



Detection of major depressive disorder using vocal acoustic analysis and machine learning—an exploratory study

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

Received: 15 April 2020 / Accepted: 28 September 2020 / Published online: 12 October 2020
© Sociedade Brasileira de Engenharia Biomedica 2020

Abstract

Purpose Diagnosis and treatment in psychiatry are still highly dependent on reports from patients and on clinician judgment. This fact makes them prone to memory and subjectivity biases. As for other medical fields, where objective biomarkers are available, there has been an increasing interest in the development of such tools in psychiatry. To this end, vocal acoustic parameters have been recently studied as possible objective biomarkers, instead of otherwise invasive and costly methods. Patients suffering from different mental disorders, such as major depressive disorder (MDD), may present with alterations of speech. These can be described as uninteresting, monotonous, and spiritless speech and low voice.

Methods Thirty-three individuals (11 males) over 18 years old were selected, 22 of which being previously diagnosed with MDD and 11 healthy controls. Their speech was recorded in naturalistic settings, during a routine medical evaluation for psychiatric patients, and in different environments for healthy controls. Voices from third parties were removed. The recordings were submitted to a vocal feature extraction algorithm, and to different machine learning classification techniques.

Results The results showed that random tree models with 100 trees provided the greatest classification performances. It achieved mean accuracy of $87.5575\% \pm 1.9490$, mean kappa index, sensitivity, and specificity of 0.7508 ± 0.0319 , 0.9149 ± 0.0204 , and 0.8354 ± 0.0254 , respectively, for the detection of MDD.

Conclusion The use of machine learning classifiers with vocal acoustic features appears to be very promising for the detection of major depressive disorder in this exploratory study, but further experiments with a larger sample will be necessary to validate our findings.

Keywords Major depressive disorder · Diagnosis · Voice · Acoustic parameters · Machine learning · Support vector machines

Introduction

Clinical assessment and treatment in psychiatry currently depend 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

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

Caroline Wanderley Espinola
caroline.espinola@ufpe.br

Juliana Carneiro Gomes
jcg@ecom.poli.br

Jessiane Mônica Silva Pereira
jmsp@ecom.poli.br

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

² Serviço de Emergência Psiquiátrica, Hospital Ulysses Pernambucano, Recife, PE, Brazil

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

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 (Baca-Garcia et al. 2007). While other medical fields hold markers of disease presence and severity, such as tumor volume measurement and biochemical blood tests, psychiatry still lacks routine objective tests (Bedi et al. 2015; Mundt et al. 2012).

Historically, evaluation and treatment in psychiatry are based on reports from patients and on clinical evaluation (Mundt et al. 2007). Thus, diagnosis and therapeutic decision are extremely sensitive to memory and subjectivity biases (Jiang et al. 2018). Considering this, over the last decades, there has been an intense search for biomarkers for diagnosis and follow-up of psychiatric patients (Iwabuchi et al., 2013; Mundt et al. 2012), most of those being expensive and invasive (Higuchi et al. 2018). Despite all efforts, instruments for assessment of mental disorders still remain a conundrum (Mundt et al. 2007).

Major depressive disorder (depression) is the most common mental disorder, affecting more than 300 million people worldwide (Sadock et al. 2017; World Health Organization 2018). It is also a leading cause of disability and economic burden (Mundt et al. 2007, 2012). The global prevalence of depression was estimated to be 4.4% in 2015 (World Health Organization 2017), with more women affected than men in a 2:1 ratio (Weinberger et al. 2017). In Brazil, the prevalence of depression is currently the fifth largest in the world, 5.8% (World Health Organization 2017), while its lifetime prevalence can be as high as 16.8% (Miguel et al. 2011).

Patients suffering from depression may present with low mood, irritability, anhedonia, fatigue, psychomotor retardation, cognitive impairment (difficulty in decision-making, poor concentration), and disturbances of somatic functions (insomnia or hypersomnia, appetite disorders, changes in body weight). These symptoms are associated with intense suffering and decline in functioning and may ultimately lead to suicide (American Psychiatric Association 2013). Depression is associated with approximately half of all suicides globally (Cummins et al. 2015).

Early depressive symptoms such as psychomotor retardation and cognitive impairment are frequently related to disturbances in speech (Hashim et al. 2016). Actually, patterns within depressed speech have been documented years ago (Mundt et al. 2007). In particular, the persistent altered emotional state in depression may affect vocal acoustic properties. As a result, depressive speech has been described by clinicians as monotonous, uninteresting, and without energy. These differences could provide the detection of depression through analysis of vocal acoustics of depressed patients (Jiang et al. 2018).

Machine learning is an intensive field of research, with successful applications to solve several problems in health sciences, like breast cancer diagnosis (Cordeiro et al. 2012;

Cordeiro et al. 2016; de Lima et al. 2016; Rodrigues et al. 2019), Alzheimer's disease diagnosis support based on neuroanatomical features (dos Santos et al. 2009; dos Santos et al. 2008; dos Santos et al. 2007), multiple sclerosis diagnosis (Commowick et al. 2018), and many applied neuroscience solutions (da Silva Junior et al. 2019; de Freitas et al. 2019).

Qualitative changes in speech from people suffering with depression have been reported decades ago (Darby and Hollien 1977), e.g., reduction in pitch range (Vanello et al. 2012), increased number of pauses (Mundt et al. 2012), slower speech (Faurholt-Jepsen et al. 2016), and reduced intensity or loudness (Hönig et al. 2014).

For the recognition of changes in mood state, prosodic, phonetic, and spectral aspects of voice are relevant, in particular fundamental frequency (F0) or pitch, intensity, rhythm, speed, jitter, shimmer, energy distribution between formants, and cepstral features. Among those features, jitter is considered important for mood state recognition due to its ability to identify rapid temporary changes in voice (Maxhuni et al. 2016). Alteration of Mel frequency cepstral coefficients (MFCC) are also found in depressed individuals. MFCCs consist of parametrical representation of the speech signal (Hasan et al. 2004) and have been extensively studied as possible features for the detection of major depressive disorder (Cummins et al. 2014; Jiang et al. 2018).

