



Vocal acoustic analysis and machine learning for the identification of schizophrenia

Caroline Wanderley Espinola^{1,2} · Juliana Carneiro Gomes³ · Jessiane Mônica Silva Pereira³ · Wellington Pinheiro dos Santos¹ 

Received: 16 May 2020 / Accepted: 22 September 2020 / Published online: 29 September 2020
© Sociedade Brasileira de Engenharia Biomedica 2020

Abstract

Purpose Psychiatry still needs objective biomarkers. In the context of schizophrenia, there are speech abnormalities such as tangentiality, derailment, alogia, neologisms, poverty of speech, and aprosodia. There is a growing interest in speech signals features as possible indicators of schizophrenia. This article aims to develop an intelligent tool for detection of schizophrenia using vocal patterns and machine learning techniques. The main advantages of this type of solution are the low cost, high performance, and for being non-invasive.

Methods Thirty-one individuals over 18 years old were selected, 20 with previous diagnosis of schizophrenia, and 11 healthy controls. Their speech was audio-recorded in naturalistic settings, during a routine medical assessment for psychiatric patients. In the case of healthy patients, the recordings were made in different environments. Recordings were pre-processed, excluding non-participant voices. We extracted 33 features. We used the particle swarm optimization algorithm for feature selection.

Results The classifiers' performance was analyzed with four metrics: accuracy, sensibility, specificity, and kappa index. Best results were achieved when considering all 33 extracted features. Within machine models, support vector machines (SVM) models provided the greatest classification performance, with mean accuracy of 91.76% for PUK kernel. Our results outperform those from most studies published so far for the detection of schizophrenia based on acoustic patterns.

Conclusion The use of machine learning classifiers using vocal parameters, in particular SVM, has shown to be very promising for the detection of schizophrenia. Nevertheless, further experiments with a larger sample will be necessary to validate our findings.

Keywords Schizophrenia · Diagnosis · Voice · Acoustic parameters · Machine learning · Support vector machines

Introduction

Current clinical practice in psychiatry depends on diagnostic criteria built entirely on expert consensus, instead of relying on objective biomarkers (Bzdok and Meyer-lindenberg 2018). Such criteria, described in the Diagnostic and Statistical Manual, 5th Edition (DSM-5), and in the International Classification of Diseases (ICD-10), are still considered the gold-standard for diagnosis in psychiatry (American Psychiatric Association 2013). Nevertheless, those diagnostic systems have been criticized due to their absence of clinical predictability and neurological validity (Bzdok and Meyer-lindenberg 2018), and their poor diagnostic stability (Bacagarcia et al. 2007). This ultimately leads to trial-and error treatment (Petzschner et al. 2017). While other medical fields hold markers of disease presence and severity, such as tumor

✉ Wellington Pinheiro dos Santos
wellington.santos@ufpe.br

Caroline Wanderley Espinola
caroline.espinola@ufpe.br

Jessiane Mônica Silva Pereira
jmep@comp.poli.br; jcg@comp.poli.br

¹ Departamento de Engenharia Biomédica, Universidade Federal de Pernambuco, Recife, Brazil

² Serviço de Emergências Psiquiátricas, Hospital Ulysses Pernambucano, Recife, Brazil

³ Núcleo de Engenharia da Computação, Escola Politécnica da Universidade de Pernambuco, Recife, Brazil

volume measurement and biochemical blood tests, psychiatry still lacks routine objective tests (Bedi et al. 2015; Mundt et al. 2012).

Assessment and treatment in psychiatry are historically based on reports from patients and on clinical evaluation (Mundt et al. 2007). This makes diagnosis and therapeutic decision extremely sensitive to memory and subjectivity biases (Jiang et al. 2018). In this context, there was an intense search for biomarkers for diagnosis and follow-up of psychiatric patients in the last decade (Iwabuchi et al. 2013; Mundt et al. 2012). However, most of them are expensive and invasive (Higuchi et al. 2018). Therefore, despite all efforts, objective measures for assessment of mental disorders are still unknown (Mundt et al. 2007).

Other major challenges psychiatry faces are that nosology and clinical practice do not benefit from advances in neurosciences. These difficulties can be tackled by computational psychiatry, which applies machine learning (ML) with focus on clinical applications and single-subject treatments (Bzdok and Meyer-lindenberg 2018; Petzschner et al. 2017). Machine learning has successful implementations in problem-solving tasks in several medical fields, like supportive diagnostic tools based on neuroanatomical structures for Alzheimer's disease (dos Santos et al. 2009; W. P. dos Santos et al. 2007), breast cancer (Cruz et al. 2018; de Lima et al. 2016; de Santana et al. 2018), and multiple sclerosis diagnosis (Commowick et al. 2018).

Schizophrenia is a group of severe psychotic disorders with heterogeneous etiologies, clinical presentations and responses to treatment (Sadock et al. 2017). It is characterized by hallucinations, delusions, thought and behavior disorder or catatonia, and “negative symptoms,” such as diminished emotional expression and avolition (American Psychiatric Association 2013). Since the first descriptions of this disorder, speech/language deficits have been described as remarkable features of schizophrenia, and are often associated with core negative symptoms and social impairment (Alberto et al. 2019). These symptoms comprise poverty of speech, disorganized speech, derailment, tangentiality, neologism, incoherence, mutism, perseveration, echolalia, thought blocking (Mac-Kay et al. 2018) inappropriate affect prosody or aprosodia (Chakraborty et al. 2018a; Covington et al. 2012; Elite et al. 2014). Also known as flattened speech intonation, aprosodia consists of diminished vocal emphasis (Alpert and Anderson 1977); reduced inflection and fluency (Alpert et al. 2000); and prosody comprehension deficits, such as difficulties in recognizing intonation patterns (Elite et al. 2014) Overall, these speech abnormalities result from disruptions in cognitive processes and contribute to the frequent communication deficits in schizophrenia (Mac-Kay et al. 2018).

In this framework, computational psychiatry has shown to be a promising method to deal with the complexity of psychiatric diagnosis, translating neuroscientific advances to clinical

applications. Its data-driven approach applies machine learning techniques to high-dimensional data in order to improve classification diagnosis, treatment selection and even treatment outcomes (Huys et al. 2016). The use of ML models is appropriate for individual-level predictions, which would provide personalized therapeutic decisions in the future (Bzdok and Meyer-lindenberg 2018). Moreover, it may also enable mobile monitoring of patients and telemedicine applications that are accessible for clinical use (Cohen et al. 2016). In the context of speech-language deficits, vocal acoustic analyses using ML classifiers appear to be a promising venue for understanding their role within mental disorders (Cohen et al. 2012).

Thinking about this, this work proposes the application of ML techniques in audio-recordings to perform binary classification. For this, we collected data from 31 patients, divided into 2 groups: group of patients diagnosed with schizophrenia, and a control group, composed of healthy patients. In this context, we pre-processed all recordings in order to minimize environment noises. After that, we extracted 33 features from each 10 s-window of the signals. Finally, multiple classifiers were tested. Our goal is to provide an intelligent tool that performs accurate and non-invasive schizophrenia diagnosis with low computational cost.

This paper is organized as follows: Section 2 describes studies related to the characterization of schizophrenia based on vocal parameters. In Section 3, an instrument for the detection of schizophrenia is introduced and implemented. Results are presented and discussed in Sections 4 and 5, respectively. Section 6 states our conclusions with suggestions for future studies on this subject.

