ORIGINAL RESEARCH

Robust Detection of COVID-19 in Cough Sounds

Using Recurrence Dynamics and Variable Markov Model

Pauline Mouawad[1] · **Tammuz Dubnov[2]** · **Shlomo Dubnov[3]**

Abstract

COVID-19, otherwise known as the coronavirus, has precipitated the world into a pandemic that has infected, as of the time of writing, more than 10 million persons worldwide and caused the death of more than 500,000 persons. Early symptoms of the virus include trouble breathing, fever and fatigue and over 60% of people experience a dry cough. Due to the devastating impact of COVID-19 and the tragic loss of lives, it is of the utmost urgency to develop methods for the early detection of the disease that may help limit its spread as well as aid in the development of targeted solutions. Coughs and other vocal sounds contain pulmonary health information that can be used for diagnostic purposes, and recent studies in chaotic dynamics have shown that nonlinear phenomena exist in vocal signals. The present work investigates the use of symbolic recurrence quantification measures with MFCC features for the automatic detection of COVID-19 in cough sounds of healthy and sick individuals. Our performance evaluation reveals that our symbolic dynamics measures capture the complex dynamics in the vocal sounds and are highly effective at discriminating sick and healthy coughs. We apply our method to sustained vowel 'ah' recordings, and show that our model is robust for the detection of the disease in sustained vowel utterances as well. Furthermore, we introduce a robust novel method of informative undersampling using information rate to deal with the imbalance in our dataset, due to the unavailability of an equal number of sick and healthy recordings. The proposed model achieves a mean classification performance of 97% and 99%, and a mean F_1-score of 91% and 89% after optimization, for coughs and sustained vowels, respectively.

Keywords COVID-19 · Nonlinear dynamics · Symbolization · Variable Markov Oracle · Recurrence quantification · Machine learning · Music information dynamic · Information theory

✉ Pauline Mouawad
pauline.mouawad@lau.edu.lb

Tammuz Dubnov
tammuz@zuzor.co

Shlomo Dubnov
sdubnov@ucsd.edu

[1] Lebanese American University, Byblos, Lebanon

[2] Zuzor, San Diego, CA, USA

[3] UC San Diego, La Jolla, CA 92093, USA

Introduction

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus that starts by infecting the mucous membranes in the throat and moves down the respiratory tract leading to the lungs, with coughing being a common symptom. Cough sounds contain underutilized pulmonary health information that can be analyzed. This paper focuses on the auditory effects that symptoms of COVID-19 can have on individuals, as detected through the analysis of audio recordings of an individual's coughs and utterances. COVID-19 is known to result in a dry cough in 67% of cases and in phlegm production in 33% of cases. Detection of these symptoms can be learned using machine learning methodologies trained on recorded samples from healthy and sick individuals.

Previous Works on Respiratory Voice Diagnostics

In recent years, several studies have proposed acoustic features for the detection of pathologies and respiratory diseases in voice signals. Common attributes include MFCCs (1-13) [41], vowel formants [49], jitter, shimmer and mean harmonics-to-noise ratio (HNR) [25], to name a few.

Santosh [59] proposes a model where speech, signal processing and image analysis techniques are integrated in healthcare. The author notably emphasizes the usefulness of speech processing techniques in helping doctors predict the presence of tuberculosis in a patient's verbal communication, identify pain in voice/speech level, as well as understand the patient's willingness to continue with treatment. The proposed model suggests the use of machine learning tools such as convolutional neural network to make unbiased decisions in healthcare.

Various diseases such as asthma, tuberculosis, bronchitis, among others, can have very prominent effects on the pulmonary system and have been shown to be identifiable through signal processing analysis of the sounds of coughs. Abeyratne et al. [1] analyzed the differences in pneumonia coughs, asthma coughs, and bronchitis coughs. Using a combination of time series statistics, formant-frequency tracking and general temporal-spectral energy-based features, they were able to achieve a sensitivity of 94% and specificity of 75% based on parameters extracted from the cough sounds alone through non-contact microphones.

Swarnkar et al. [63] use a variety of signal processing methods to look at the cough sounds including analyzing the spectral energy, temporal envelope, and time-independent waveform statistics and was able to reach a recall of 55% and specificity of 93% of classifying wet versus dry coughs when trained on 536 samples. A study in [2] showed that the coughs from asthmatic patients had a measurably higher energy signatures, especially in the low-frequency bins, than the coughs from the non-asthmatic patients. A comprehensive review of the detection and analysis of voice pathology from acoustic analysis

may be found in [8], and an overview of the initiatives so far taken to combat with the present COVID-19 pandemic is found in [14]. Table 1 reports recent related works from the literature.

As the amount of research focusing on COVID-19 increased, recent works started to investigate the usage of deep neural networks to classify individuals as sick based on cough sounds. Previously, Nallanthighal et al. [51] utilized CNN and RNN architectures for breath event detection as a potential indicator for COPD, asthma, and general respiratory infections. More recently, Imran et al. [32] utilized CNN architectures to perform direct COVID-19 diagnostic classifications based on cough sounds. Unlike the approach discussed in this paper, [32] utilizes transfer learning approaches with deep learning to achieve performance similar to ours with an F1 score of 0.929. Their work is based on a model pre-trained on general sounds and then fine-tuned on COVID-19 data. Their COVID-19 dataset is larger than the dataset used in this study and could not be accessed. Overall, their system has many components that are hard to evaluate individually therefore making it difficult to compare directly [32, 51].

Non-linear Dynamic Approaches to Vocal Analysis

Although acoustic analysis studies of pathological voice signals are widespread, they mostly assume implicitly that the information required to characterize disorders in voice are depicted in the instantaneous signal's acoustic properties rather then in their temporal dynamics. Chaos theory, an area of nonlinear dynamics systems theory, has been recently applied to time series analysis, opening a new approach to nonlinear speech signal processing [68]. Numerous studies have recently shown that nonlinear phenomena exist in vocal signals [27, 29] and evidence of nonlinear dynamical behaviour in speech signals and pathological voices had been found in [30, 31]. In [68], nonlinear dynamics features as well as entropy measures were used to discriminate between healthy and pathological voices.

Table 1 Cough classification tasks showing results, methods and classifier types: K-nearest neighbors (KNN) and logistic regression model (LRM)

Publication	Task	Results	Methods
Al-khassaweneh [2]	Asthma	Sensitivity: 88%	Spectral estimation with KNN
Subburaj [62]	Intensity	Accuracy: 98%	Temporal energy-based regression
Swarnkar [63]	Wet/Dry	Sensitivity: 55% Specificity: 93%	A variety of features with a LRM
Chatrzarrin [10]	Wet/Dry	Sensitivity: 100% Specificity: 100%	Spectral / temporal theresholding
Abeyratne [1]	Pneumonia	Sensitivity: 80% Specificity: 73%	A variety of features with an LRM
Botha [7]	Tuberculosis	Sensitivity: 82% Specificity: 95%	Log-spectral bands with an LRM and clinical metrics

One of the basic properties of nonlinear dynamic systems is recurrence, or approximate repetition, of the system states over time. To capture such recurrence patterns, the phase space that represents system states and a distance measure between the states needs to be defined. This allows creating so-called Recurrence Plot (RP), a graphical representation of a square matrix where patterns correspond to points in time at which recurrence of the system's state occurs, and from which additional statistics can be extracted.