In a sample of 57 depressed patients, Cohn et al. (2009) analyzed prosodic and facial expression elements using two machine learning classifiers: support vector machines (SVM) and logistic regression. Their accuracy for the identification of depression was 79–88% for facial expressions and 79% for prosodic features.

Due to its good generalization power, SVM is considered a state-of-the-art classifier (Alghowinem et al. 2013a), showing great performance in the identification of speech pathologies (Arjmandi and Pooyan 2012; Sellam and Jagadeesan 2014; Wang et al. 2011); along with GMM, SVM is the most widely used classification technique using voice parameters (Jiang et al. 2017). The greatest performances from different SVM kernels in this dataset support findings from previous studies in the literature. For instance, SVM RBF kernel was successfully used for the detection of MDD and PTSD using voice quality parameters (Scherer et al. 2013); it has also provided better performance than other classifiers (MLP, HFS) in handling raw data and has shown strong discriminative power using features like intensity and root mean square (Alghowinem et al. 2013b). Another study has reported the superiority of linear models (SVM linear kernel and LR) for the detection of depressive and anxiety disorders in early childhood using high-quality data vocal features (McGinnis et al. 2019).

In a study with adolescents, Ooi, Lech, and Brian Allen (2013) used glottal, prosodic, and spectral features and Teager energy operator for the prediction of early symptoms

of depression in that age group and reported accuracy of 73% (sensitivity: 79%; specificity: 67%). In another study with a larger sample of adolescents, Low, Maddage, Lech, Sheeber, and Allen (2011) utilized the above attributes with the addition of cepstral parameters and submitted to SVM and Gaussian mixture model (GMM) classifiers. They described significant differences in classifier performances for detecting depression based on gender: 81–87% for males and 72–79% for females.

With emphasis on vocal features for the identification of depression, Hönig et al. (2014) used automatic feature selection to study 34 features: spectral (MFCCs, formants F1 to F4), prosodic (pitch, energy, duration, rhythm), and vocal quality or phonetic features (jitter, shimmer, raw jitter, raw shimmer, logarithm harmonics-to-noise ratio, spectral harmonicity, and spectral tilt). In agreement with findings from Low et al. (2011), they reported a slightly higher correlation in males ($r = 0.39$ males vs. 0.36 females). This suggests that clinical depression can possibly lead to more significant changes in vocal features in men than in women. Similarly, Jiang et al. (2017) also noticed gender differences in classifier performances, with superior results in males. In a sample of 170 subjects, they investigated the discriminative power of three classifiers for the detection of depression: SVM, GMM, and k-nearest neighbors (kNN). SVM achieved the best results in that study, with accuracy of 80.30% (sens.: 75%; spec.: 85.29%) for males and 75.96% (sens.: 77.36%; spec.: 74.51%) for females.

Adversely, Higuchi et al. (2018) analyzed pitch (F0), spectral centroid, and five attributes of MFCC using polytomous logistic regression for the classification of depression, bipolar disorder, and healthy controls. They did not find any difference between genders. Moreover, they also reported an overall accuracy of 90.79%, the highest among the studies revised for this work. This discrepancy may be due to the fact that some features, such as voice quality parameters, could be more gender-independent than other vocal features like F0 (Scherer et al., 2013). In addition to that, several previous studies did not perform gender-based classification experiments (Alpert et al. 2001; Cannizzaro et al. 2004; Liu et al. 2015; Mundt et al. 2007, 2012; Ooi et al. 2013). The same approach is adopted in this study.

Another attribute of study design that might influence the performance of an automated instrument is the type of speech task. Spontaneous speech, e.g., social interactions or interviews, tend to yield higher classification performances than reading tasks. This finding suggests that spontaneous speech provides more acoustic variability, improving the recognition of depression (Alghowinem et al. 2013a; Jiang et al. 2017). Moreover, it is likely that depressed individuals can suppress their emotional state during reading tasks, because of the unimportant nature of the read content or their concentration on reading or even both (Mitra and Shriberg 2015).

This work is organized as follows: “Methods” briefly discusses studies related to the identification of depression using voice. “Methods” describes in detail the implementation of a proposed voice-based instrument for the detection of depression. In “Results,” we present and discuss our results. “Discussion” states our conclusion and suggestions for future work on this subject.

Methods

For this exploratory study, 33 volunteers over 18 years old from both genders were selected and separated into one of the following groups:

- Control group: 11 healthy participants (5 females) were selected through the Self-Reporting Questionnaire (SRQ-20) screening for common mental disorders (Gonçalves et al. 2008; K. O. B. Santos et al., 2010).
- Depression group: 22 patients with previous diagnosis of major depressive disorder (17 females), in conformity with Hamilton Depression Rating Scale—HAM-D 17 (Hamilton 1960).

All individuals from the depression group fulfilled DSM-5 diagnostic criteria for major depressive disorder and were diagnosed by an independent professional prior to this study. Data for this group was collected in outpatient settings and psychiatric wards in the Hospital das Clínicas, Federal University of Pernambuco, and in the Hospital Ulysses Pernambucano, both in Recife, Northeast Brazil. Participants with coexistent neurological disorders or who made professional use of their voices were excluded. The use of validated psychometric scales aimed to verify previous diagnostic consistency and assess clinical severity. The mean age of the control group was 30.1 years (± 12.6 years), whereas the mean age of the depression group was 42.9 years (± 13.0 years). There is no standardization on age controlling in studies: some selected age-matched controls to their samples (Alghowinem et al. 2013b; Alghowinem et al. 2012; Cummins et al. 2015); and some did not (Afshan et al. 2018; Cannizzaro et al. 2004; Higuchi et al. 2018; Jiang et al. 2017; Joshi et al. 2013; Liu et al. 2015; Ozdas et al., 2004; Scherer et al. 2013). Given this heterogeneity, in this work, we assume the perspective of the majority of revised studies in which age between groups was not controlled. All participants have given written consent, and this study was conducted only after approval of a local Research Ethical Board. Table 1 provides a summary of mean age and scale scores for both groups.

