (3) is a parameter that ponders the influence of the second-order and higher-order moments. For sufficiently large values, the second-order moments are dominant and the measure gets closer to the correlation.

When only a finite amount of data is available, that is, $(x_n, x_n)_{n=1}^N$, it is possible to define the estimator to the autocorrentropy function as [18]

$$\widehat{U}(m) = \frac{1}{N-m} \sum_{n=m}^{N-1} k_\sigma(x(n), x(n-m)). \quad (4)$$

The autocorrentropy with an estimator defined by (4) does not ensure zero average even when the input data are centralized due to the nonlinear transforms produced by the kernel. However, a centralized estimator for correntropy has been defined in [18], which is given as

$$\begin{aligned} \widehat{V}(m) &= \frac{1}{N-m} \sum_{n=m}^{N-1} k_\sigma(x(n), x(n-m)) \\ &\quad - \sum_{n=1}^N \sum_{m=1}^N k_\sigma(x(n), x(n-m)). \end{aligned} \quad (5)$$

This work explores the spectral properties of centralized autocorrentropy in voice signals in order to detect and classify vocal pathologies. Specifically, the spectral descriptors of healthy pathological voices used in the classification are extracted from the signals analyzed by the correntropy spectral density (CSD) function, defined as [19]

$$P_\sigma(w) = \sum_{m=-\infty}^{\infty} \widehat{V}(m) \exp(-jwm), \quad (6)$$

where w is the digital frequency given in radians. The CSD function may be considered a generalization of the power spectral density (PSD) of a signal.

One of the advantages of using correntropy measures in the classification of vocal signals lies in the robustness of such measures against impulsive noise due to the use of the Gaussian kernel in (5), which is close to zero, that is, $k_\sigma = (x(n), x(n - m))$ when $x(n)$ or $x(n - m)$ is an outlier. In addition, the correntropy function extract information of higher-order statistical moments present in the data, thus potentially increasing the classification efficiency the proposed technique.

3. Proposed Architecture

The classification method of vocal pathologies proposed in this paper is composed of two stages. The first stage is characterized by the extraction of descriptors for the voice signals based on the CSD defined in (6). The second stage is responsible for the classification of voices by simple metrics of the Euclidian distance.

Since the analyzed voice signals may vary in amplitude, the normalization of all signals in the database is performed. Then, the signals are clustered into two sets: one set with healthy voice signals and another one with pathological voice signals, which contain voices with edema and nodules. After the calculation of the CSD for all signals in each set, the average value of CSD is calculated for each set. The average CSDs of the sets are used as descriptors for the healthy and pathological voices. This procedure is depicted in Figure 1. An equivalent methodology is applied to obtain the descriptors for the voices with edema and nodules. This procedure is detailed in Figure 2.

The CSDs of the healthy and pathological voices and the respective descriptors stored for the classification stage are shown in Figures 3 and 4. It is possible to observe in Figure 3 that the average CSDs in the healthy voices have correntropy and frequency values that are different from those in the average CSDs of pathological voices. On the other hand, it may be observed in Figure 4 that the average CSDs for voices with edema and nodules have different correntropy values, even though the frequencies are similar. In order to decrease computational cost and improve the success rate, the descriptors of different classes considered in this work are constituted by the first fifty samples because there is a clear distinction involving the respective average CSDs of each class within this interval.

The classification architecture proposed in this work is presented in Figure 5. Initially, the desired voice signal to be calculated is normalized and its respective CSDs are obtained from (6). Then, the Euclidian distance between the voice CSD and each descriptor for the classes defined as healthy and pathological voices is calculated. The closer descriptor obtained according to the Euclidian distance criterion defines the class of such signal. If the analyzed voice is classified as pathological, it is necessary to apply a new classification stage in order to distinguish between an edema and a nodule.