An analysis of speech signals based on quantification metrics of recurrence plots showed that recurrence quantification measures significantly discriminate healthy from pathological voices [74]. In [39], the authors use recurrence and fractal scaling measures to distinguish between pathological disorders in voices. In [48], a symbolic nonlinear dynamics analysis using recurrence quantification measures was applied for the detection of affect in vocal and musical stimuli, as well as in auditory scenes.

In this paper, we apply the so called Recurrence Quantification Analysis (RQA) to characterize healthy and COVID-19 diagnosed subjects from the recurrence plot of their audio recordings. RQA has attracted increasing attention in the time series community, and measures based on recurrence plots have been applied in various scientific disciplines [5, 20, 33, 42, 45, 55–57, 69, 75]. As will be discussed in the theory section, there are two main methods to obtain recurrence features—by extracting statistics directly from RP, or by symbolization of the data and considering the equivalent symbolic dynamics [16, 38]. Adaptive symbolization of audio signals was previously used for repeated theme discovery in musical recordings in [72].

With such various new systems being developed, to date there is still no benchmark model for the identification of diseases from cough sounds or other vocalisations with the main issue being the less than satisfactory detection precision. In this paper, we apply a combined method from nonlinear dynamics that uses adaptive time series symbolization with RQA features to detect COVID-19 in vocal signals.

By the time of writing and to the best of our knowledge, no prior work has so far attempted to diagnose COVID-19 in cough sounds in a comparably reliable manner. The robust results of our model have the potential to substantially aid in the early detection of the disease on the onset of seemingly harmless coughs, and is central towards the development of a broad range vocal-based ambulatory analysis system.

Theoretical Background

Concepts and theories of chaotic dynamics have been recently applied to complex music signals [24, 43], voice [30] such as infant cries [47], pathological voices [4, 27, 31] and emotional speech [28, 29, 54]. They are shown to be powerful methods for characterizing complex signals in terms of their inherent nonlinear dynamics. A time series comprises time-ordered measurements computed from a large sample of data that represent a small observation of the underlying dynamical system. Nonlinear time series analysis (NLTSA) [34] consists of a set of methods that describe dynamical information from temporal series of real-world systems. NLTSA considers that all the information needed to determine the future behaviour of the system's state is independent of its past and can be predicted based on knowledge of the present state. Methods for the quantification of the underlying dynamics in time series are found in the literature [45] and are detailed next.

Mel Frequency Cepstral Coefficients

Standard ways to consider similarity in audio signals is through time–frequency representation. The Mel frequency cepstral coefficients (MFCCs) are spectral attributes that are obtained using a frequency transform of the log spectrum. MFCCs have a frequency resolution similar to that of the human ear [67] and, therefore, represent the nonlinear human auditory response to audio frequency. MFCCs are widely successful in the recognition of various types of audio signals as well as in various human speech processing tasks and are standard in such studies. Some examples include the use of MFCC-based features for language identification systems [50], speech emotion recognition [53], and speaker identification [65]. In [50], a comprehensive review on the use of MFCC-based features for language identification is presented. Authors further propose a second-level MFCC-based feature (MFCC-2) that handles the large and uneven dimensionality of MFCC.

Recurrence Quantification Analysis

Recurrence is a fundamental property of most dynamical systems. Due to a system's recurrence to former states, we know how to make predictions about its future state. Recurrence takes place in the system's phase space, and the tool that measures a recurrence of a trajectory in phase space is called a recurrence plot (RP) [45]. Recurrence plots (RPs) and recurrence quantification analysis (RQA) are robust methods for analyzing recurrences in temporal series. A recurrence plot is a graphical representation of a squared matrix with black and white dots at two time axes, highlighting recurrent states in the structural dynamics underlying the signal.

Given a trajectory $\mathbf{x}_i \in \mathbb{R}^d$ in a d-dimensional phase space of a dynamical system, the RP is a two-dimensional visualization of the square recurrence matrix of the embedded time series defined by

$$R_{i,j}^{m,\varepsilon} = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i,j = 1, ..., N, \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are phase space trajectories in an m-dimensional phase space, N is the number of measured points in a trajectory, ε is a threshold distance, $\Theta(.)$ the Heaviside function such that: $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ otherwise, and $\|.\|$ is some appropriate choice of a norm, such as the L_2-norm otherwise known as the Euclidean distance. Both axes of the RP are time axes. The dots or pixels located at (i, j) and (j, i) on the RP are black if the distance between points x_i and x_j in the phase space fall inside a *ball* or *threshold corridor* of radius ε, the threshold distance [9, 55]. In this case, the black points refer to *recurring* states also termed ε-recurrent states since they occur in an ε-neighborhood. The ε-recurrent states are represented by the relation [45]:

$$\mathbf{x}_i \approx \mathbf{x}_j \iff R_{i,j} \equiv 1. \tag{2}$$

The dots are white if $R_{i,j} \equiv 0$. The RP always displays a main black diagonal line called the line of identity (LOI) since $R_{i,i} \equiv 1$ by definition. For more in-depth description of the RP properties, the reader is referred to [45].

Recurrence Quantification Measures

Recurrence Quantification Analysis (RQA) is a nonlinear statistical technique that quantifies the structures of an RP by means of various complexity measures [9, 61, 75]. RQA measures are based on the density of recurrence points, the diagonal and the vertical line structures in the RP. Fifteen RQA measures are computed in this work and they are detailed next.

Measures Based on the Density of Recurrence Points

Given an RP thresholded at ε (Eq. 1), the Recurrence Rate (*RR*) measures the density of recurrence points in the RP:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}. \tag{3}$$

The RR corresponds to the correlation sum (D2) measure, but D2 excludes the main diagonal line (LOI):

$$D2 = \frac{1}{N(N-1)} \sum_{\substack{i,j = 1 \\ j \neq i}}^{N} R_{i,j}. \tag{4}$$

Diagonal Line-Based Measures

The following RQA measures are computed from the histogram $P(l)$ of diagonal lines of length l [61]:

Determinism (*DET*) is the ratio of recurrence points in the diagonals to all recurrence points. *DET* measures the predictability of the system.