For the control group, the SRQ-20 cutoff score was 6/7 (K. O. B. Santos et al. 2010), and for the depression group, the eligibility criterion was HAM-D score above 7. Consequently, patients suffering from mild to severe depression were

Table 1 Mean age and rating scale scores

| Group | Age | Rating scale | Mean scores |
|------------|--------------------------------|--------------|----------------------------|
| Control | 30.1 years (\pm 12.6 years) | SRQ-20 | 3.00 points (\pm 1.86) |
| Depression | 42.9 years (\pm 13 years) | HAM-D 17 | 19.32 points (\pm 7.36) |

HAM-D 17 Hamilton Depression Rating Scale, *SRQ-20* Self-Reporting Questionnaire

selected, as this study aimed to encompass different diagnostic scenarios within depression. The average HAM-D 17 score of the depression group 19.32 corresponds to moderate illness (Zimmerman et al. 2013). By including depressed patients irrespective of their disease severity, we believe to have created a database that represents a real-world clinical dataset. To the best of our knowledge, the same approach was made in the works of Afshan et al. (2018), Alghowinem et al. (2012), Alpert et al. (2001), Hönig et al. (2014), Jiang et al. (2018), and Ooi et al. (2013).

Acquisitions of voice samples

We used a Tascam™ 16-bit linear PCM recorder at 44.1 KHz sampling rate, in WAV format, without compression. Audio acquisitions were made during an interview with a psychiatrist in naturalistic settings, i.e., patients from the depression group were recorded during a routine medical evaluation in an outpatient office or hospital ward. More specifically, 17 patients were recorded in an outpatient office, four patients were recorded in a psychiatry ward, and one in an internal medicine ward. After each interview, a clinician utilized HAM-D 17 scale to verify diagnostic suitability and assess clinical severity. Participants from control group were asked to answer SRQ-20 in order to verify their eligibility. Recordings for this group were made in

different environments, as follows: six in offices or classrooms, 3 in gyms, and two in their residences. Despite being made in different facilities, recordings from both groups shared similar environmental conditions, such as closed rooms and little background noise interference. 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. Six from 22 depressed patients had companions during the recording process, one from which did not speak; four made punctual remarks about medications in use; only one companion had a significant interference in the clinical evaluation, whose corresponding excerpt was entirely discarded. Having said that, we believe the presence of possible companions did not hinder the process of data acquisition. The total time recorded was 425.1 min (7.09 h). Figure 1 summarizes the process of data acquisition.

Audio editing

After voice acquisition, we used Audacity™ audio software to remove voice signals from the interviewer and any potential companion. The edition process was manually made and yielded 271 min of voice signals from participants (4.52 h) as follows: 96.9 min for the control group and 174.1 min for

Fig. 1 Block diagram of voice acquisition

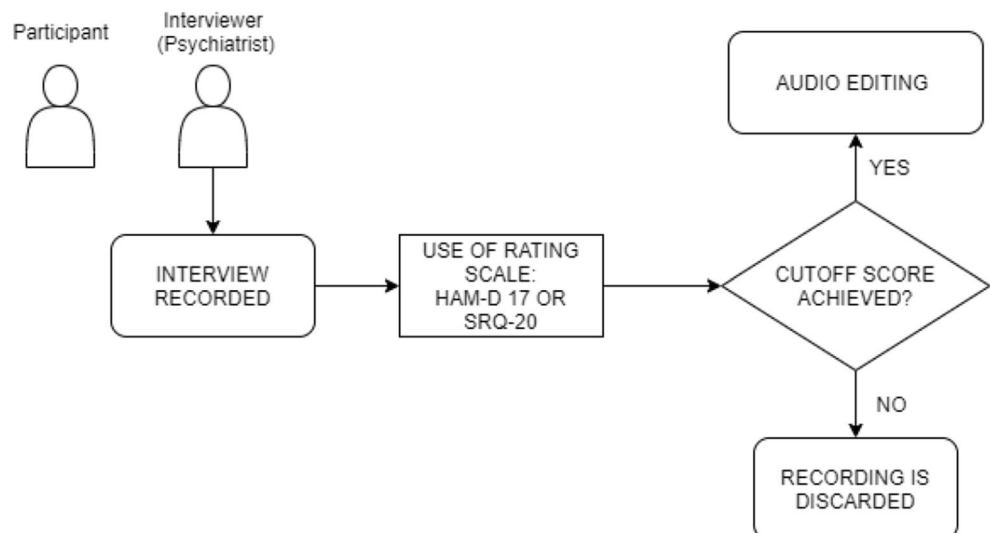


Table 2 Recording duration after audio editing

| Group | Number of participants | Total recording duration | Mean recording duration |
|------------|------------------------|--------------------------|--------------------------------|
| Control | 11 | 5816 s (96.9 min) | 528 s (8.8 min) \pm 138.7 s |
| Depression | 22 | 10,444 s (174.1 min) | 4675 s (7.8 min) \pm 213.3 s |

the depression group. Table 2 provides detailed information for both groups.

Feature extraction

All recordings were submitted to vocal feature extraction on GNU Octave™, a free open-source signal-processing software. We used rectangular windows, frame length of 10 s with 50% overlap. As raw audio data was used, no filtering process was applied. During this stage, we extracted the following 33 features: skewness; kurtosis; zero crossing rates; 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. We wanted to add a scope of features as broad as possible, avoiding human selection, which could introduce some a priori knowledge not desirable for this study.