Related works

As speech-language abnormalities are a hallmark in schizophrenia, several related studies have been published, most of which on natural language processing and semantics/syntax (Bedi et al. 2015; Chakraborty et al. 2018a; Elvevåg et al. 2010; Kayi et al. 2017; Tovar et al. 2019), and a limited number of studies about vocal patterns in schizophrenia (Tahir et al. 2019).

Patients with schizophrenia tend to show slowed speech, reduced pitch variability, significantly increased number of pauses, and decreased variability in syllable timing than healthy individuals. These characteristics were observed in a semi-automatic analysis of vocal pitch or fundamental frequency (F0) during an emotionally neutral reading task performed by Martínez-sánchez et al. (2015). In a sample of 80 subjects, they reported a discrimination accuracy of 93.8% between schizophrenic patients and controls using signal processing algorithms. They also observed remarkable intergroup

differences, with patients exhibiting slowed speech, low volume, and many pauses.

Likewise, Rapcan et al. (2010) compared vocal pitch, temporal, and energy parameters of 39 schizophrenic patients and 18 healthy controls during an emotionally neutral reading task. Their results demonstrated significant differences between groups, with patients showing decreased mean utterance duration, and increased values in number of pauses, proportion of silence, mean pause duration, total length of pauses, and relative variation in energy. On the other hand, no statistical significance was reported for total length of utterances and relative variation in vocal pitch. However, the lack of educational level matching between groups with reading task may represent an important limitation to their findings, because different educational status may translate into different reading speed and fluency between patient and control samples.

Vocal acoustic analysis is also capable of measuring the severity of negative symptoms such as aprosodia. Compton et al. (2018) analyzed audio recordings of schizophrenic patients with aprosodia, schizophrenic patients without aprosodia, and healthy controls, and compared variability in pitch (F0), first (F1) and second (F2) formants, and intensity/loudness. Their results showed significant differences among groups, with the group with aprosodia showing reduced variability in pitch, F2, and intensity/loudness than other groups.

Similarly, Covington et al. (2012) analyzed F0, F1, and F2 of 25 video-recorded interviews. They investigated tongue movement as an indicator of the severity of negative symptoms in first-episode schizophrenia-spectrum patients. Their study concluded that F2, a measure of variability of tongue anterior or posterior position, was significantly correlated with the severity of negative symptoms.

Chakraborty et al. (2018b) employed low-level speech signals (or low-level descriptors, LLD) alone or in combination with body movements to predict negative symptoms of schizophrenia using automatic classifiers. For that purpose, they applied support vector machines (SVM), a supervised

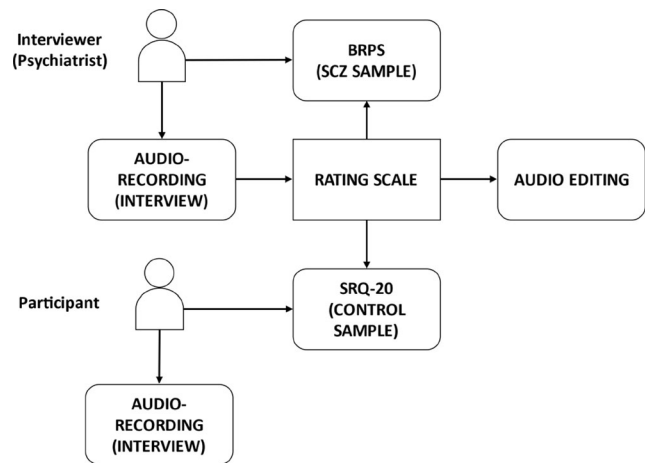


Fig. 1 Block diagram of data acquisition: audio-recordings of the schizophrenia sample were acquired during an interview with a psychiatrist. After that, a trained clinician assessed their symptoms using BPRS. Healthy controls were audio-recorded in different environments. They were also asked to answer SRQ-20 questionnaire. BPRS and SRQ-20 scores were calculated. The SRQ-20 cut-off score of 6/7 was considered, while all diagnosed patients with SCZ were included, regardless of the obtained scores. After participants’ selection, we did the audios editing, aiming to remove voices from other people besides the patients

machine learning technique widely used in classification problems (Russell and Norvig 2016). They reported a classification accuracy of 79.49% using low-level speech signals alone, and of 86.36% for their combination with body movements.

Likewise, Tahir et al. (2019) investigated conversational and prosodic features as objective measures of negative symptoms in schizophrenia. Conversational features relate to duration of speech, speaking turns, interruptions, and interjections, while prosodic features comprise F0, F1, F2, and F3; mel frequency cepstral coefficients (MFCCs); and amplitude (minimum, maximum and mean volume, entropy). The performance of some ML algorithms in discriminating between patients and healthy controls was evaluated in their article: SVM, multilayer perceptron (MLP), random forest (RF), and ensemble (bagging). The best results were reported for

Table 1 Sample characteristics: the 31 participants were divided into two groups: control group composed of healthy patients, and the group of people diagnosed with schizophrenia. In both groups, there is a predominance of males. The average age of the control group is 30 years, while in the second group it is 36 years

Group	Number of participants	Age	Gender (male:female)	Rating scale	Mean scores
Healthy control	11	30.09 years (± 12.58 years)	6:5	SRQ-20	3.00 points (± 1,86)
Schizophrenia	20	36.00 yrs. (± 12.39 years)	12:8	BPRS	44.55 points (± 11.28)
<i>p</i> value	–	<i>p</i> = 0.223 (<i>t</i> ₂₉ = 1.2463)	<i>p</i> = 0.768 χ^2	–	–

Notes: BPRS, Brief Psychiatric Rating Scale; SRQ-20, Self-Reporting Questionnaire

Table 2 Recording duration after audio editing

Group	Number of participants	Total recording duration	Mean recording duration
Control	11 (6 males)	5816 s (96.9 min)	528.7 s (8.8 min) ± 138.7 s
Schizophrenia	20 (12 males)	7541 s (125.7 min)	377.1 s (6.28 min) ± 270.4 s

MLP (accuracy = 81.3%), with speaking rate, frequency, and volume entropy showing significant differences between groups.

In a meta-analysis of 46 papers about acoustic patterns in schizophrenia, Alberto et al. (2019) compared three categories of study design: qualitative ratings, quantitative univariate analyses, and multivariate ML investigations. Machine learning studies provided superior results, with overall out-of-sample accuracy of 76.5–87.5%, and appeared to be more promising. They also identified remarkable differences in acoustic patterns between schizophrenic patients and healthy controls, with the patient group showing decreased proportion of spoken time, reduced speech rate, and increased duration of pauses. These abnormalities were directly related to flat affect and alogia. Additionally, they observed that studies with dialogical and free speech provided the greatest differences between groups, in contrast with studies using constrained monologs.

Methods

In this study, a sample of 31 volunteers over 18 years old was selected and divided into two subsamples:

- Healthy control: 11 healthy participants (6 males) were selected through the Self-Reporting Questionnaire (SRQ-

20), a screening instrument for common mental disorders (Gonçalves et al. 2008; K. O. B. Santos et al. 2010);

- Schizophrenia: 20 patients previously diagnosed with schizophrenia (12 males) were assessed using the Brief Psychiatric Rating Scale (BPRS; Overall and Gorham 1962), one of the most widely used instruments for the evaluation of symptom severity in schizophrenia (Leucht et al. 2005).