4. Experiments and Results

The proposed architecture has been validated by computer simulation performed in MATLAB. The architecture

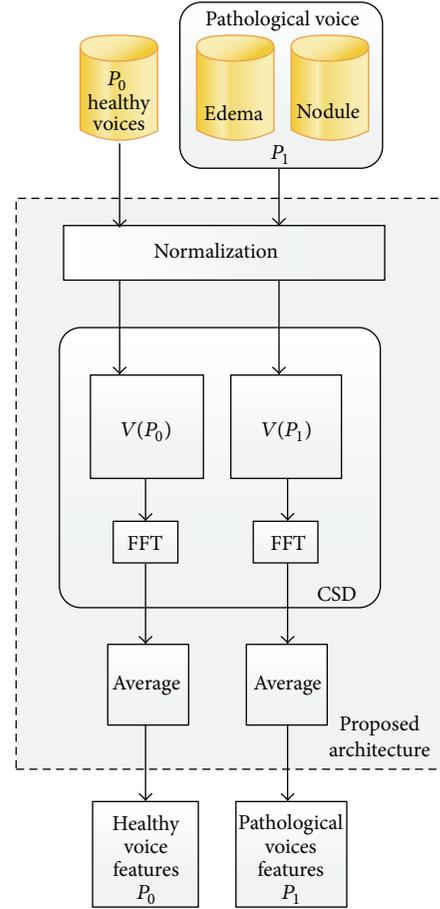


FIGURE 1: Extractor for the characteristics of healthy and pathological voices.

performance was evaluated according to Monte Carlo's method. For each experiment a minimum number of 100 trials were used. The set of voices for each class is divided into two subsets: one to extract the descriptors and another for the classification test. Crossvalidation is used in this case, where 50% of the voices are for the extraction of descriptors and the remaining 50% are used for test purposes. The evaluation of the architecture is based on the average success rates and also maximum, minimum, and standard deviation.

4.1. Database. The database used for this work was developed by Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech [20]. This database has been widely used in international research works, in acoustic analysis of disordered voices, and in the discrimination between healthy and pathological voices. It contains the sustained pronunciation of vowel "a" with 116 files from distinct speakers, where 53 are healthy voices, 43 are voices affected by edema, and 20 are voices with nodules. All used signals have the duration from 1 to 3 seconds, sampling frequency of 25 kHz, and resolution of 16 bits.

4.2. Experiments. Initially, 43 voices affected by edema and 20 voices with nodules are joined in one database called pathological. The first experiment is characterized by extracting

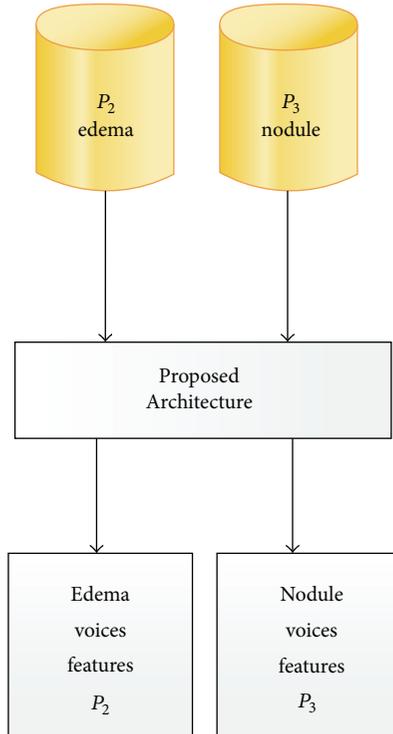


FIGURE 2: Extractor for the characteristics of voices with edema and nodules.

the success rate between healthy and pathological voices. Then the success rate is assessed among 43 voices with edema and 20 voices with nodules.

The only adjustable parameter in the architecture is the variance of the Gaussian kernel, that is, the kernel size. Many systems use a heuristic known as Silverman's rule to determine such variance [21]. However, this work has employed a numerical evaluation method to determine the influence of the kernel size on the accurate classification rate for the proposed architecture. The performed experiments are represented in Figures 6 and 7.