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)}. \tag{5}$$

The average length of diagonal line length L refers to the mean prediction time:

$$L = \frac{\sum_{l=l_{min}}^{N} lP(l)}{\sum_{l=l_{min}}^{N} P(l)}. \tag{6}$$

The length L_{max} of the longest diagonal line in the RP excluding LOI is

$$L_{max} = max(\{l_i\}_{i=1}^{N_l}). \tag{7}$$

The inverse of L_{max} indicates the divergence (*DIV*) of the phase space trajectory. The faster the divergence of the trajectory segments, the shorter the diagonal lines, and *DIV* has higher value:

$$DIV = \frac{1}{L_{max}}. \tag{8}$$

The next measure is the Shannon entropy of diagonal line length distribution in the RP (S_{RP}), which is the probability $p(l) = P(l)/N_l$ to find a diagonal line of exactly length l in the RP. It is a measure of complexity in the RP in terms of the diagonal lines, such that, for uncorrelated noise the value of S_{RP} will be small, which indicates a low complexity.

It is defined as

$$S_{RP} = -\sum_{l=l_{min}}^{N} p(l) \ln p(l). \tag{9}$$

The *RATIO* is a measure that uncovers transitions in the system's dynamics:

$$RATIO = \frac{DET}{RR}. \tag{10}$$

Vertical Line-Based Measures

Measures based on vertical structures in the RP uncover chaos–chaos transitions [46] in a dynamical system that are not found using diagonal line-based measures. These are *laminarity* and *trapping time*.

The laminarity (*LAM*) refers to the occurrence of laminar states in the system. If the RP contains less vertical lines and more single recurrence points, then the value of *LAM* will be low:

$$LAM = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}. \tag{11}$$

The trapping time measure (TT) is the average length of vertical lines, and estimates the mean time that the system's state will be trapped:

$$TT = \frac{\sum_{v=v_{min}}^{N} vP(v)}{\sum_{v=v_{min}}^{N} P(v)}. \tag{12}$$

Our Model

When computing RP-based measures, a key factor is to construct an RP that exhibits enough recurrence points. Another difficulty to address is the length of the sequence used to generate the RP. A widely known approach for RP construction is based on phase space reconstruction from time series [34, 60] using Takens time-delay embedding theorem [64]. However, one important limitation is that accurate embeddings are not easy to construct [33], and the recurrence quantification measures that describe the underlying dynamics are not invariant for different values of the embedding parameters, dimension m and time delay τ. This is a major concern for any system that relies on RQA estimates to make accurate predictions. It was previously suggested that with methods of recurrence plot analysis it may not be necessary to embed the data, and that NLTSA methods that circumvent the need for embedding are desired [33]. Another concern is the case of discrete-valued time series. Most of the time series analysis methods have been developed for continuous-valued time series only, and when they are applied to discrete-valued observables, problems are encountered in interpreting the RQA results. Additional problems are also encountered if the variability of a system happens at very different time-scales [16]. These issues can be addressed using symbolic time series analysis that encodes a time series into a sequence of discrete symbols, from which recurrence statistics are estimated to characterize the dynamics of the system.

Our method has the following key advantages: first, it detects recurrences in symbolic recurrence plots and quantifies them using symbolic recurrence quantification analysis (RQA) measures while bypassing phase space reconstruction and time-delay embedding. Second, it allows the construction of symbolic RPs that prune irrelevant information by filtering only the interesting aspects of the system's dynamics; hence, the symbolic RQA measures describe only the essential information in the signal. A third and novel aspect of our approach is the use of an information theoretic measure, the information rate (IR), in an informative undersampling method using a global thresholding approach. This novel aspect is detailed in Sect. Information Rate.

We start by transforming the signal into Mel frequency cepstral coefficients (MFCCs). Next we perform a symbolization step using the Variable Markov Oracle (VMO) [73]. Then we compute the symbolic RP from the self-similarity matrix, and in a final step, we derive the symbolic RQA measures that constitute our final feature set used in classification. A framework of the method is shown in Fig. 1.

Variable Markov Oracle

In a previous work [48], we develop the argument on the robustness of our nonlinear dynamics method using the Variable Markov Oracle (VMO) in detecting recurrences. We provide here a brief description of the model.

The VMO is a suffix automaton that reduces a multivariate time series down to a symbolic sequence while retaining the most informative recurrent sub-sequences. The advantage of VMO symbolization is that it detects the best recurrence structure by searching over all possible symbolized sequences and selects the threshold θ that achieves the best compression effect, which offers the most revealing symbolization.

To find the best symbolization of the signal, different VMO models can be created with different θ values. There is a tradeoff to consider when choosing θ values since no structure in the time series can be captured by the VMO if the value of θ is either too low or too high. Hence, θ should be determined before VMO construction to obtain the most informative symbolization. Dubnov et al. have shown that the value of θ can be resolved by computing the Information Rate (IR) over candidate θ values and the optimal θ is the one that yields a highest IR value [19]. In short, recurrences of symbolic sequences are evaluated based on a mutual information criterion that estimates the optimal threshold θ in terms of maximizing Information Rate (IR) that considers mutual information between past and present in a signal [17].

The RP derived from VMO represents the structure of recurrent motifs of variable length. Detailed description of
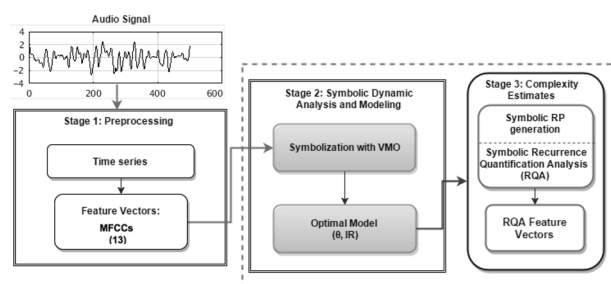

Fig. 1 Framework for symbolic RQA estimation

the algorithm is found in [3, 6, 18, 71, 73]. A previous application of our model in characterizing affective non-verbal vocal signals, music and auditory scenes can be found in [48].

Information Rate

Information rate is an information theoretic metric that measures the information content that a signal carries into its future. It is the mutual information between past and present observation in a signal $O[t]$ and is maximized when there is balance between variation and repetition in the symbolized signal. A pure random signal as well as a constant signal carry little information, resulting in an overall small IR value. This means that a VMO with a higher IR value captures more of the repeating patterns; hence, more information, than a VMO with a lower IR value. Complete details about the information theoretic framework of IR and its application to VMO construction can be found in [17, 72].

In addition to the optimal VMO symbolization, IR is significant for addressing the issue of learning from imbalanced data, a situation that arises when there is severe difference between the number of observations belonging to the minority class and those belonging to the majority class. We propose a novel method of informative undersampling using IR, in a global thresholding approach, that significantly improves the overall classification performance, described in Sect. Learning from Imbalanced Data.