All individuals from the schizophrenia sample (mean age = 36.00; SD = 12.39; 54.5% male) fulfilled DSM-5 diagnostic criteria for schizophrenia and were previously diagnosed by an independent psychiatrist. Data for this group were collected at outpatient settings and at inpatient psychiatric units in Hospital das Clínicas, Federal University of Pernambuco, and in Hospital Ulysses Pernambucano, both in Recife, Northeast Brazil. Participants with coexistent neurological disorders or who made professional use of their voices were excluded.

Meanwhile, the control sample (mean age = 30.09; SD = 12.58; 60.0% male) was matched with the patient sample for age, gender and region of origin (Brazilian Northeast). The same exclusion criteria were applied to this group. Participants from both groups were literate, but the control sample had a higher educational level ($p < 0.001$). Unfortunately, it was not possible to match subsamples with reference educational level, as this was a challenging co-variable to match for in this particular population. Although this represents a limitation to our study, a similar approach was made in some previous studies (Cannizzaro et al. 2005; Cohen et al. 2008; Rapcan et al. 2010). Sample characteristics are presented in Table 1.

The use of SRQ-20 was designed to remove participants with current mental illnesses from the control sample. The SRQ-20 cut-off score of 6/7 was considered (Santos et al. 2010), whereas in the schizophrenia sample, participants with prior diagnosis were included, irrespective of their BPRS score. The mean BPRS score of schizophrenic patients in this sample was 44.55 and corresponded to moderate illness severity (Leucht et al. 2005). All participants have given written consent, and this study was conducted only after approval of a local Research Ethical Board.

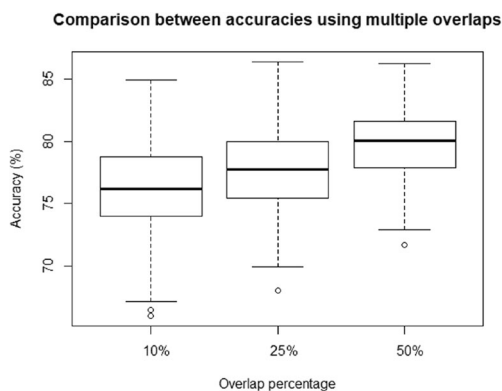


Fig. 2 Boxplots with comparison of performance using different window overlaps: 10%, 25%, and 50%. The boxplot shows that an overlap of 50% achieved greater accuracy values and less dispersion of values

Table 3 Equations of the 33 extracted parameters

Parameter	Equation	Parameter	Equation
Mean (μ)	$\mu = \frac{1}{N} \sum_{n=1}^N x_n$	Waveform length	$WL = \sum_{n=1}^{N-1} x_{n+1} - x_n $
Variance	$var = \frac{1}{N-1} \sum_{n=1}^N (x_n - \mu)^2$	Zero crossing	$ZC = \sum_{n=1}^{N-1} [sgn(x_n \times x_{n+1}) \cap x_n - x_{n+1} \geq \text{threshold}]$ $sgn(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
Standard deviation (σ)	$\sigma = \sqrt{\frac{1}{N-1} \sum_{n=1}^N x_n - \mu ^2}$	Slope Sign Changes	$SSC = \sum_{n=1}^{N-1} [f(x_n - x_{n-1}) \times (x_n - x_{n+1})]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
Root mean square	$RMS = \sqrt{\frac{\sum_{n=1}^N (x_n)^2}{N}}$	Hjorth parameter activity	$Hjorth_{activity} = \frac{1}{N-1} \sum_{n=1}^N (x_n - \mu)^2$
Average amplitude changes	$AAC = \frac{1}{N} \left(\sum_{n=1}^N \left \frac{d x(t)}{dt} \right \right)$	Hjorth parameter mobility	$Hjorth_{mobility} = \sqrt{\frac{var(\frac{d x(t)}{dt})}{var(x(t))}}$
Difference absolute deviation	$DASDV = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{d x(t)}{dt} \right)^2}$	Hjorth parameter complexity	$Hjorth_{complexity} = \frac{Hjorth_{mobility}(\frac{d x(t)}{dt})}{Hjorth_{mobility}(x(t))}$
Integrated absolute value	$IAV = \sum_{n=1}^N x_n$	Mean frequency	$MNF = \frac{\sum_{j=1}^M f_j P_j}{\sum_{j=1}^M P_j}$ Where f_j, P_j are the frequencies and power of the spectrum, respectively, and M is the length of the frequencies
Logarithm detector	$LOGD = e^{(\frac{1}{N} \sum_{n=1}^N \log(x_n))}$	Median frequency	$MDF = \frac{1}{2} \sum_{j=1}^M P_j$
Simple square integral	$SSI = \sum_{n=1}^N x_n^2$	Mean power	$MNP = \sum_{j=1}^M \frac{P_j}{M}$
Mean absolute value	$MAV = \frac{1}{N} \sum_{n=1}^N x_n $	Peak frequency	$PKF = \max(P_j)$
Mean logarithm kernel	$MLOGK = \frac{1}{N} \left \sum_{n=1}^N x_n \right $	Power Spectrum ratio	$PSR = \frac{PKF}{\sum_{j=1}^M P_j}$
Skewness (s)	$s = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^3}{\sigma^3}$	Total power	$TP = \sum_{j=1}^M P_j$
Kurtosis	$kurt = \frac{\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^4}{\sigma^4}$	First spectral moment	$SM1 = \sum_{j=1}^M f_j P_j$
Maximum amplitude	$MAX = \max(x_n)$	Second spectral moment	$SM2 = \sum_{j=1}^M f_j^2 P_j$
Third moment	$M3 = \left \frac{1}{N} \sum_{n=1}^N (x_n)^3 \right $	Third spectral moment	$SM3 = \sum_{j=1}^M f_j^3 P_j$
Fourth moment	$M4 = \left \frac{1}{N} \sum_{n=1}^N (x_n)^4 \right $	Variance of central frequency	$VCF = \frac{SM2}{TP} - \left(\frac{SM1}{TP} \right)^2$
Fifth moment	$M5 = \left \frac{1}{N} \sum_{n=1}^N (x_n)^5 \right $		

Acquisitions of voice samples

A Tascam™ 16-bit linear PCM recorder was used, at 44.1 KHz sampling rate, in WAV format, without file compression. Audio-recordings of the schizophrenia sample were acquired during an interview with a psychiatrist in naturalistic settings, i.e., patients were recorded during a routine medical assessment at outpatient offices or inpatient units. After each

interview, a trained clinician assessed their symptoms using BPRS. Meanwhile, healthy controls were audio-recorded in different environments (e.g., office, classroom, gym). Participants from this sample were asked to answer SRQ-20, as this questionnaire is self-applied. No duration limit was set for the recordings. As conversations were thoroughly recorded, voices from the clinician and possible third parties were also acquired and needed to be further removed. The total

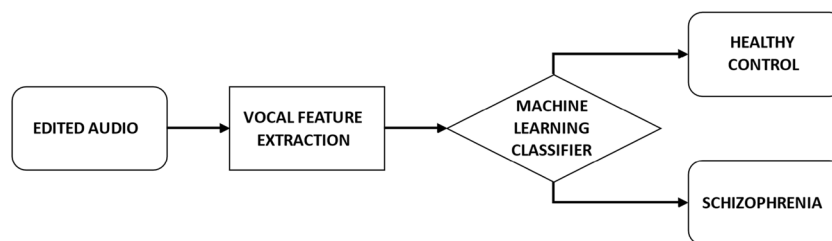


Fig. 3 Block diagram of proposed solution: After editing each of audio-recordings, 33 attributes were extracted from each sample window (10 s with 50% overlap) in Octave environment. Then, and .ARFF file was

generated, and multiple classifiers were tested in the Weka software. The tested classifications were binary, seeking to differentiate healthy patients from patients diagnosed with schizophrenia

duration of the recordings of both samples was 407.3 min (6.79 h). The process of data acquisition is summarized in Fig. 1.