This broad set of characteristics combines representation in the time domain and in the frequency domain. It is an assumption of this work to use a sign representation approach that does not depend on human specialist knowledge, avoiding potential representation biases. Given that (1) this is an approach that does not assume prior knowledge regarding the origin and nature of the signal over time, (2) good results were obtained with electroencephalographic and electromyographic signals (da Silva Júnior et al., 2019; de Freitas et al. 2019), (3) rectangular windows were used with both

electroencephalography and electromyography, (4) although audio signals have very different characteristics from other signals over time that would not guarantee acceptable results for the same approach, the results with electroencephalographic and electromyographic signals are a good motivation for investigating this approach in audio signals.

In order to investigate the influence of a reduced set of features and possible model generalization improvements, we added a feature selection stage based on genetic algorithms using a J48 decision tree as objective function. The parameters of the genetic algorithms were set empirically as follows: crossover probability of 0.6, mutation probability of 0.1, population size of 20 individuals, and 500 generations. Individuals are represented as vectors of 33 binary chromosomes; each chromosome is related to the selection of a specified feature. The objective/fitness function conducts the heuristic search according to the overall classification accuracy.

Classification experiments

Classes were balanced by adding artificial instances on Weka™ artificial intelligence environment. Class balancing was performed using Weka's class balancer method, consisting of the estimation of a uniform distribution parameters from the instances of the minority class to generate random instances until the number of instances of both classes equalize. This step is important so as to avoid computational biases towards the class with more representativeness, in this case the depression class. All experiments were performed using tenfold cross-validation. Figure 2 summarizes the steps of our proposed solution. The hyper-parameters of the

Fig. 2 Block diagram of proposed solution, considering audio edition, vocal feature extraction, and the binary classification between depressed subjects and the ones integrating the healthy control group

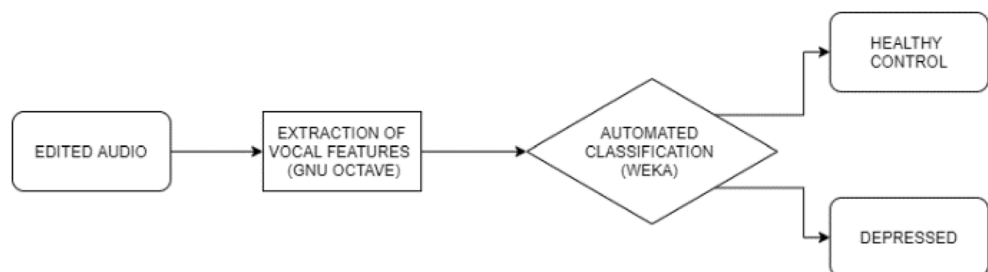


Table 3 Classification performance of multiple classifier configurations without automatic feature selection for binary classification (control vs. depression)

| Classifier | Configuration | Accuracy (%) | | Kappa index | | Sensibility | | Specificity | | |
|------------------------|--------------------------------|--------------------|---------|-------------|--------|-------------|--------|-------------|--------|--------|
| | | Mean | Std | Mean | Std | Mean | Std | Mean | Std | |
| MLP CS | 50 neurons in the hidden layer | 82.5980 | 2.0382 | 0.6516 | 0.0408 | 0.8704 | 0.0506 | 0.7808 | 0.0548 | |
| Random forest | 10 trees | 84.2124 | 1.6628 | 0.6838 | 0.0333 | 0.9057 | 0.0207 | 0.7773 | 0.0299 | |
| | 50 trees | 87.0486 | 1.5877 | 0.7407 | 0.0318 | 0.1932 | 0.0206 | 0.8269 | 0.0261 | |
| | 100 trees | 87.5505 | 1.5902 | 0.7508 | 0.0319 | 0.9149 | 0.0204 | 0.8354 | 0.0254 | |
| Support vector machine | Polynomial kernel | exp = 1; c = 0.01 | 61.2140 | 1.8346 | 0.2201 | 0.0370 | 0.8932 | 0.0234 | 0.3258 | 0.0288 |
| | | exp = 1; c = 0.1 | 64.9012 | 1.6526 | 0.2940 | 0.0334 | 0.9513 | 0.0160 | 0.3410 | 0.0287 |
| | | exp = 1; c = 1.0 | 66.7969 | 1.7595 | 0.3328 | 0.0355 | 0.9149 | 0.0197 | 0.4164 | 0.0296 |
| | | exp = 2; c = 0.01 | 65.0891 | 1.6300 | 0.2976 | 0.0330 | 0.9620 | 0.0141 | 0.3339 | 0.0287 |
| | | exp = 2; c = 0.1 | 68.4280 | 1.6520 | 0.3654 | 0.0334 | 0.9425 | 0.0167 | 0.4212 | 0.0291 |
| | | exp = 2; c = 1.0 | 73.4987 | 1.8160 | 0.4679 | 0.0366 | 0.9403 | 0.0160 | 0.5258 | 0.0320 |
| | | exp = 3; c = 0.01 | 67.3124 | 1.6824 | 0.3428 | 0.0340 | 0.9515 | 0.0154 | 0.3895 | 0.0300 |
| | | exp = 3; c = 0.1 | 71.0645 | 1.6231 | 0.4185 | 0.0328 | 0.9596 | 0.0139 | 0.4570 | 0.0300 |
| | | exp = 3; c = 1.0 | 76.1384 | 1.7817 | 0.5211 | 0.0359 | 0.9398 | 0.0175 | 0.5796 | 0.0315 |
| | Puk Kernel | c = 0.01 | 66.9048 | 2.1044 | 0.3358 | 0.0423 | 0.8524 | 0.0285 | 0.4822 | 0.0310 |
| | | c = 0.1 | 76.5444 | 1.8409 | 0.5295 | 0.0370 | 0.9218 | 0.0187 | 0.6061 | 0.0309 |
| | | c = 1.0 | 84.5768 | 1.6206 | 0.6911 | 0.0325 | 0.9090 | 0.0190 | 0.7813 | 0.0265 |
| | RBF Kernel | c = 0.01; g = 0.01 | 50.4675 | 0.1133 | 0.0000 | 0.0000 | 1.0000 | 0.0000 | 0.0000 | 0.0000 |
| | | c = 0.1; g = 0.01 | 57.9765 | 1.5706 | 0.1537 | 0.0319 | 0.9428 | 0.0193 | 0.2099 | 0.0245 |
| | | c = 1.0; g = 0.01 | 62.5230 | 1.8477 | 0.2464 | 0.0373 | 0.9087 | 0.0224 | 0.3364 | 0.0292 |
| c = 0.01; g = 0.25 | | 59.6523 | 1.9336 | 0.1888 | 0.0390 | 0.8710 | 0.0255 | 0.3169 | 0.0301 | |
| c = 0.1; g = 0.25 | | 64.8053 | 1.6797 | 0.2920 | 0.0340 | 0.9529 | 0.0150 | 0.3375 | 0.0296 | |
| c = 1.0; g = 0.25 | | 69.5692 | 1.6450 | 0.3885 | 0.0332 | 0.9432 | 0.0159 | 0.4435 | 0.0297 | |
| | c = 0.01; g = 0.5 | 60.7336 | 2.0596 | 0.2108 | 0.0415 | 0.8641 | 0.0275 | 0.3457 | 0.0307 | |
| | c = 0.1; g = 0.5 | 66.8632 | 1.7168 | 0.3338 | 0.0347 | 0.9443 | 0.0162 | 0.3877 | 0.0302 | |
| | c = 1.0; g = 0.5 | 74.3307 | 1.7107 | 0.4846 | 0.0345 | 0.9471 | 0.0151 | 0.5357 | 0.0312 | |
| Bayes Net | | 69.2008 | 2.2104 | 0.3830 | 0.0443 | 0.7808 | 0.0291 | 0.6015 | 0.0327 | |
| Naive Bayes | | 60.6848 | 1.9109 | 0.2096 | 0.0386 | 0.8774 | 0.0230 | 0.3312 | 0.0312 | |
| Logistic | | 71.1300 | 2.0332 | 0.4216 | 0.0408 | 0.7991 | 0.0276 | 0.6218 | 0.0325 | |