Let $x_1^N = x_1, x_2, ..., x_N$ be a time series x with N observations, where $H(x) = -\sum P(x) log_2 P(x)$ is the entropy of x, then the definition of IR is

$$IR(x_1^{n-1}, x_n) = H(x_n) - H(x_n|x_1^{n-1}). \tag{13}$$

And it is approximated by replacing the entropy terms in Eq. 13 by a complexity measure C associated with a compression algorithm [72]. The complexity measure is the number of bits used to compress x_n independently using the past observations x_1^{n-1}:

$$IR(x_1^{n-1}, x_n) \approx C(x_n) - C(x_n|x_1^{n-1}). \tag{14}$$

Recurrence Plots from VMO

From the generated VMO-symbolized time series, we obtain the symbolic RP ($RP_{\mathscr{S}}$ hereafter), plotted from the binary self-similarity matrix. The index of a suffix link is a point on the $RP_{\mathscr{S}}$ and a repeated sequence is detected as a line since it includes repetitions of length 1, 2,... up to the longest repeated length. This makes VMO effectively find a repetition for variable length non-uniform embedding.

We redefine the RP as a symbolic recurrence plot $RP_{\mathscr{S}}$ obtained from the optimal VMO-symbolization of the signal's time series:

$$R_{i,j}^{\sigma_M,\theta} = \Theta(\theta - d(\sigma_{q_i}, \sigma_{q_j})) \quad i,j = 1, ..., N, \tag{15}$$

such that

$$R_{i,j}^{\sigma_M,\theta} = \begin{cases} 1 & if\ d(\sigma_{q_i}, \sigma_{q_j})\ is \leq \theta \\ \\ 0 & otherwise, \end{cases} \tag{16}$$

where N is the number of states considered, σ_M refers to the symbolized substring, Θ is the Heaviside step function (i.e. $\Theta(x) = 0$ if $x < 0$, and $\Theta(x) = 1$ otherwise). θ is a threshold distance, and $d(\sigma_{q_i}, \sigma_{q_j})$ is a distance metric between pairs of symbolized substrings q_i at $t = i$ and q_j at $t = j$.

Figure 2 shows six symbolic $RP_{\mathscr{S}}$ of six cough sounds. The sick and healthy coughs $RP_{\mathscr{S}}$ are in the left and right columns, respectively. In the plots to the left, it is possible to detect the presence of clear white bands or disruptions in the sick coughs. The white bands indicate abrupt changes in the dynamics as well as extreme events, showing that some states are far from normal. In the plots to the right, the white bands are fading. To get an objective evaluation of the structures found in $RP_{\mathscr{S}}$, we apply a symbolic recurrence quantification analysis ($RQA_{\mathscr{S}}$ hereafter) based on (Eq. 15), and we obtain symbolic measures in line with Eq. 3 to 12.

COVID-19 Detection Analysis

Dataset

Our machine learning models are trained upon the data collected by the Corona Voice Detect project in partnership with Voca.ai and Carnegie Mellon University. The data samples include self-submitted audio recordings, age, gender, country of residence (optional), smoking habits (optional) and more. The recordings contain the audio of the individual coughing three times, uttering sustained vowels "ah", "oh", "eh" among other vocal prompts. Owing to the unanticipated onset of the disease, benchmark data for analysis are not yet available. Though a controlled setting is not enforced, such as a laboratory setting that ensures the absence of noise, the self-submitted nature of audio recordings offers the advantage of being real-life occurring audio events collected in real-life settings. An early detection of the disease demands a speedy analysis of vocal prompts, hence the necessity

Fig. 2 Recurrence plots of cough sounds. Left: with COVID-19. Right: without COVID-19
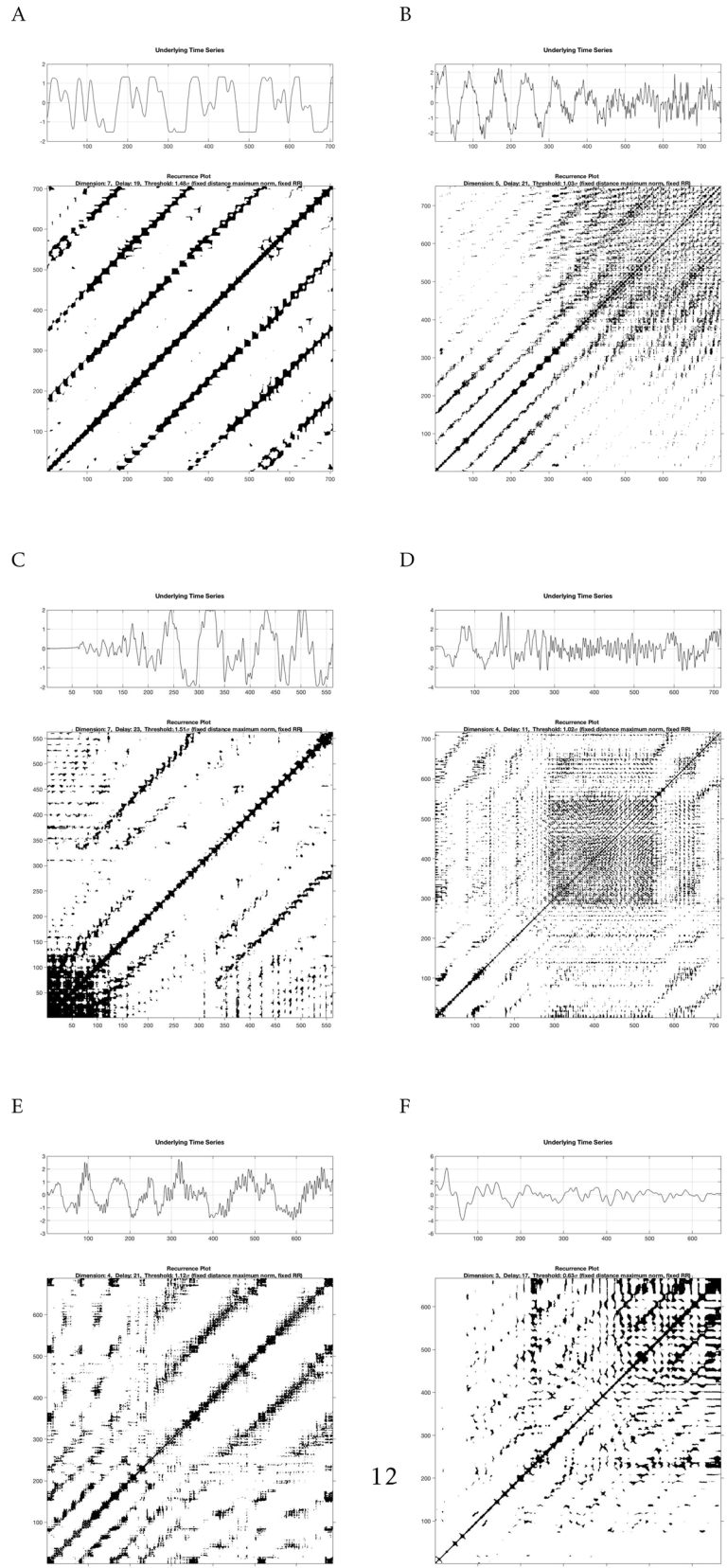
of a voluntary self-recordings of vocal symptoms in the spur of the moment. This provides a realistic analysis of the sounds as they occur when there is an apprehension of being infected.