Audio editing

After data collection, voice signals from the interviewer and any potential companion were manually removed using *Audacity* audio software (version 2.3.2). This process yielded 222.6 min of recorded audio from participants (3.71 h) as follows: 96.9 min for the control sample and 125.7 min for the schizophrenia sample. Recording duration of both samples after audio editing is shown in Table 2.

Feature extraction

All recordings were submitted to a vocal feature extraction on GNU Octave™; a free open-source signal-processing software. Rectangular windows, with frame length of 10 s. In order to determine the window overlap percentage, three overlap sizes were tested: 10% (1 s), 25% (2.5 s), and 50% (5 s). For this, the random forest classifier was used. We performed these experiments 30 times, using 10-folds cross validation in Weka environment. Boxplots in Fig. 2 shows the accuracy results for these three scenarios. As shown in the figure, 50% overlap outperforms the others. It reached higher mean accuracy value, as well as less dispersion.

As raw audio data were used, no filtering process was applied. Consequently, background noise was also captured. However, we believe such noise would not be able to interfere significantly, given the homogeneous spectral behavior of the acoustic features selected for extraction. At this stage, the following 33 parameters were extracted: skewness; kurtosis; zero crossing; slope sign changes; variance; standard deviation; mean absolute value; logarithm detector; root mean square; average amplitude change; difference absolute deviation; integrated absolute value; mean logarithm kernel; simple square integral; mean value; third, fourth and fifth moments; maximum amplitude; power spectrum ratio; peak frequency; mean power; mean frequency; median

frequency; total power; variance of central frequency; first, second and third spectral moments; Hjorth parameter activity, mobility and complexity; and waveform length. The corresponding mathematical expressions of these attributes are presented in Table 3.

The choice of the above parameters relies on their accurate representation of input signals to computational models, once decision-making process of machine learning classifiers is not associated with human interpretation. Additionally, attributes from different domains (e.g., temporal and spectral) were selected so as to avoid feature selection biases. Furthermore, such parameters have already been successfully used for representing other biomedical signals, such as electroencephalography. Subsequently, the most relevant parameters were selected using particle swarm optimization (PSO), a feature selection method for dimensionality reduction within classification problems (Xue et al. 2012).

Feature selection using particle swarm optimization

Particle swarm optimization (PSO) algorithms were created by James Kennedy and Russel Eberhart in 1995,

Comparison between classifiers using different number of attributes

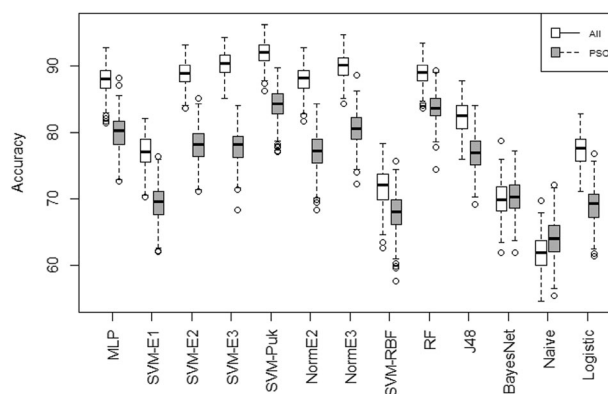


Fig. 4 Accuracy boxplots for comparison of classifiers performance. In most classification experiments, the performance of the classification with all the extracted attributes exceeds the classification performance using the attributes selected with PSO. Considering the boxplots with all the attributes (white color), we can see that the Support vector machines have higher accuracy values

Table 4 List of 12 attributes after selection with particle swarm optimization

PSO selected attributes
Zerocrossings
Hjorth parameter complexity
Average amplitude changes
Mean absolute value
Kurtosis
Third moment
Fourth moment
Maximum amplitude
Peak frequency
Power spectrum ratio
Mean power
Total power

respectively a social psychologist and an electrical engineer (Kennedy and Eberhart 1995). PSOs are based on the behavior and movement of flocks of animals, such as fish and birds, therefore being algorithms based on theories that describe animal social behavior, having elements in common with genetic algorithms and with evolutionary programming (Eberhart and Kennedy 1995; Kennedy and Eberhart 1995; Santos and Assis 2013).

Similar to genetic algorithms, PSO is initialized with a random initial population. However, while in the genetic algorithms, the individuals in this initial population are represented by chromosomes, in the PSO a position vector and a

velocity vector are associated with each individual. In addition, in the PSO there are no mutations or selection of individuals. Thus, at each iteration, only positions and speeds of different individuals are adjusted in the direction of the best global position and the best individual position, according to a certain objective function, according to the following canonical expression (Eberhart and Shi 2011; Chuanwen and Bompard 2005; Van der Merwe and Engelbrecht 2003; Hu et al. 2003; Trelea 2003; Shi and Krohling 2002):

$$\mathbf{x}_i(t + 1) = \mathbf{x}_i(t) + \mathbf{v}_i(t + 1), \tag{1}$$

since

$$\mathbf{v}_i(t + 1) = w\mathbf{v}_i(t) + c_1r_1[\mathbf{p}_i(t) - \mathbf{x}_i(t)] + c_2r_2[\mathbf{p}_g(t) - \mathbf{x}_i(t)], \tag{2}$$

for $1 \leq i \leq m$, where m is the number of particles in the cluster; w is the inertia factor, where $0 < w < 1$; $r_1(t)$ and $r_2(t)$ are numbers randomly uniformly distributed in the interval $[0, 1]$; c_1 and c_2 are constriction constants, also called coefficients of acceleration, so that $c_1 + c_2 = 4$ (typically, $c_1 = 2 + D$ and $c_2 = 2 - D$, where $D \approx 0$), where c_1 is the weight due to consciousness of the particle, individual consciousness or local consciousness, depending on the implementation, while c_2 is the weight due to global awareness; \mathbf{x}_i is the position, while \mathbf{v}_i is the speed of i th particle; \mathbf{p}_g is the best global position, while \mathbf{p}_i is the best individual or local position in relation to the i th particle.

Local and global best positions are considered according to local and global maxima of a determined objective function, whilst the position \mathbf{x}_i defines the i -th solution candidate. In this classification problem, we defined \mathbf{x}_i as a n -dimensional binary vector in which each coordinate is associated to the presence (“1” values”) or absence (“0” values) of the corresponding selected characteristic. Therefore, each solution candidate is associated to training and test sets composed by dimension-reduced feature vectors. As objective function, we used a J48 decision tree returning classification accuracies. The parameters w , c_1 , c_2 , r_1 , and r_2 were all set to 0.33. We used a population of 20 individuals evolving in 500 generations. This solution was implemented in Java using the Java machine learning library Weka (Moraglio et al. 2007; García-Nieto et al. 2009).