From the aforementioned experiments, it is possible to determine suboptimal values for the kernel width associated with each class of vocal disease. The kernel size was tested by using a logarithmic scale with values ranging from 0.01 to 10. The kernel adjustment has provided an effective mechanism to eliminate outliers. The right choice of its size may increase the success rate considerably. The amount of samples is also a fundamental factor in the classification rates, according to the results given in Figures 6 and 7.

The results shown in Figure 6 demonstrate that the independently of the kernel size the classifier for healthy and pathological voices represents an unsatisfactory result when only a few samples are used. However, when the kernel is equal to 0.77 and about 1,000 signal samples are used, the classifier has presented a success rate of 93%.

One of the goals of this work is to reduce the complexity associated with the architecture, while a classifier based on the Euclidean distance is used. However, it is sensitive to

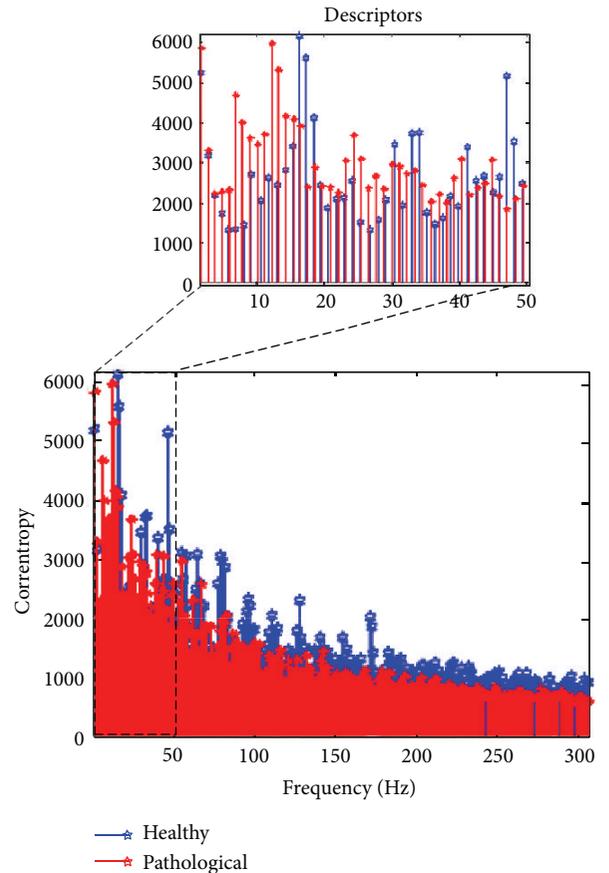


FIGURE 3: Correntropy spectral density for healthy and pathological voices obtained for a kernel size equal to 0.003.

TABLE 1: Performance of classification for healthy and pathological classes with kernel size of 0.77 and 1,000 samples.

Method	Recognition rate (%)			
	Minimum	Average	Maximum	Standard deviation
Correntropy	91.22	93.81	95.65	1.30
Correlation	60.42	71.43	82.33	9.12

the sample size as it is necessary optimize such parameter. Thus, it is possible to state that amounts of samples higher than 1,000 affect success rates between healthy and pathological voices.

The success rate in the classification between voices with edema and nodules varied from 65% (when the kernel size is 0.31 and the number of samples is 700) to 98% (when the kernel size is 1.77 and 1,300 samples are adopted), according to Figure 7.

After adjusting the architecture with the adequate values for the kernel size and number of samples, the correntropy measure was investigated regarding its capability of classifying and characterizing the statistical independencies of the considered voice signals. Accordingly, correlation is adopted as a reference measure since it can be seen as a particular case of correntropy and is very often used in classification