It has been shown that the sustained vowel 'ah' is sufficient for various voice assessment applications [66, 70], voice disorders [15, 52] and spectral characteristics of vowels to efficiently gauge complex content in vocal sounds [37]. In [49], an automatic voice disorder classification task using two vowel recordings is made to discriminate cyst, paralysis and polyp. In our experiment, we use two types of vocal recordings from the dataset: coughs as well as the sustained vowel 'ah'. The cough sounds are on average about 5.46 s of length, sampled at 22050 Hz. The sustained vowels are on average about 12.66 s of length, sampled at 22050 Hz.

Learning from Imbalanced Data

The dataset of cough samples consists of 1895 healthy and 32 sick samples, and the dataset of 'ah' vowels consists of 1468 healthy and 20 sick samples. Considering that our dataset is imbalanced with the number of healthy samples considerably exceeding the number of sick samples, it is necessary to take steps to ensure that the classification model is not biased towards the majority class. To that end, we experiment with a wide range of data sampling techniques such as oversampling the minority class or undersampling the majority class. Random undersampling (RUS) of the majority class was tested; however, since it is known to lose information that may be important to fit a robust decision boundary, it was not retained.

Oversampling: Oversampling was tested using state-of-the-art variants of the Synthetic Minority Oversampling Technique (SMOTE) as well as some of its extensions that perform data cleaning of synthetic samples. Methods tested include the Adaptive Synthetic Sampling (ADASYN) algorithm, SMOTE with data cleaning using the Wilson's Edited Nearest Neighbour rule (SMOTE + ENN), SMOTE with data cleaning using Tomek Links (SMOTE + TL), Majority Weighted Minority Oversampling Technique (MWMOTE), filteriNg of ovErsampled dAta using non cooperaTive gamE theoRy (NEATER), among others. NEATER is a filtering method based on game theory, that is applied after oversampling the minority class. It discards instances with high probability of belonging to the opposite class, based on each instance's neighborhood. SMOTE + ENN generates synthetic samples from the minority class, and then cleans those samples by removing the examples that are misclassified by its three nearest neighbours. SMOTE + TL applies Tomek Links for cleaning synthetic samples. A comprehensive review of sampling methods and their applications is provided in [13, 58].

The oversampling methods were applied only to the training set during a cross-validation procedure. The test sets are never oversampled nor seen by the model during training, thus avoiding the problem of overfitting. This approach ensures a good evaluation of the classifier's generalizability for the test data. Unfortunately, despite the state-of-the-art oversampling methods tested, the classification performance was lower for the test set compared to the training set. Next we describe a novel method of informative undersampling with IR.

Informative Undersampling with IR: We present a novel informative undersampling method, based on a criterion of information rate (IUS + IR). As mentioned in Sect. Variable Markov Oracle, the symbolization step with the VMO selects the symbolic representation of a signal that has the highest IR value, thus capturing a higher amount of information carried in the signal. If the optimal IR value is very low, this means that the optimal symbolization of a signal carries little information, in which case it is discarded. We apply a global thresholding approach to the samples belonging to the majority class, such that all the samples having an IR value beneath a threshold are deleted. In this case, the mean of all IR values of the samples in the majority class is used as a threshold. This results in an undersampled majority class that consists of the samples that have a high optimal IR value, which means their symbolization captures a considerable amount of information that will be subsequently quantified by our symbolic $RQA_{\mathscr{S}}$ measures. In Sect. Results: COVID-19 Classification Performance, we show that this novel undersampling method improves remarkably the overall classification performance metrics.

Feature Set

Our feature set consists of symbolic $RQA_{\mathscr{S}}$ extracted from the VMO-symbolized MFCC vectors. We used the standard MFCC feature extraction algorithm and 13 MFFC coefficients for each frame per signal are considered. Using this feature set we are able to analyse the temporal evolution of the features by computing $RP_{\mathscr{S}}$ from overlapping windows of fixed size for long time series.

In our previous work [48], we showed that $RQA_{\mathscr{S}}$ features achieve fairly high scores in identifying emotion in musical sounds as well as in auditory scene. In the present work, we show that $RQA_{\mathscr{S}}$ metrics extracted from symbolized MFCC vectors constitute a robust feature set for the identification of COVID-19 in two types of vocal sounds: coughs and sustained vowels [42].

Classification

Considering our imbalanced dataset, we experiment with different classification algorithms as well as various sampling

techniques. Classifiers included decision trees, support vector machines, K-nearest-neighbor, random forest (RF) and XGBoost. We report the results of the best performing classifier, the weighted XGBoost classifier.

XGBoost: XGBoost (eXtreme Gradient Boosting) is an advanced implementation of the stochastic gradient boosting algorithm. It is a powerful and sophisticated method that deals with data irregularities and helps reduce overfitting. One of the main advantages of XGBoost is that it is effective on datasets that have severe imbalance in class distribution. The weighted XGBoost for class imbalance provides a hyperparameter, the **scale-pos-weight**, designed to tune the behavior of the algorithm for imbalanced classification problems. The weighted XGBoost estimates this value by the total number of examples in the majority class divided by the total number of examples in the minority class:

$$scale - pos - weight = \frac{\sum(Negative\,Examples)}{\sum(Positive\,Examples)}. \tag{17}$$

Experimental work: In a preprocessing stage, the time series is transformed into a vector of 13 MFCCs per frame per sound, at 22050 Hz sampling rate using an orthonormal discrete cosine transform (DCT). In stage 2, the MFCC feature vector is input to the VMO construction algorithm, that generates several symbolizations of the features in terms of their recurrence properties. Next by means of information rate (IR), the optimal threshold θ is evaluated to obtain the optimal VMO symbolization model $\mathcal{M}_S$. Finally, the symbolic $RP_{\mathcal{S}}$ is generated from the self-similarity matrix created from the longest repeated substrings (LRS) of $\mathcal{M}_S$, and a recurrence analysis of the $RP_{\mathcal{S}}$ infers the $RQA_{\mathcal{S}}$ estimates (Fig. 1).

The dataset is normalized prior to training so that column features are scaled to have standard deviation 1, and centered to have mean 0. A binary classification task is conducted where the positive class is positive diagnosis of the disease and the negative class is negative diagnosis (healthy). A 5 times repeated 20-fold stratified cross-validation was performed to evaluate the generalization of the model.