Comparison between classifiers using different number of attributes

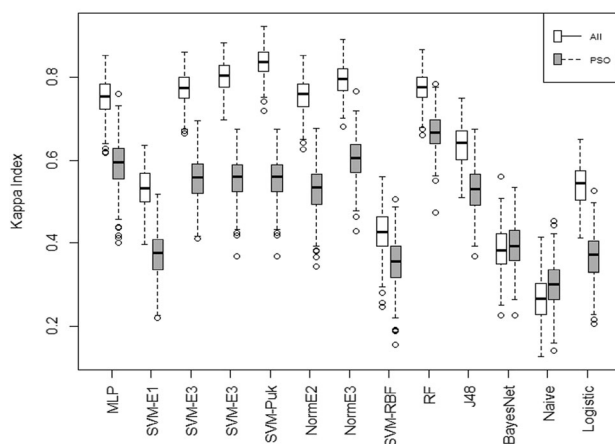


Fig. 5 Kappa index boxplots for comparison of classifiers performance. Similarly to the accuracy values, the results of the Kappa index are higher for cases in which all 33 attributes extracted initially were considered. In addition, SVMs also performed better

Classification

Both databases (with all features extracted and after PSO selection) were balanced through the addition of artificial instances on Weka™ artificial intelligence environment. This is essential to avoid computational biases towards the class with more representativeness, in this case the schizophrenia sample. Edited audio samples were submitted to classification experiments using

Table 5 Classification performances of machine learning models (schizophrenia vs. healthy control). SVM with PUK kernel presented the best results in all four evaluated metrics (accuracy, kappa index, sensitivity, and specificity). It achieved an average accuracy of 91.76%, mean kappa index of 0.8352, sensibility of 91.9%, and specificity of 91.6%

Model	Accuracy	Kappa Index	Sensibility	Specificity
Multilayer perceptron (MLP)	88.22%	0.7644	88.8%	87.6%
Random forest	89.53%	0.7907	90.0%	89.0%
Decision trees	82.13%	0.6427	83.6%	80.7%
Logistic regression	78.47%	0.5694	75.1%	81.9%
SVM linear	77.96%	0.5592	72.4%	83.6%
SVM 2-degree polynomial	89.07%	0.7815	86.7%	91.4%
SVM 3-degree polynomial	89.94%	0.7987	88.3%	91.6%
SVM PUK	91.76%	0.8352	91.9%	91.6%
SVM normalized polynomial	90.37%	0.8073	89.2%	91.5%
SVM RBF	72.93%	0.4585	64.7%	81.2%
Bayes net	69.12%	0.3825	74.4%	63.9%
Naive Bayes	63.97%	0.2794	47.2%	80.7%

Abbreviations: *MLP* multilayer perceptron, *PUK* Pearson universal VII kernel, *RBF*, radial basis function, *SVM* support vector machines

the following ML algorithms on Weka™: multilayer perceptron (MLP), logistic regression, random forest (RF), decision trees, Bayes net, Naïve Bayes, and SVM with different kernels (linear, polynomial kernel, radial basis function or RBF, PUK, and normalized polynomial kernel). Given the relatively small number of subjects in each sample, experiments were performed with 10-fold cross-validation in order to maximize training samples. Figure 3 illustrates the steps of the prediction system.

Results

Initially, computational experiments were performed using classifiers in their default settings. Subsequently, different setups for all algorithms with adjustable settings were tested (MLP; polykernel and normalized polykernel SVM, SVM PUK kernel, and random forest). The best performances for each classifier type are presented in the boxplots of the Figs. 3 and 4 below. Figures 3 and 4 show the accuracy and kappa index values, respectively. They also compare the classifiers' performances using all 33 extracted attributes (white boxplots), and using the

attributes selected by the PSO method (gray color). Using PSO, 12 attributes were selected, which are listed in the Table 4. As can be seen in Figs. 4 and 5, most classifiers have a better performance when considering all attributes. The exception occurs only for classifiers based on Bayes' theory. However, the latter are classifiers with low performance for this problem. Thereby, they were not chosen. Furthermore, Table 5 presents accuracy, kappa index, sensibility, and specificity values for the best classifiers.

The results above demonstrate that classification accuracy for SVM models varied significantly (72.93–91.76%), depending on which kernel was used. SVM PUK kernel achieved mean accuracy of 91.76% (sensibility 91.9%; specificity 91.6%), which was the best performance of all classifiers used in this study. The confusion matrix of this kernel is shown in Table 6. SVM normalized polynomial kernel also achieved accuracy above 90%. The greatest performances of different SVM kernels in this dataset support findings from previous studies, which possibly indicate the superiority of this algorithm for classification tasks using vocal parameters.

Discussion

This paper presents a study on discriminating schizophrenic patients and healthy subjects based on vocal features and machine learning classifiers. The process of data acquisition was designed to provide high translational power, as this is the first study to collect audio-recordings during actual psychiatric interviews. A feature extraction algorithm has been locally developed for the reliable extraction of 33 acoustic features, which have

Table 6 Confusion matrix for the model with the highest performance (SVM PUK): 91.59% instances from the control group were correctly classified, while 91.89% instances from the Schizophrenia group were correctly classified

	Classified as control	Classified as schizophrenia
Control	91.59%	8.41%
Schizophrenia	8.11%	91.89%

Table 7 Comparative analysis of previous studies and this paper

Authors	Sample Size	Task	Features extracted	Findings/comments
Tahir et al. (2019)	54 SCZ (46.3% male); 26 HC (46.2% male)	Social interaction (interview with a psychologist)	Conversational (number of natural turns; speaking percentage; mutual silence percentage; turn duration; natural interjections; speaking interjections; interruptions; failed interruptions; speaking rate and response time) Prosodic (F0; formants F1, F2, and F3; MFCCs; mean volume; max volume; min volume; entropy)	Correlations between non-verbal speech cues and NSA-16 indicators: Turn duration and reduced expressive gestures (NSA 15): -0.466 Turn duration and restricted speech quantity (NSA 2): -0.461 ; Response time and communication (NSA 19): 0.42 Classifiers tested: SVM, RF, MLP, ensemble (bagging). Best performance: MLP without feature selection Classification acc (SCZ vs. HC): 81% (precision: 0.9; recall: 0.82) Comments: Audio acquisitions were made in a research setting. Findings may not be replicable in a real-life environment
Martínez-Sánchez et al. (2015)	45 SCZ (71.1% male); 35 HC (62.9% male)	Constrained production (reading)	Task duration (s); intensity; pause rate ≥ 300 ms (%); mean F0 (Hz); F0 SD (Hz); F0 range (ST); syllabic dynamics (%); prosodic peaks (%); prosodic valleys (%); intra-syllabic trajectory (ST/s); inter-syllabic trajectory (ST/s); phonation trajectory (ST/s).	Parameters with increased values in the SCZ group: Task duration (s): $p = 0.001$; Pause rate ≥ 300 ms (%): $p = 0.001$; Parameters with decreased values in the SCZ group: Intensity (dB): $p = 0.001$; Syllabic dynamics (%): $p = 0.017$; Prosodic valleys (%): $p = 0.011$; Intra-syllabic trajectory (ST/s): $p = 0.001$; Inter-syllabic trajectory (ST/s): $p = 0.014$; Phonation trajectory (ST/s): $p = 0.001$. NOT significantly different: Mean F0; F0 SD; F0 range; prosodic peaks. Correlations: duration of illness and prosodic variables: Number of years elapsed since diagnosis and intra-syllabic trajectories: $r = -0.377$; $p = 0.028$) and phonation ($r = -0.422$; $p = .013$) Time elapsed since the last relapse correlated and phonation trajectory ($r = 0.404$; $p = 0.018$). Comments: Only literate patients may perform the task, which reduces its applicability as a screening test. An average illness duration of 21 years. in the SCZ group possibly indicates a high prevalence of patients with residual symptoms of SCZ. Therefore, classification performance in an acute or first-episode SCZ sample is unknown.
Chakraborty et al. (2018a)	50 SCZ (50% male)	Semi-structured interview	988 acoustic features from OpenSMILE “emobase” set, including 26 LLDs and their delta regression coefficients (intensity; loudness; F0; 12 MFCCs; ZCR; F0 envelope; probability of voicing; 8 LSF).	Predictions of clinician-assigned NSA-16 ratings of negative symptoms: Prolonged time to respond: acc. = 82%; Restricted speech quantity: acc. = 78%; Impoverished speech content: acc. = 64%;