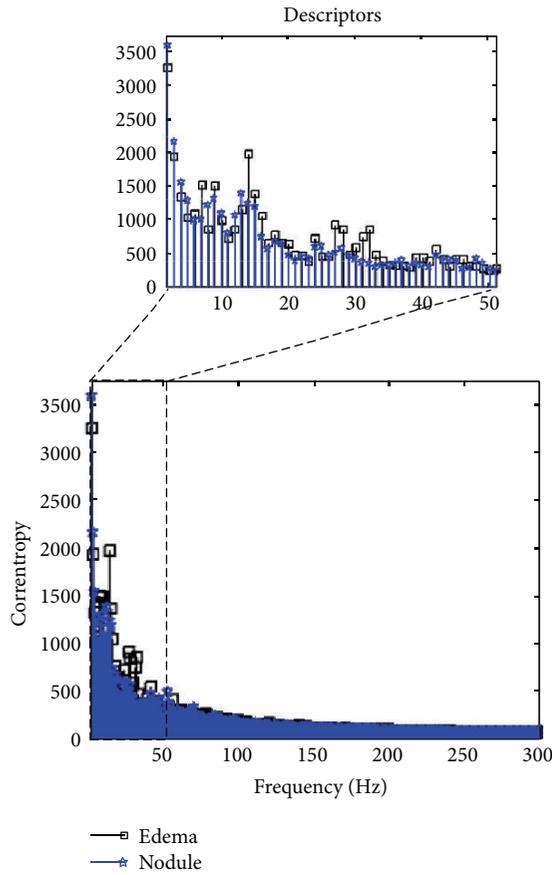


FIGURE 4: Correntropy spectral density for voices with edema and nodules obtained for a kernel size equal to 0.0129.

problems. The obtained results are presented in Tables 1 and 2. The average success rate is determined considering 100 trials.

Based on Table 1, it is possible to observe that the architecture based on correntropy is able to distinguish between healthy and pathological voices with a success rate between 91.22% and 95.65% and average rate of 93.81%, while the standard deviation is 1.30%. On the other hand, when an architecture based on correlation is used, the success rate of the classifier is reduced considerably. Besides, it can be stated that the standard deviation of the recognition rate for the proposed architecture is much lower if compared to that for the architecture based on correlation, thus indicating higher reliability of the classifier with correntropy.

From Table 2, it can be seen that the recognition rate between the voices with edema and nodules is considerably higher in the architecture based on correntropy. Once again, the success rate of the proposed architecture is within an acceptable range of values and with low standard deviation.

4.3. Comparison with Existing Methods. Several methods for the detection of vocal pathologies have been proposed in the literature. With the aim of better assessing the classifier developed in this work, the obtained results are compared with those regarding the works in [6, 8, 10], which were also obtained using the MEEI database [20].

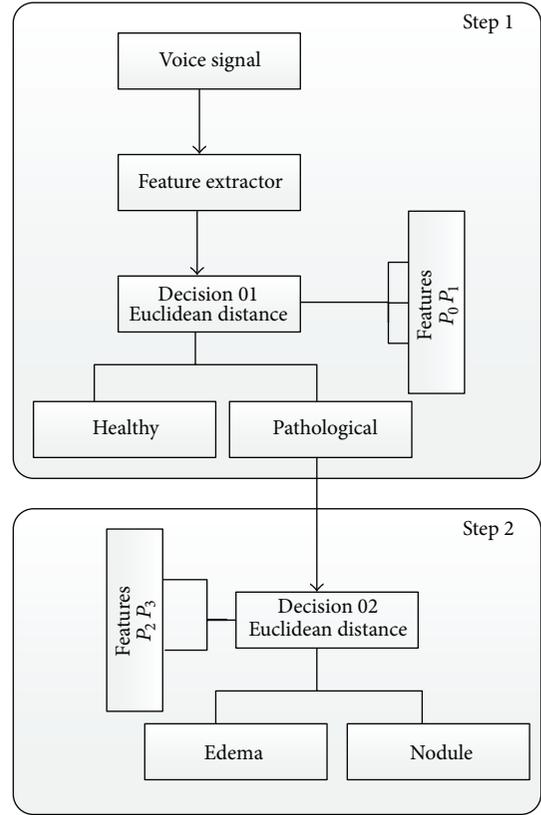


FIGURE 5: Automatic classification process of pathological voices.