The table shows the mean and standard deviation values of four performance metrics (accuracy, kappa index, sensitivity, and specificity). Among these 28 configurations, the random forest classifier with 100 trees presented the best results, with a mean accuracy of 87.5505, mean kappa index of 0.7508, and mean sensitivity and specificity of 0.9132 and 0.8269, respectively

MLP multilayer perceptron, *PUK* Pearson Universal VII Kernel, *RBF* radial basis function, *exp* exponent, *c* the complexity parameter, *g* gamma

classifiers were empirically determined based on the experience of the research group and the results of the related literature.

Feature vectors were submitted to experiments with the following ML algorithms on Weka™: multilayer perceptron, logistic regression, random forests, decision trees, Bayes net, Naïve Bayes, and support vector machines with different kernels (linear, polynomial kernel, radial basis function or RBF, and PUK), described as following:

- Multilayer perceptron (MLP): learning rate 0.3, momentum 0.2, 50 neurons in the hidden layer
- Random forests (RF): 10, 50, and 100 trees

- Support vector machines (SVM):
 - Parameter C for 0.01, 0.1, and 1.0
 - Polynomial kernel: 1 (linear), 2 and 3 degree
 - Radial basis function (RBF) kernel: Gamma for 0.01, 0.25, and 0.5, i.e., a small value, a medium value, and a value corresponding to a pure Gaussian curve, respectively
 - Pearson Universal VII Kernel (PUK) with preset hyperparameters
- Bayes Network
- Naïve Bayes classifier

- Logistic regression classifier

Results

Experiments for this exploratory study were initially made under default settings on Weka™. After this, we tested different setups for all algorithms with adjustable settings (MLP; polynomial kernel and normalized polynomial kernel SVM, SVM PUK kernel, and random forest) with tenfold cross-validation and 30 repetitions for each configuration. Tables 3 and 4 describe in details our best results for each ML model, considering overall accuracy, kappa index, sensitivity, and specificity, both sample

mean and standard deviation, for datasets without and with automatic feature selection, respectively. Figures 3 and 4 present boxplots of multiple configurations for overall accuracy, kappa index, sensitivity, and specificity, for datasets without and with automatic feature selection using genetic algorithms, in this order. For SVM with polynomial, PUK, and RBF kernels, we only presented the best results for each type of kernel.

Discussion

Through analysis of Table 3, when using all extracted attributes, we notice that classification mean accuracy varied significantly for SVM classifier (50.4675–

Table 4 Classification performance for multiple machine learning techniques with automatic feature selection for binary classification (control vs. depression)