Table 7 (continued)

Authors	Sample Size	Task	Features extracted	Findings/comments
Chakraborty et al. (2018b)	52 SCZ (48.1% male); 26 HC (46.2% male)	Semi-structured interview	Statistical measures (min and max values; positions of max and min values; range; arithmetic mean; SD; 2 linear regression coefficients and linear and quadratic error; skewness; kurtosis; quartiles 1, 2 and 3; interquartile ranges 1–2, 2–3 and 1–3). 14 conversational features (e.g., number of natural turns, interjections, mutual silence). 78 Linguistic features processed with LIWC 2015.	Affect–reduced modulation of intensity: acc = 72%; Reduced expressive gestures: acc. = 72%. <u>Comments:</u> This study could have benefitted from the inclusion of a control sample for comparison of acoustic features Details about the severity of negative symptoms within SCZ sample are not provided. A symptom rating scale could have been used to ensure diagnosis. Several algorithms and feature selection methods were tested for the prediction of subjective NSA-16 ratings. <u>Best performances:</u> Prolonged time to respond: acc = 82.69% (kNN with <i>F</i> -score); Restricted speech quantity: acc. = 82.69% (Gaussian SVM with PCA); Impoverished speech content: acc. = 80.77% (linear SVM with linear SVM); Emotion–reduced range: acc. = 61.54% (kNN with RF); Affect–reduced modulation of intensity: acc. = 78.85% (adaboosted decision trees with decision trees); Reduced expressive gestures: acc. = 84.62% (adaboosted decision trees with χ^2). <u>SCZ vs. HC classification:</u> Acoustic features: acc. = 79.49% (linear SVM with PCA); Acoustic features and body movement signals: acc. = 86.36% (linear SVM with F score). <u>Comments:</u> Details about the severity of negative symptoms within SCZ sample are not provided. A symptom rating scale could have been used to ascertain diagnosis in the patient sample.
Compton et al. (2018)	25 SCZ patients with aprosody (68.0% male); 29 SCZ patients without aprosody (79.3% male); 102 HC (55.4% male)	Task 1: spontaneous speech (picture description); Tasks 2 and 3: spontaneous speech (answers to open-ended questions); Tasks 4 and 5: constrained production (excerpt reading).	F0 variability (SD); intensity/loudness variability (SD); F1 variability (SD); F2 variability (SD)	SANS and CAINS research-rated scales were used to derive subsamples of patients with and without aprosody. <u>Findings:</u> F0 SD: decreased values in SCZ with aprosody than HC in task 5 after adjusting tests for sociodemographics ($p < 0.05$), F1 variability: NS in adjusted tests. F2 variability: decreased values in SCZ with aprosody than HC in tasks 3, 4 and 5 ($p < 0.05$); <u>Intensity/loudness variability:</u> decreased values in SCZ with aprosody than SCZ without aprosody and HC across all tasks both in unadjusted and in adjusted tests ($p < 0.05$).

Table 7 (continued)

Authors	Sample Size	Task	Features extracted	Findings/comments
Covington et al. (2012)	25 first-episode psychosis patients (76% male)	Semi-structured interview	Mean log (F0); F0 variability: SD _N of log (F0) F1 variability: SD _N (F1) F2 variability: SD _N (F2)	<p>Comments: Subsamples were not matched for gender ($p = 0.05$) or educational level ($p < 0.01$). Due to the inclusion of reading tasks, differences in educational level may limit interpretations of findings.</p> <p>Correlations with PANSS negative symptom ratings: Mean log (F0): NS ($r = -0.071$; $p = 0.73$); SD_N of log (F0): NS ($r = -0.107$; $p = 0.61$); SD_N (F1): NS ($r = -0.339$, $p = 0.098$); SD_N (F2): Negative symptom severity: ($r = -0.446$; $p = 0.03$); Blunted affect ($r = -0.398$, $p = 0.05$); Emotional withdrawal ($r = -0.423$, $p = 0.04$); Poor rapport ($r = -0.404$, $p = 0.05$); Lack of spontaneity and flow of conversation ($r = -0.523$, $p = 0.007$).</p> <p>Comments: Small sample size: 25 min of recordings were processed, because only a one-minute sample from each patient was considered.</p> <p><u>Comparison SCZ vs. HC for feature values:</u> <u>Increased in SCZ:</u> Number of pauses ($p < 0.0003$); Mean pause duration ($p < 0.01$); Proportion of silence ($p < 0.00001$); Total recording time ($p < 0.008$); Total length of pauses ($p < 0.000007$); Relative variation in energy ($p < 0.04$).</p> <p><u>Reduced in SCZ:</u> Mean utterance duration ($p < 0.04$). <u>NOT statistically significant:</u> <u>Total length of utterances</u> ($p < 0.07$); Relative variation in vocal pitch ($p < 0.4$). <u>Correlations between features and symptom scales:</u> <u>BPRS:</u> Proportion of silence ($r = -0.34$; $p = 0.05$); Mean utterance duration ($r = 0.46$; $p = 0.01$); Total Length of Utterances ($r = 0.35$; $p = 0.05$); Relative Variation in Energy ($r = 0.37$; $p = 0.05$). <u>SANS:</u> Proportion of silence ($r = -0.35$; $p = 0.05$); Mean utterance duration ($r = 0.37$; $p = 0.05$); Total recording time ($r = 0.38$; $p = 0.05$); Total length of utterances ($r = 0.46$; $p = 0.01$);</p>
Rapcan et al. (2010)	39 SCZ (69.2% male); 18 HC (44.4% male)	Constrained production (reading)	Temporal features (number of pauses; mean pause duration; proportion of silence; mean utterance duration; total recording time; total length of pauses; total length of utterances). Energy measures (relative variation in energy). Vocal pitch estimation (relative variation in vocal pitch)	

Table 7 (continued)

Authors	Sample Size	Task	Features extracted	Findings/comments
This article	20 SCZ (60.0% male) 11 HC (54.5% male)	Social interaction (clinical interview)	33 features (mentioned in Section 3)	<p>Relative variation in energy ($r = 0.37$; $p = 0.05$). Classification acc. SCZ vs. HC: 79.42% (LDA); feature subset: all features minus total length of utterances and relative variation in vocal pitch).</p> <p>Comments: Samples were not matched for gender and educational level. The use of reading task without matching samples for educational level may hamper interpretations of results.</p> <p>SCZ vs. HC classification: Best performance: SVM PUK acc. = 91.77% (sens.: 91.9%; spec. 91.6%); Patients were recorded during a real psychiatric interview; small sample size. Samples were not matched for educational level.</p>