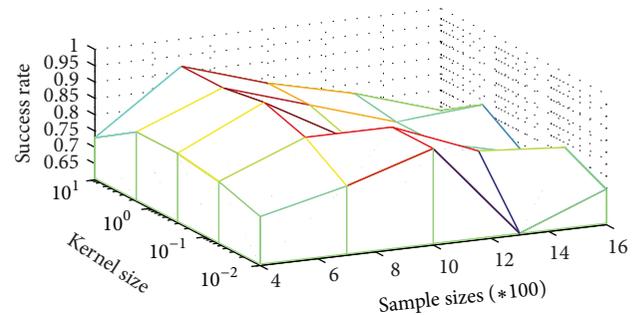


FIGURE 6: Influence of the kernel size and amount of samples on the success rates between healthy and pathological voices.

TABLE 2: Performance of classification for edemas and nodules with kernel size of 1.77 and 1,300 samples.

Method	Recognition rate (%)			
	Minimum	Average	Maximum	Standard deviation
Correntropy	95.55	96.43	98.23	1.43
Correlation	77.83	79.75	83.12	6.43

The work in [6] considers a transform initially applied to the voice signal to obtain a space of smaller dimension by using the decomposition in singular values, that is, higher-order singular value decomposition (HOSVD). In the classification stage, measure of mutual information and a

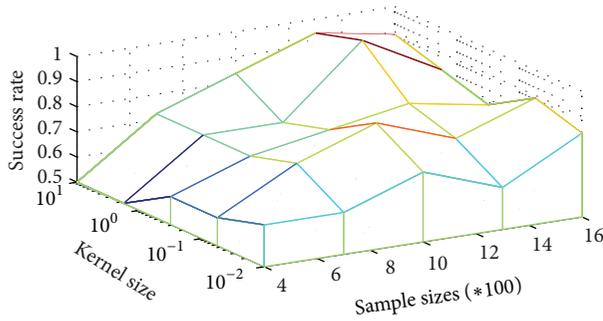


FIGURE 7: Influence of the kernel size and amount of samples on the success rates between voices with edema and nodule.

SVM network are used. The success rate is around 94.1% with an interval of reliability of 0.28%.

The classification architecture in [8] is developed from eleven characteristics extracted by means of nonlinear analysis of temporal series, where two are based on conventional nonlinear statistics, other two are based on the analysis of recurrence and fractal scheduling, and the rest of them are obtained from different estimations of the entropy. The achieved success rate is 98% by using a SVM and Gaussian mixture models (GMM) in the classification stage.

Information measures are employed in [10], for example, Shannon entropy, correlation entropy, approximate entropy, Tsallis entropy, Hurst exponent, maximal Lyapunov exponent, and the first minimum of the mutual information function, in addition to LPC (linear prediction coding) coefficients [10]. In the classification process, a quadratic discriminant analysis (QDA) is applied with a success rate equal to about 96.50%.

Thus, it is possible to state that all aforementioned methods employ a complex stage for the extraction of characteristics, with the calculation of a large set of variables that, in general, are sent to a neural network for classification purposes.

On the other hand, the architecture proposed in this paper uses only one extractor defined by the CSD and also a very simple classification stage based on Euclidian distance. Therefore, the introduced strategy presents low computational complexity, which implies simple implementation in real-time embedded systems. Besides, the proposed system presents high success rate, that is, about 97%.

5. Conclusion

This paper has presented a novel method of automatic classification for pathological voices based on correlogram spectral density (CSD). It has been demonstrated that CSD is adequate to characterize dynamic interdependencies among the voice signal samples, being able to extract distinct characteristics between healthy and pathological voices. Among the main characteristics of such method, it is worth mentioning that the classification stage becomes simpler by the use of Euclidian distance, which effectively reduces its computational complexity. From the obtained results, it has been shown