| Classifier | Configuration | Accuracy (%) | | Kappa index | | Sensibility | | Specificity | | | |
|------------------------|--------------------------------|--------------------|--------------------|-------------|--------|-------------|--------|-------------|--------|--------|--------|
| | | Mean | Std | Mean | Std | Mean | Std | Mean | Std | | |
| MLP CS | 50 neurons in the hidden layer | 71.8244 | 2.3886 | 0.4353 | 0.0477 | 0.8292 | 0.0923 | 0.6051 | 0.0870 | | |
| Random forest | 10 trees | 78.5829 | 2.0767 | 0.5710 | 0.0417 | 0.8590 | 0.0246 | 0.7113 | 0.0338 | | |
| | 50 trees | 80.3147 | 2.0433 | 0.6059 | 0.0410 | 0.8555 | 0.0239 | 0.7498 | 0.0323 | | |
| | 100 trees | 80.5193 | 1.9490 | 0.6100 | 0.0391 | 0.8548 | 0.0243 | 0.7547 | 0.0323 | | |
| Support vector machine | Polynomial kernel | exp = 1; c = 0.01 | 59.9273 | 1.2527 | 0.1924 | 0.0255 | 0.9976 | 0.0032 | 0.1934 | 0.0251 | |
| | | exp = 1; c = 0.1 | 64.4162 | 1.7884 | 0.2844 | 0.0361 | 0.9288 | 0.0194 | 0.3541 | 0.0295 | |
| | | exp = 1; c = 1.0 | 64.5513 | 2.0778 | 0.2881 | 0.0418 | 0.8616 | 0.0287 | 0.4253 | 0.0310 | |
| | Puk Kernel | exp = 2; c = 0.01 | 55.0747 | 0.9435 | 0.0938 | 0.0190 | 0.9998 | 0.0009 | 0.0932 | 0.0188 | |
| | | exp = 2; c = 0.1 | 62.3456 | 1.4367 | 0.2418 | 0.0291 | 0.9758 | 0.0102 | 0.2645 | 0.0267 | |
| | | exp = 2; c = 1.0 | 65.0123 | 1.7493 | 0.2965 | 0.0353 | 0.9305 | 0.0189 | 0.3645 | 0.0299 | |
| | RBF Kernel | exp = 3; c = 0.01 | 52.2193 | 0.6509 | 0.0357 | 0.0131 | 1.0000 | 0.0000 | 0.0354 | 0.0130 | |
| | | exp = 3; c = 0.1 | 59.8450 | 1.2794 | 0.1908 | 0.0261 | 0.9961 | 0.0040 | 0.1933 | 0.0253 | |
| | | exp = 3; c = 1.0 | 63.8248 | 1.5419 | 0.2720 | 0.0312 | 0.9653 | 0.0132 | 0.3051 | 0.0284 | |
| | | c = 0.01 | 64.8589 | 1.6794 | 0.2932 | 0.0339 | 0.9450 | 0.0165 | 0.3465 | 0.0294 | |
| | | c = 0.1 | 65.9921 | 1.8279 | 0.3163 | 0.0370 | 0.9302 | 0.0197 | 0.3845 | 0.0322 | |
| | | c = 1.0 | 68.6981 | 1.9250 | 0.3711 | 0.0389 | 0.9257 | 0.0207 | 0.4437 | 0.0338 | |
| | Bayes Net | c = 0.01; g = 0.01 | c = 0.01; g = 0.01 | 50.4675 | 0.1133 | 0.0000 | 0.0000 | 1.0000 | 0.0000 | 0.0000 | 0.0000 |
| | | | c = 0.1; g = 0.01 | 50.4675 | 0.1133 | 0.0000 | 0.0000 | 1.0000 | 0.0000 | 0.0000 | 0.0000 |
| | | | c = 1.0; g = 0.01 | 62.7699 | 1.4690 | 0.2505 | 0.0298 | 0.9733 | 0.0109 | 0.2756 | 0.0267 |
| c = 0.01; g = 0.25 | | c = 0.01; g = 0.25 | 50.6417 | 0.2406 | 0.0036 | 0.0042 | 1.0000 | 0.0000 | 0.0035 | 0.0042 | |
| | | c = 0.1; g = 0.25 | 64.2915 | 1.6409 | 0.2816 | 0.0332 | 0.9511 | 0.0162 | 0.3289 | 0.0285 | |
| | | c = 1.0; g = 0.25 | 64.8501 | 1.8807 | 0.2935 | 0.0379 | 0.9101 | 0.0246 | 0.3820 | 0.0299 | |
| c = 0.01; g = 0.5 | | c = 0.01; g = 0.5 | 59.1792 | 1.1618 | 0.1772 | 0.0236 | 0.9984 | 0.0026 | 0.1775 | 0.0233 | |
| | | c = 0.1; g = 0.5 | 64.6463 | 1.6783 | 0.2889 | 0.0339 | 0.9449 | 0.0169 | 0.3424 | 0.0290 | |
| | | c = 1.0; g = 0.5 | 65.4375 | 1.8791 | 0.3054 | 0.0379 | 0.9061 | 0.0232 | 0.3979 | 0.0315 | |
| Naive Bayes | | 70.2294 | 2.2015 | 0.4029 | 0.0442 | 0.8458 | 0.0331 | 0.5560 | 0.0389 | | |
| Logistic | | 61.5736 | 1.6840 | 0.2269 | 0.0342 | 0.9286 | 0.0182 | 0.2970 | 0.0301 | | |
| | | 65.0308 | 2.1657 | 0.2991 | 0.0434 | 0.7609 | 0.0324 | 0.5376 | 0.0316 | | |

In general, after selecting attributes, the random forest classifier with 100 trees showed the best results of accuracy, kappa index, and specificity. Despite the SVMs stand out in terms of sensitivity, the other metrics showed poor performance

MLP multilayer perceptron, PUK Pearson Universal VII Kernel, RBF Radial Basis Function, exp exponent, c the complexity parameter, g gamma