Note: acc. accuracy, *BPRS* Brief Psychiatric Rating Scale, *CAINS* Clinical Assessment Interview for Negative Symptoms, *HC* healthy controls, *LDA* linear discriminant analysis, *LWC* Linguistic Inquiry and Word Count, *LLD* low-level descriptors, *LSF* line spectral frequencies, *MFFC*: Mel frequency cepstral coefficients, *NSA-16* Negative Symptom Assessment, *PAANS* Positive and Negative Syndrome Scale, *PCA* principal component analysis, *PUK* Pearson Universal VII Kernel, *RF* random forest, *SCZ* schizophrenia, *ST* semitones, *SVM* support vector machines

successfully been used for modeling classification problems in neurology and psychiatry. Some machine learning models tested in this paper have achieved high performances; in particular, SVM with PUK kernel yielded high classification accuracy both for schizophrenic patients and healthy controls. With the exception of Martínez-sánchez et al. (2015), our results outperformed those from similar studies using vocal parameters for the detection of schizophrenia.

Nevertheless, although promising, findings reported in this article should be considered preliminary due to limitations in study design. For instance, the small sample size and not controlling for possible confounding factors, such as smoking history and use of medications, may limit statistical analyses. Additionally, an important caveat is the difference in educational level between samples, given the fact that educational background is related to speech fluency. In future studies, we aim to address these limitations and perform the same experiments on a larger number of subjects. In Table 7 below a comparative analysis between some of the studies mentioned in this article and this study is presented.

Conclusion and future works

Current psychiatric diagnosis still lacks objective biomarkers and relies mostly on specialist opinion based on diagnostic systems. Nevertheless, these criteria have been criticized due to their lack of correlation with the neurobiology and etiopathogenesis of mental disorders, leading to trial-and-error treatments. In this context, patients with schizophrenia may present with vocal acoustic abnormalities that may be used as objective parameters for the identification and assessment of this disorder.

Therefore, this paper aimed at the development of objective measures of schizophrenia to aid clinical practice in the future. For this purpose, we extracted vocal acoustic features and performed experiments using different automated classification techniques based on machine learning. Some of the most widely used machine learning classifiers were tested in this work. Our results demonstrate the viability of an inexpensive and non-invasive tool for the detection of schizophrenia based on vocal acoustic analysis through machine learning algorithms. In future studies, we intend to perform the same experiments in a larger sample, and also with gender-based datasets. We would like to evaluate if schizophrenia affects vocal acoustic properties from men and women in a different fashion, and if so, how these differences influence the performance of automated classifiers.

Acknowledgments We are grateful to the Brazilian research-funding agency CNPq, for the partial support of this research.

Compliance with ethical standards

Conflict of interest Authors do not have any conflicts of interest to declare.

References

- Alberto P, Amdis S, Vibeke B, Riccardo F. Voice patterns in schizophrenia: a systematic review and Bayesian Meta-analysis. *Voice Schizophrenia Rev Meta-anal.* 2019;1–40.
- Alpert M, Anderson LT. Imagery mediation of vocal emphasis in flat affect. *Arch Gen Psychiatry.* 1977;34(2):208–12.
- Alpert M, Rosenberg SD, Pouget ER, Shaw RJ. Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Res.* 2000;97:107–18.
- American Psychiatric Association. (2013). *DSM-5 - Manual Diagnóstico e Estatístico de Transtornos Mentais. Artmed (5.)*. Porto Alegre: Artmed. 1011769780890425596.
- Baca-Garcia E, Perez-Rodriguez MM, Basurte-Villamor I, Fernandez Del Moral AL, Jimenez-Arriero MA, Gonzalez De Rivera JL, et al. Diagnostic stability of psychiatric disorders in clinical practice. *Br J Psychiatry.* 2007;190(MAR):210–6. <https://doi.org/10.1192/bjp.bp.106.024026>.
- Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *Nature Partner Journals.* 2015;1:15030. <https://doi.org/10.1038/npjschz.2015.30>.
- Bzdok D, Meyer-lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biologic Psychiat Cognit Neurosci Neuroimag.* 2018;3:223–30. <https://doi.org/10.1016/j.bpsc.2017.11.007>.
- Cannizzaro MS, Cohen H, Rappard F, Snyder PJ. Bradyphrenia and Bradykinesia both contribute to altered speech in schizophrenia: a quantitative acoustic study. *Cogn Behav Neurol.* 2005;18(4):206–10. <https://doi.org/10.1097/01.wnn.0000185278.21352.e5>.
- Chakraborty D, Xu S, Yang Z, Han Y, Chua V, Tahir Y, et al. Prediction of negative symptoms of schizophrenia from objective linguistic, acoustic and non-verbal conversational cues. In: IEEE 2018 international conference on Cyberworlds prediction; 2018a. p. 280–3. <https://doi.org/10.1109/CW.2018.00057>.
- Chakraborty, D, Yang, Z, Tahir, Y, Maszczyk, T, Dauwels, J, Thalmann, N, ... Lee, J (2018b). Prediction of Negative Symptoms of Schizophrenia From Emotion Related Low-Level Speech Signals. *IEEE*, 6024–6028.
- Chuanwen J, Bompard E. A hybrid method of chaotic particle swarm optimization and linear interior for reactive power optimisation. *Math Comput Simul.* 2005;68(1):57–65.
- Cohen AS, Alpert M, Nienow TM, Dinzeo TJ, Docherty NM. Computerized measurement of negative symptoms in schizophrenia. *J Psychiatr Res.* 2008;42:827–36. <https://doi.org/10.1016/j.jpsychires.2007.08.008>.
- Cohen AS, Mitchell KR, Docherty NM, Horan WP. Vocal expression in schizophrenia: less than meets the ear. *J Abnorm Psychol.* 2016;125(2):299–309. <https://doi.org/10.1037/abn0000136>.
- Cohen AS, Najolia GM, Kim Y, Dinzeo TJ. On the boundaries of blunt affect/alogia across severe mental illness: implications for research domain criteria. *Schizophr Res.* 2012;140(1–3):41–5. <https://doi.org/10.1016/j.schres.2012.07.001>.
- Commowick O, Istace A, Kain M, Laurent B, Leray F, Simon M, et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Sci Rep.* 2018;8(1):1–17.
- Compton MT, Lunden A, Cleary SD, Pauselli L, Alolayan Y, Halpern B, et al. The aprosody of schizophrenia: computationally derived acoustic phonetic underpinnings of monotone speech. In: *Schizophrenia Research*; 2018. p. 1–8. <https://doi.org/10.1016/j.schres.2018.01.007>.
- Covington MA, Lunden SLA, Cristofaro SL, Wan CR, Bailey CT, Broussard B, et al. Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. *Schizophr Res.* 2012;142:93–5.
- Cruz T, Cruz T, Santos W. Detection and classification of lesions in mammographies using neural networks and morphological wavelets. *IEEE Lat Am Trans.* 2018;16(3):926–32.
- de Lima SM, da Silva-Filho AG, dos Santos WP. Detection and classification of masses in mammographic images in a multi-kernel approach. *Comput Methods Prog Biomed.* 2016;134:11–29.
- de Santana MA, Pereira JMS, da Silva FL, de Lima NM, de Sousa FN, de Arruda GMS, et al. Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Res Biomed Eng.* 2018;34(1):45–53.
- dos Santos WP, De Assis FM, De Souza RE, Mendes PB, De Souza Monteiro HS, Alves HD. A dialectical method to classify Alzheimer's magnetic resonance images. *Evol Comput.* 2009;473.
- dos Santos, WP, de Souza, RE, & dos Santos Filho, PB (2007). Evaluation of Alzheimer's disease by analysis of MR images using multilayer perceptrons and Kohonen SOM classifiers as an alternative to the ADC maps. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2118–2121).
- Eberhart, R, & Kennedy, J (1995). A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science* (pp. 39–43). IEEE.
- Eberhart RC, Shi Y. *Computational intelligence: concepts to implementations*. Amsterdam: Elsevier; 2011.
- Elite A, Pedrão LJ, Zamberlan-Amorim NE, Carvalho AMP, Bárbaro AM. Comportamento comunicativo de indivíduos com esquizofrenia. *Rev CEFAC.* 2014;16(4):1283–93.
- Elvevåg B, Foltz PW, Rosenstein M, DeLisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *J Neurolinguistics.* 2010;23(3):270–84. <https://doi.org/10.1161/CIRCULATIONAHA.110.956839>.
- García-Nieto J, Alba E, Jourdan L, Talbi E. Sensitivity and specificity based multiobjective approach for feature selection: application to cancer diagnosis. *Inf Process Lett.* 2009;109(16):887–96.
- Gonçalves DM, Stein AT, Kapczinski F. Avaliação de desempenho do Self-Reporting Questionnaire como instrumento de rastreamento psiquiátrico: Um estudo comparativo com o Structured Clinical Interview for DSM-IV-TR. *Cad Saude Publica.* 2008;24(2):380–90. <https://doi.org/10.1590/S0102-311X2008000200017>.
- Higuchi M, Tokuno S, Nakamura M, Shinohara S. Classification of bipolar disorder, major depressive disorder, and healthy state using voice. *Asian J Pharm Clin Res.* 2018;11(3):89–93. <https://doi.org/10.22159/ajpcr.2018.v11s3.30042>.
- Hu, X, Eberhart, RC, & Shi, Y (2003). Engineering optimization with particle swarm. In *Proceedings of the 2003 IEEE Swarm Intelligence Symposium*. SIS'03 (cat. No. 03EX706) (pp. 53–57). IEEE.
- Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci.* 2016;19(3):404–13. <https://doi.org/10.1038/nn.4238>.
- Iwabuchi SJ, Liddle PF, Palaniyappan L. Clinical utility of machine-learning approaches in schizophrenia: improving diagnostic