that the proposed classifier presents high recognition rate, which is achieved after a simple adjustment in the kernel size employed by the feature extractor. The proposed method can be used as a valuable tool by researchers and speech pathologists. Future work aims at the development of experiments using other databases and also the implementation of an online diagnosing system.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] B. G. Aguiar Neto, S. C. Costa, J. M. Fechine, and M. Muppa, "Feature estimation for vocal fold edema detection using short-term cepstral analysis," in *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE '07)*, pp. 1158–1162, January 2007.
- [2] J. I. Godino-Llorente, P. Gómez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 10, pp. 1943–1953, 2006.
- [3] G. Muhammad and M. Melhem, "Pathological voice detection and binary classification using MPEG-7 audio features," *Biomedical Signal Processing and Control*, vol. 11, pp. 1–9, 2014.
- [4] R. T. S. Carvalho, C. C. Cavalcante, and P. C. Cortez, "Wavelet transform and artificial neural networks applied to voice disorders identification," in *Proceedings of the 3rd World Congress on Nature and Biologically Inspired Computing (NaBIC '11)*, pp. 371–376, IEEE, October 2011.
- [5] E. S. Fonseca, R. C. Guido, A. C. Silvestre, and J. C. Pereira, "Discrete wavelet transform and support vector machine applied to pathological voice signals identification," in *Proceedings of the 7th IEEE International Symposium on Multimedia (ISM '05)*, pp. 785–789, December 2005.
- [6] M. Markaki and Y. Stylianou, "Voice pathology detection and discrimination based on modulation spectral features," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 1938–1948, 2011.
- [7] E. S. Fonseca and J. C. Pereira, "Normal versus pathological voice signals," *IEEE Engineering in Medicine and Biology Magazine*, vol. 28, no. 5, pp. 44–48, 2009.
- [8] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez, "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 370–379, 2011.
- [9] S. Jothilakshmi, "Automatic system to detect the type of voice pathology," *Applied Soft Computing*, vol. 21, pp. 244–249, 2014.
- [10] W. C. A. Costa, S. L. N. C. Costa, F. M. Assis, and B. G. Aguiar, "Healthy and pathological voice assessment by means of nonlinear dynamic analysis measures and linear predictive coding," *Brazilian Journal of Biomedical Engineering*, vol. 29, no. 1, pp. 3–14, 2013.
- [11] L. Salhi and A. Cherif, "Robustness of auditory teager energy cepstrum coefficients for classification of pathological and normal voices in noisy environments," *The Scientific World Journal*, vol. 2013, Article ID 435729, 8 pages, 2013.

- [12] I. Santamaría, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: definition, properties, and application to blind equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6 I, pp. 2187–2197, 2006.
- [13] A. Gunduz and J. C. Principe, "Correntropy as a novel measure for nonlinearity tests," *Signal Processing*, vol. 89, no. 1, pp. 14–23, 2009.
- [14] I. Park and J. C. Principe, "Correntropy based Granger causality," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3605–3608, April 2008.
- [15] K. Jeong, W. Liu, S. Han, E. Hasanbelliu, and J. C. Principe, "The correntropy MACE filter," *Pattern Recognition*, vol. 42, no. 5, pp. 871–885, 2009.
- [16] R. Li, W. Liu, and J. C. Principe, "A unifying criterion for instantaneous blind source separation based on correntropy," *Signal Processing*, vol. 87, no. 8, pp. 1872–1881, 2007.
- [17] A. I. R. Fontes, L. A. Pasa, V. A. De Sousa Jr., F. M. Abinader Jr., J. A. F. Costa, and L. F. Q. Silveira, "Automatic modulation classification using information theoretic similarity measures," in *Proceedings of the 76th IEEE Vehicular Technology Conference, (VTC '12)*, pp. 1–5, September 2012.
- [18] J. C. Principe, *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*, Springer, 2010.
- [19] J. Xu, *Nonlinear signal processing based on reproducing Kernel Hilbert space [Ph.D. thesis]*, University of Florida, 2007.
- [20] M. Eye and E. Infirmay, *Elemetrics Disordered Voice Database (Version 1.03)*, Voice and Speech Lab., Boston, Mass, USA, 1994.
- [21] B. W. Silverman, "Density estimation for statistics and data analysis," *Technometrics*, vol. 37, 1986.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