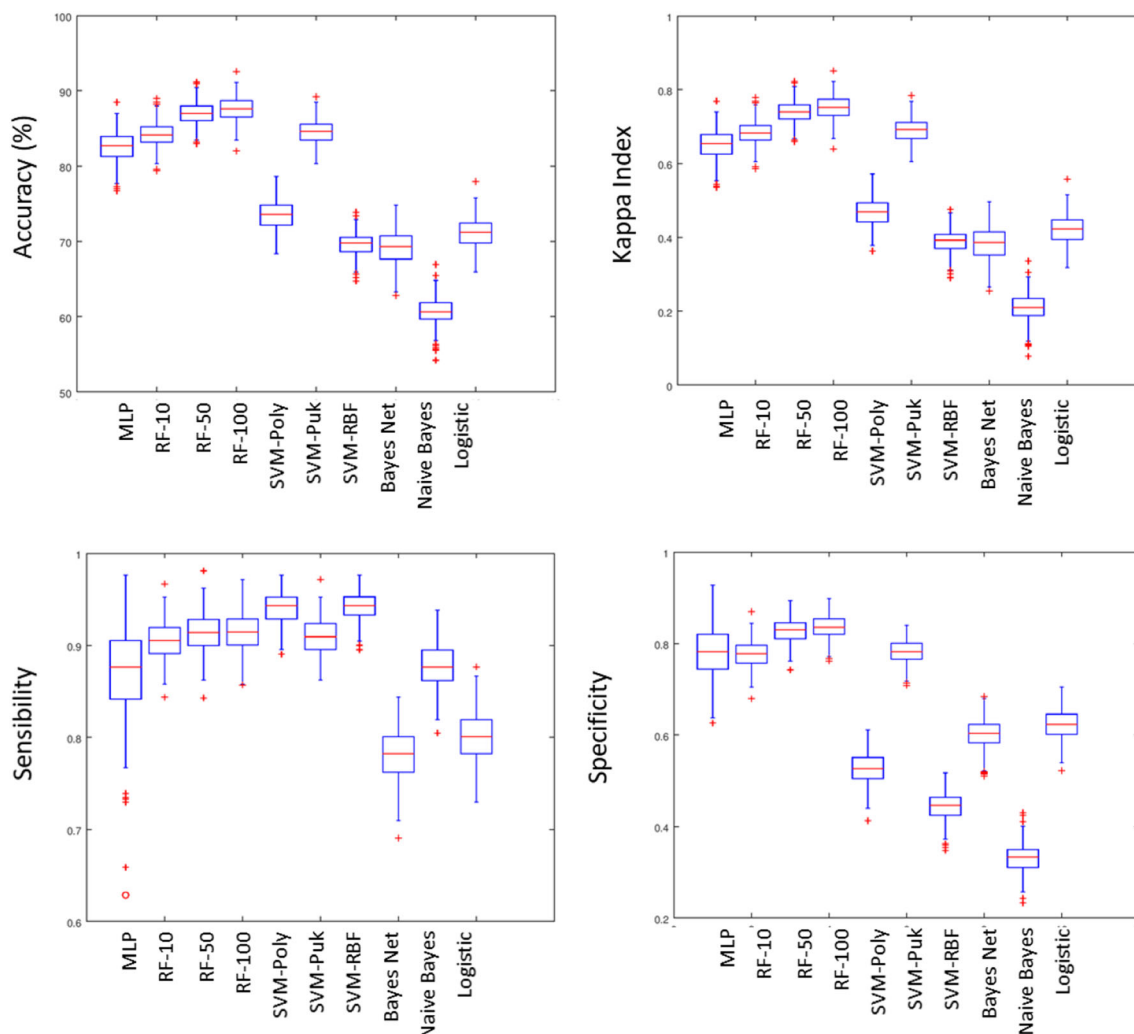


Fig. 3 Boxplots with classification performance to distinguish depressive patients from the control group using all extracted attributes. Random forest with 100 trees performed better, with highest values of accuracy,

kappa index, and specificity. In addition, it achieved a reasonable value of sensitivity. In the case of SVMs, only the configurations with the best results for each type of kernel were plotted

84.5768%), depending on the kernel used. Furthermore, to the best of our knowledge, this is the first study to compare performances of several SVM kernels for the detection of depression. It also highlights the need of further studies for discriminating the impact of different kernels in the discriminative power of SVM classifiers.

However, random forest with 100 trees provided the highest accuracy ($87.5575\% \pm 1.9490$) among all classifiers in this study. Similarly, the kappa index and specificity values were the highest for this configuration: 0.7508 ± 0.0319 and 0.8354 ± 0.0254 , respectively. In contrast, SVMs with an RBF kernel showed the highest sensitivity values, with a mean of 1 in some cases. Despite this, SVMs performed poorly in other metrics. In addition, we can see that classifiers based on Bayes theory showed inferior results. This may indicate that

the attributes considered in this study are statistically dependents.

It is also important to highlight that the automatic selection of attributes resulted in a worsening of the classification performance (Table 4). This suggests that the use of all 34 extracted attributes is important in the binary classification. As in the original scenario, random forests with 100 trees showed better results after feature selection with genetic algorithms, with mean accuracy of 80.5193 ± 1.9490 , mean kappa index of 0.6100 ± 0.0391 , and mean sensitivity and specificity of 0.8548 ± 0.0243 and 0.7547 ± 0.0323 , respectively.

The confusion matrix for the best classifier (random forest with 100 trees for all attributes) is shown in Table 5. As can be seen, 89.72% of participants belonging to the control group were classified as control,

the best results for each type of kernel were plotted.