Classification tasks: The following classification tasks are implemented with weighted XGBoost, using recordings of coughs and vowels. First, a baseline reporting the performance on the imbalanced dataset. A second task reports the significantly improved results with undersampling of the majority class with RUS + IR. Finally, we optimize the performance using a threshold-moving method.

Performance Evaluation

The most frequently used metrics are based on the confusion matrix. However, these metrics do not provide accurate information on a classifier's functionality and are ineffective

when there is data imbalance [26]. Here we use metrics that account for dataset imbalance, and that are not particularly sensitive to data distributions.

Precision is sensitive to changes in data, while recall is not. Recall does not give an insight about how many examples are incorrectly labelled as positive. Precision cannot assert how many positive examples are incorrectly labelled. The F_1-measure is the weighted harmonic mean of precision and recall that tends towards the lowest of the two, and by means of a single value, it provides a better insight into the functionality of a classifier [26, 40, 44]. Cohen's Kappa (κ) takes chance into account and overcomes the problem of overestimating the accuracy, it measures the extent to which the agreement between observed and predicted values is higher than that expected by chance alone [11, 12, 40]. Cohen's kappa ranges from -1 (total disagreement), through 0 (random classification) to 1 (perfect agreement) [22, 23]. Particularly, [36] considers 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81-1 as almost perfect. Finally, the area under the receiver operating characteristic curve (ROC) [21, 26, 40] provides a visual representation of the relative trade-offs between recall (TPR) and the false-positives rate (FPR), and measures the ability of a classifier for various threshold settings [26].

For a binary classification problem, the classification performance is typically measured by the geometric mean (G-Mean) of the true-positive and the true-negative rates [35]. G-Mean is a measure for imbalanced classification that can be optimized to achieve a balance between sensitivity and specificity. A common approach is to test the learning model with each threshold returned from the ROC curve, and select the threshold with the maximal G-Mean value:

$$G - Mean = \sqrt{Sensitivity * Specificity}. \tag{18}$$

Evaluation of our model is done using G-Mean, mean accuracy (mACC) with cross-validation, precision or positive predictive value (PPV), sensitivity or true-positive rate (TPR), $F1$-measure, area under the ROC, as well as Cohen's Kappa (κ).

Results: COVID-19 Classification Performance

The performance results are reported in Tables 2, 3 and 4 for coughs and sustained vowel.

Table 2 shows that the imbalanced data negatively affect the baseline performance results, for both types of vocal sounds. Table 3 shows the results of the classification after undersampling with IR. It is obvious that the results significantly improve after our novel undersampling method. However, both precision and recall have low values. A key

Table 2 Baseline performance with XGBoost on imbalanced data

Datasets	ACC	PPV	TPR	F_1	κ	AUC
Coughs	0.98	0.12	0.11	0.10	0.10	0.46
Vowel	0.98	0.08	0.07	0.07	0.07	0.47

Table 3 Best achieved performance with XGBoost and IUS+IR

Datasets	ACC	PPV	TPR	F_1	κ	AUC
Coughs	0.97	0.78	0.65	0.62	0.61	0.84
Vowel	0.99	0.69	0.7	0.69	0.69	0.86

Table 4 Optimal threshold metrics with XGBoost

Datasets	Threshold	G-Mean	PPV	TPR	F_1
Coughs	0.265	0.97	0.83	1.0	0.91
Vowel	0.76	0.99	0.80	1.0	0.89

point concerning imbalanced classification problems is that the accuracy metric is not a good measure for assessing model performance since the positive class is greatly outnumbered by the negative class. In scenarios such as ours, it is important to accurately identify the positive cases, which in our case refer to the COVID-19 diagnosed patients. Intuitively, one tends to maximize recall in our scenario to minimize unknowingly infectious individuals from spreading it further. However increasing recall will suffer from low precision. Conversely, as we increase precision recall suffers. Hence, we want to maximize both recall and precision. In such a case, a common practice is to optimize the performance results by means of threshold-moving, such that we find the optimal threshold for which the F_1-score is maximized. We search for the threshold that maximizes the geometric mean (G-Mean). We obtain a precision of 83%, a recall of 100%, and an F_1 of 91% for cough sounds, and a precision of 80%, a recall of 100% and an F_1 measure of 89% for vowel sounds (Table 4).

Discussion and Future Work

This work displays high potential for evaluating and automatically detecting COVID-19 from web-based audio samples of an individual's coughs and vocalisations. Using symbolic nonlinear recurrence dynamics, we showed the robustness of our method in detecting COVID-19 in cough sounds as well as in sustained vowel 'ah', with an overall mean accuracy of 97% and 99%, respectively, and a mean F_1 measure of 91% and 89%, respectively, obtained after threshold-moving. The steps taken to obtain the features included the preprocessing of the audio signal using a Variable Markov Oracle method that detects the best

recurrence structure by searching over all possible symbolized sequences and selects the threshold that achieves the highest predictive information. Then computing a recurrence plot from the symbolic dynamics. Followed by extracting a set of recurrence quantification features from the recurrence plot that are used as input to an XGBoost classifier.

The unique aspect of our approach is that it considers a variety of statistical complexity measures that are known to be powerful characteristics of non-linear dynamics of chaotic physical systems, and apply this to the analysis of a time series comprising of sequences of cepstral audio features. The combination of mel-frequency cepstrum preprocessing with long-term repetition structure detection method and extraction of multiple complexity measures from repetition statistics provide a good representation of sounds that is capable of distinguishing between sick and healthy individuals. Furthermore, we presented a novel method of informative undersampling using an information theoretic measure of information rate, and applied it to undersample the majority class of healthy sounds in our dataset. We showed that the overall classification performance was significantly improved after informative undersampling with IR, compared to the baseline.

We intend to expand our analysis to other speech signals such as sustained vowels 'oh' and 'eh', as well as spontaneous speech signals. As is commonly the case, a larger data-set and cleaner audio samples may lead to enhanced performance. Additional work needs to be done to consider the sensitivity and understand how many COVID-19 sick diagnosis our model may miss as this is critical to evaluate the risk of having misdiagnosed individuals spread the disease further.