- confidence for translational neuroimaging. *Front Psych.* 2013;4(August):1–9. <https://doi.org/10.3389/fpsy.2013.00095>.
- Jiang H, Hu B, Liu Z, Wang G, Zhang L, Li X, et al. Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features. *Comput Math Methods Med.* 2018;2018:6508319. <https://doi.org/10.1155/2018/6508319>.
- Kayi, ES, Diab, M, Pauselli, L, Compton, M, & Coppersmith, G (2017). Predictive linguistic features of schizophrenia. *Proceedings Of the 6th Joint Conference on Lexical and Computational Semantics*, 241–250.
- Kennedy, J, & Eberhart, R (1995). Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE.
- Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel R. Clinical implications of Brief psychiatric rating scale scores. *Br J Psychiatry.* 2005;187(2):366–71. <https://doi.org/10.1016/j.physbeh.2017.03.040>.
- Mac-Kay A, Jerez I, Pesenti P. Speech-language intervention in schizophrenia: an integrative review. *Rev CEFAC.* 2018;20(2):238–46. <https://doi.org/10.1590/1982-0216201820219317>.
- Martínez-Sánchez F, Muela-Martínez JA, Cortés-soto P, José J, Meilán G, Antonio J, et al. Can the acoustic analysis of expressive prosody discriminate schizophrenia? *Span J Psychol.* 2015;18(86):1–9. <https://doi.org/10.1017/sjp.2015.85>.
- Moraglio A, Di Chio C, Poli R. Geometric particle swarm optimisation. In: *European conference on genetic programming*. Berlin, Heidelberg: Springer; 2007. p. 125–36.
- Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geraltz DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *J Neurolinguistics.* 2007;20:50–64. <https://doi.org/10.1016/j.jneuroling.2006.04.001>.
- Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. *Biol Psychiatry.* 2012;72(7):580–7. <https://doi.org/10.1016/j.biopsych.2012.03.015.Vocal>.
- Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. *Psychol Rep.* 1962;10:799–812.
- Petzschner FH, Weber LAE, Gard T, Stephan KE. Review computational psychosomatics and computational psychiatry : toward a joint framework for differential diagnosis. *Biol Psychiatry.* 2017;82:1–10. <https://doi.org/10.1016/j.biopsych.2017.05.012>.
- Rapcan V, D'Arcy S, Yeap S, Afzal N, Thakore J, Reilly RB. Acoustic and temporal analysis of speech: a potential biomarker for schizophrenia. *Med Eng Phys.* 2010;32:1074–9. <https://doi.org/10.1016/j.medengphy.2010.07.013>.
- Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach* (third). Harlow: Pearson Education; 2016.
- Sadock B, Sadock V, Ruiz P. *Compêndio de Psiquiatria: Ciência do Comportamento e Psiquiatria Clínica* (11.). Porto Alegre: Artmed; 2017.
- Santos WP, Assis FM. *Algoritmos dialéticos para inteligência computacional*. Recife: Editora Universitária UFPE; 2013.
- Santos KOB, Araújo TM, Pinho PS, Silva ACC. Avaliação de um Instrumento de Mensuração de Morbidade Psíquica. *Revista Baiana de Saúde Pública.* 2010;34(3):544–60.
- Shi, Y, & Krohling, RA (2002). Co-evolutionary particle swarm optimization to solve min-max problems. In *Proceedings of the 2002 Congress on Evolutionary Computation*. CEC'02 (cat. No. 02TH8600) (Vol. 2, pp. 1682-1687). IEEE.
- Tahir Y, Yang Z, Id DC, Thalmann N, Thalmann D, Maniam Y, et al. Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. *PLoS One.* 2019;14:1–17. <https://doi.org/10.1371/journal.pone.0214314>.
- Tovar A, Fuentes-Claramonte P, Soler-Vidal J, Ramiro-Sousa N, Rodríguez-Martínez A, Sarri-Closa C, et al. The linguistic signature of hallucinated voice talk in schizophrenia. *Schizophr Res.* 2019;206:111–7.
- Trelea IC. The particle swarm optimization algorithm: convergence analysis and parameter selection. *Inf Process Lett.* 2003;85(6):317–25.
- Van der Merwe, DW, & Engelbrecht, AP (2003). Data clustering using particle swarm optimization. In *The 2003 Congress on Evolutionary Computation*, 2003. CEC'03. (Vol. 1, pp. 215-220). IEEE.
- Xue B, Zhang M, Member S, Browne WN. Particle swarm optimization for feature selection in classification: a multi-objective approach. In: *Ieee Transactions on Cybernetics*; 2012. p. 1–16.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.