While this model is not intended to be used as an independent diagnosis method and should be used alongside other standard tests, it offers a non-invasive immediate diagnosis method that can reach millions of people. This can help address the shortage of tests worldwide and can be easily accessible to anyone with an internet-connected computer or smartphone. We recommend the model be used as a tool to guide whether further tests at a hospital

or testing center are recommended. This can have a very valuable contribution in helping individuals take knowledgeable steps to protect themselves and others in their community.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abeyratne UR, Swarnkar V, Setyati A, Triasih R. Cough sound analysis can rapidly diagnose childhood pneumonia. Ann Biomed Eng. 2013;41(11):2448–62.
2. Al-khassaweneh M, Abdelrahman Ra'ed Bani. A signal processing approach for the diagnosis of asthma from cough sounds. J Med Eng Technol. 2013;37(3):165–71.
3. Cyril A, Maxime C, Mathieu R Factor oracle: A new structure for pattern matching. In: *International conference on current trends in theory and practice of computer science*, pages 295–310. Springer, (1999).
4. Alonso Jesús B, María Fernando Díaz-de, Travieso Carlos M, Ferrer Miguel Angel. Using nonlinear features for voice disorder detection. In *ISCA tutorial and research workshop (ITRW) on nonlinear speech processing*, (2005).
5. Humberto A, Ayari, F, Hortensia GG Recurrence analysis of cardiac restitution in human ventricle. In *Recurrence plots and their quantifications: expanding horizons*, pages 169–183. Springer, (2016).
6. Assayag Gérard, Dubnov Shlomo. Using factor oracles for machine improvisation. Soft Comput. 2004;8(9):604–10.
7. Botha Tania, Ryffel Bernhard. Reactivation of latent tuberculosis infection. J Immunol. 2003;171:3110–8.
8. Boyanov Boyan, Hadjitodorov Stefan. Acoustic analysis of pathological voices. a voice analysis system for the screening of laryngeal diseases. IEEE Eng Med Biol Mag. 1997;16(4):74–82.
9. Bradley Elizabeth, Kantz Holger. Nonlinear time-series analysis revisited. Chaos: an interdisciplinary. J Nonlinear Sci. 2015;25(9):097610.
10. Chatrzarrin Hanieh, Arcelus Amaya, Goubran Rafik, Knoefel Frank. Feature extraction for the differentiation of dry and wet cough sounds. In: 2011 IEEE international symposium on medical measurements and applications, pages 162–166. IEEE, (2011).
11. Cohen Jacob. A coefficient of agreement for nominal scales. Educ Psychol Measurement. 1960;20(1):37–46.
12. Cohen Jacob. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull. 1968;70(4):213.
13. Cordón I, García S, Fernández A, Herrera F. Imbalance: oversampling algorithms for imbalanced classification in r. Knowl-Based Syst. 2018;161:329–41.
14. Gauri D, Björn S An overview on audio, signal, speech, & language processing for covid-19. *arXiv preprint*arXiv:2005.08579, 2020.
15. Dibazar Alireza A, Berger Theodore W, Narayanan Shrikanth S. Pathological voice assessment. In *2006 international conference of the IEEE engineering in medicine and biology society*, pages 1669–1673. IEEE, 2006.
16. Donner R, Hinrichs U, Scholz-Reiter B. Symbolic recurrence plots: a new quantitative framework for performance analysis of manufacturing networks. Euro Phys J Special Topics. 2008;164(1):85–104.
17. Dubnov Shlomo. Spectral anticipations. Comput Music J. 2006;30(2):63–83.
18. Dubnov S, Assayag G, Cont A Audio oracle: a new algorithm for fast learning of audio structures; 2007.
19. Dubnov Shlomo, Assayag Gérard. and Arshia Cont. Audio oracle analysis of musical information rate. In: *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on*, pages 567–571. IEEE, 2011.
20. Faure Philippe, Lesne Annick. Recurrence plots for symbolic sequences. Int J Bifurcation Chaos. 2010;20(06):1731–49.
21. Fawcett T. An introduction to roc analysis. Pattern Recognition Lett. 2006;27(8):861–74.
22. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Measurement. 1973;33(3):613–9.
23. García S, Fernández A, Luengo J, Herrera F. A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Comput. 2009;13(10):959.
24. Gibiat V, Castellengo M. Period doubling occurrences in wind instruments musical performance. Acta Acustica United with Acustica. 2000;86(4):746–54.
25. Gómez-Vilda P, Fernández-Baillo R, Rodellar-Biarge V, Lluis Víctor Nieto, Álvarez-Marquina A, Mazaira-Fernández LM, Martínez-Olalla R, Godino-Llorente JI. Glottal source biometrical signature for voice pathology detection. Speech Commun. 2009;51(9):759–81.
26. He Haibo, Garcia Edwardo A. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21(9):1263–84.
27. Henríquez P, Alonso JB, Ferrer MA, Travieso CM, Godino-Llorente JI, Díaz-de María Fernando. Characterization of healthy and pathological voice through measures based on nonlinear dynamics. IEEE Trans Audio Speech Language Process. 2009;17(6):1186–95.
28. Henríquez Patricia, Alonso Jesús B, Ferrer Miguel A, Travieso Carlos M, Orozco-Arroyave Juan R. Application of nonlinear dynamics characterization to emotional speech. In: *International conference on nonlinear speech processing*, pages 127–136. Springer, 2011.
29. Patricia H, Alonso JB, Ferrer MA, Travieso CM, Orozco-Arroyave JR. Nonlinear dynamics characterization of emotional speech. Neurocomputing. 2014;132:126–35.
30. Herzel Hanspeter. Bifurcations and chaos in voice signals. Appl Mech Rev. 1993;46(7):399–413.
31. Herzel Hanspeter, Berry David, Titze Ingo R, Saleh Marwa. Analysis of vocal disorders with methods from nonlinear dynamics. J Speech Lang Hearing Res. 1994;37(5):1008–19.
32. Ali Imran, Iryna Posokhova, Haneya N Qureshi, Usama Masood, Sajid Riaz, Kamran Ali, Charles N John, and Muhammad Nabeel. Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app. *arXiv preprint* arXiv:2004.01275, 2020.
33. Iwanski JS, Bradleya E. Recurrence plots of experimental data: To embed or not to embed? Chaos. 1998;8(4):861.
34. Kantz Holger, Schreiber Thomas. Nonlinear time series analysis, vol. 7. Cambridge University Press; 2004.
35. Kuncheva Ludmila I, Arnaiz-González Álvar, Díez-Pastor José-Francisco, Gunn Iain AD. Instance selection improves geometric mean accuracy: a study on imbalanced data classification. Progress Artif Intell. 2019;8(2):215–28.

36. Landis J Richard, Koch, Gary G. The measurement of observer agreement for categorical data. *biometrics*, 1977; pages 159–174.

37. Min LC, Serdar Y, Murtaza B, Abe K, Carlos B, Zhigang D, Sungbok L, Shrikanth N Emotion recognition based on phoneme classes. In *Interspeech*, 2004; pages 205–211.

38. Letellier C. Estimating the shannon entropy: recurrence plots versus symbolic dynamics. Phys Rev Lett. 2006;96(25):254102.

39. Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online. 2007;6(1):23.

40. Liu C, White M, Newell G. Measuring the accuracy of species distribution models: a review. In *Proceedings 18th world IMACs/ MODSIM congress. Cairns, Australia*, 2009; pages 4241–4247.

41. Liu Jia-Ming, You Mingyu, Wang Zheng, Li Guo-Zheng, Xianghuai Xu, Qiu Zhongmin. Cough event classification by pretrained deep neural network. BMC Med Inform Decision Making. 2015;15(S4):S2.

42. Lombardi Angela, Guccione Pietro, Guaragnella Cataldo. Exploring recurrence properties of vowels for analysis of emotions in speech. Sens Transducers. 2016;204(9):45.

43. Maganza Christian, Caussé René, Laloë Franck. Bifurcations, period doublings and chaos in clarinetlike systems. EPL (Europhysics Letters). 1986;1(6):295.

44. Maratea Antonio, Petrosino Alfredo, Manzo Mario. Adjusted f-measure and kernel scaling for imbalanced data learning. Inform Sci. 2014;257:331–41.

45. Marwan Norbert, Romano M Carmen, Thiel Marco, Kurths Jürgen. Recurrence plots for the analysis of complex systems. Phys Rep. 2007;438(5):237–329.

46. Marwan Norbert, Wessel Niels, Meyerfeldt Udo, Schirdewan Alexander, Kurths Jürgen. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. Phys Rev E Phys Rev E. 2002;66:026702.

47. Mende Werner, Herzel Hanspeter, Wermke Kathleen. Bifurcations and chaos in newborn infant cries. Phys Lett A. 1990;145(8–9):418–24.

48. Mouawad Pauline, Dubnov Shlomo. On modeling affect in audio with non-linear symbolic dynamics. Adv Sci Technol Eng Sys J. 2017;2(3):1727–40.

49. Muhammad Ghulam, Alsulaiman Mansour, Mahmood Awais, Ali Zulfiqar. Automatic voice disorder classification using vowel formants. In: 2011 IEEE international conference on multimedia and expo, 2011; pages 1–6. IEEE.

50. Mukherjee Himadri, Obaidullah Sk Md, Santosh KC, Phadikar Santanu, Roy Kaushik. A lazy learning-based language identification from speech using mfcc-2 features. Int J Mach Learn Cybernet. 2020;11(1):1–14.

51. Venkata Srikanth N, Strik H Deep sensing of breathing signal during conversational speech. 2019.

52. Vijay P , Jamieson Donald G. Acoustic discrimination of pathological voice. Journal of Speech, Language, and Hearing Research, 2001.

53. Peerzade GN, Deshmukh RR, Waghmare SD. A review Speech emotion recognition. Int J Comput Sci Eng. 2018;6:400–2.

54. Patricia Henríquez Rodríguez, Jesús B Alonso Hernández, Miguel A Ferrer Ballester, Carlos M Travieso González, and Juan R Orozco-Arroyave. Global selection of features for nonlinear dynamics characterization of emotional speech. *Cognitive Computation*, 5(4):517–525, 2013.

55. Rolink J, Kutz M, Fonseca P, Long X, Misgeld B, Leonhardt S. Recurrence quantification analysis across sleep stages. Biomed Signal Process Control. 2015;20:107–16.

56. Roma G, Nogueira W , Herrera P, de Boronat R Recurrence quantification analysis features for auditory scene classification. *IEEE AASP challenge: detection and classification of acoustic scenes and events, Tech. Rep*, 2013.

57. Rui R, Bao Changchun. Musical instrument classification based on nonlinear recurrence analysis and supervised learning. Radioengineering. 2013;22(1):61.

58. Santos MS, Soares JP, Abreu PH, Araujo H, Santos J. Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches [research frontier]. IEEE Comput Intell Mag. 2018;13(4):59–76.

59. KC Santosh. Speech processing in healthcare: Can we integrate? In *Intelligent Speech Signal Processing*, pages 1–4. Elsevier, 2019.

60. Schreiber T. Interdisciplinary application of nonlinear time series methods. Phys Rep. 1999;308(1):1–64.

61. Schultz D, Spiegel S, Marwan Norbert, Albayrak S. Approximation of diagonal line based measures in recurrence quantification analysis. Phys Lett A. 2015;379(14):997–1011.

62. Subburaj S, Parvez L, Rajagopalan TG. Methods of recording and analysing cough sounds. Pulmonary Pharmacol. 1996;9(5–6):269–79.

63. Vinayak Swarnkar, Udantha R Abeyratne, Anne B Chang, Yusuf A Amrulloh, Amalia Setyati, and Rina Triasih. Automatic identification of wet and dry cough in pediatric patients with respiratory diseases. *Annals of biomedical engineering*, 41(5):1016–1028, 2013.

64. Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

65. Sreenivas Sremath Tirumala, Seyed Reza Shahamiri, Abhimanyu Singh Garhwal, and Ruili Wang. Speaker identification features extraction methods: A systematic review. *Expert Systems with Applications*, 90:250–271, 2017.

66. Ingo R Titze. *Principles of voice production*. National Center for Voice and Speech, 2000.

67. Keiichi Tokuda, Takao Kobayashi, Takashi Masuko, and Satoshi Imai. Mel-generalized cepstral analysis-a unified approach to speech spectral estimation. In *Third International Conference on Spoken Language Processing*, 1994.

68. Carlos M Travieso, Jesús B Alonso, Juan Rafael Orozco-Arroyave, Jesús Francisco Vargas-Bonilla, Elmar Nöth, and Antonio G Ravelo-García. Detection of different voice diseases based on the nonlinear characterization of speech signals. *Expert Systems with Applications*, 82:184–195, 2017.

69. Trulla LL, Giuliani A, Zbilut JP, Webber CL. Recurrence quantification analysis of the logistic equation with transients. Phys. Lett. A. 1996;223(4):255–60.

70. Athanasios Tsanas, Max A Little, Patrick E McSharry, Jennifer Spielman, and Lorraine O Ramig. Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 59(5):1264–1271, 2012.

71. Cheng-i Wang and Shlomo Dubnov. Guided music synthesis with variable markov oracle. In *The 3rd international workshop on musical metacreation, 10th artificial intelligence and interactive digital entertainment conference*, 2014.

72. Cheng-i Wang and Shlomo Dubnov. Pattern discovery from audio recordings by variable markov oracle: A music information dynamics approach. In *2015 IEEE International conference on acoustics, speech and signal processing (ICASSP)*, pages 683–687. IEEE, 2015.

73. Wang C-i, Dubnov S. The variable markov oracle: algorithms for human gesture applications. IEEE MultiMedia. 2015;22(4):52–67.

74. C de A Washington, FM de Assis, BG Aguiar Neto, Silvana C Costa, and Vinıcius JD Vieira. Pathological voice classification based on recurrence quantification measures. 2012.

75. Charles L Webber, Norbert Marwan, Angelo Facchini, and Alessandro Giuliani. Simpler methods do it better: success of recurrence quantification analysis as a general purpose data analysis tool. Phys Lett A, 373(41):3753–3756, 2009.