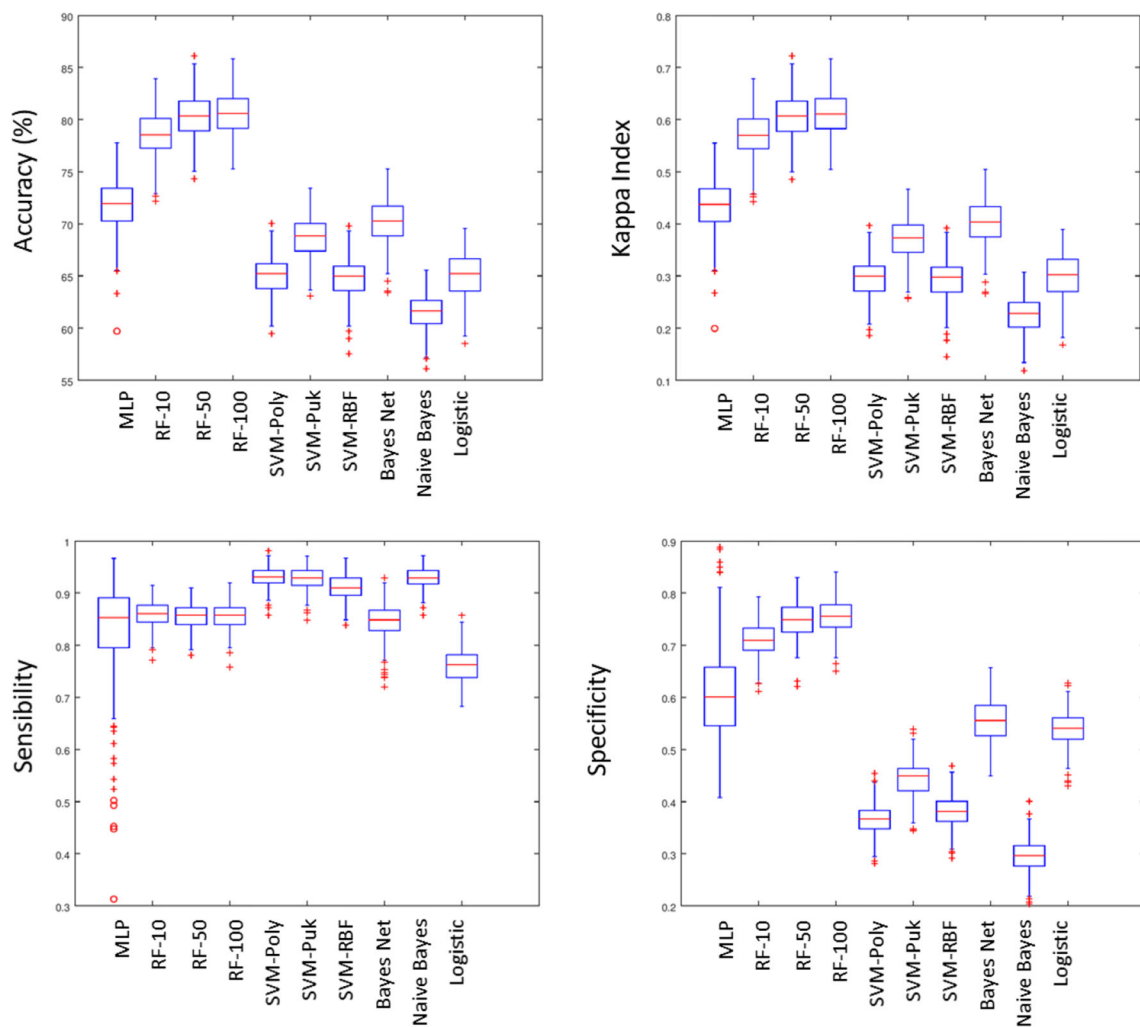


Fig. 4 Classification performance to distinguish depressive patients from the control group using selected attributes by an automated method. The random forest achieved better results, being visually similar when tested with 50 or 100 trees

while 83.74% of patients with depression were classified correctly. It is important to notice that there is greater confusion in the group of people with depression, with 16.26% being classified as healthy.

As mentioned earlier, except for the work of Higuchi et al. (2018), we achieved better results than other previously published studies. Our work provided high

classification accuracy both for depressed and healthy individuals. However, it is important to note that our small sample size may limit statistical interpretations. Factors that may influence vocal acoustic properties, such as smoking history, pharmacotherapy, and demographic variables such as age, gender and educational level, were not controlled and represent a limitation. In

Table 5 Confusion matrix for the model with the highest performance classifier (random forest with 100 trees), considering all extracted attributes

| | Classified as control | Classified as depression | Classified as depression |
|------------|-----------------------|--------------------------|--------------------------|
| Control | 89.72% | 10.28% | 11.75% |
| Depression | 16.26% | 83.74% | 90.03% |

a prospective study, we aim to control possible confounders and repeat the same experiments with more participants.

Conclusion

Current psychiatric diagnosis still lacks objective biomarkers and relies mostly on specialist opinion based on diagnostic manuals. Nevertheless, such diagnostic systems have been heavily criticized due to their absence of correlation with the neurobiology and etiopathogenesis of mental disorders. Among these, depression presents with vocal acoustic alterations that may be used as objective parameters for the identification of this disorder.

Therefore, this exploratory study focused on the development of an auxiliary instrument for the diagnosis of depressive disorders. To this end, we extracted vocal acoustic features and performed experiments using different automated classification techniques. Some of the most widely used classifiers were also tested in this work. Our results suggest the viability of a machine learning tool for the detection and even screening of major depressive disorder in a cost-effective and non-invasive manner. In future studies, we intend to perform the same experiments in a larger sample with age-matched groups, as well as with controlled disease severity and gender-based datasets. With this approach, we aim to evaluate the impact of depression severity on vocal acoustic parameters. We also would like to assess if depression unequally affects vocal acoustic properties from men and women and how such differences influence the performance of automated classifiers.

Acknowledgments The authors are grateful to the Brazilian research agency CNPq for the partial financial support of this research.

Funding This study was partially funded by the Brazilian research agency CNPq.

Compliance with ethical standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

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

References

Afshan A, Guo J, Park S J, Ravi, V, Flint, J, Alwan, A (2018). Effectiveness of voice quality features in detecting depression. *Proceedings of the Annual Conference of the International Speech*

Communication Association, INTERSPEECH, 2018-Septe(September), 1676–1680. <https://doi.org/10.21437/Interspeech.2018-1399>

- Alghowinem, S, Goecke, R, Wagner, M, Epps, J., Breakspear, M, Parker, G (2012). From joyous to clinically depressed: mood detection using spontaneous speech. In *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25* (pp. 141–146).
- Alghowinem S, Goecke R, Wagner M, Epps J. Detecting depression: a comparison between spontaneous and read speech. *IEEE*. 2013a: 7547–51.
- Alghowinem, S, Goecke, R, Wagner, M, Epps, J., Gedeon, T, Breakspear, M, Parker, G (2013b). A comparative study of different classifiers for detecting depression from spontaneous speech. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (pp. 8022–8026). <https://doi.org/10.1109/ICASSP.2013.6639227>.
- Alpert M, Pouget ER, Silva RR. Reflections of depression in acoustic measures of the patient's speech. *J Affect Disord*. 2001;66:59–69.
- American Psychiatric Association. (2013). *DSM-5 - Manual Diagnóstico e Estatístico de Transtornos Mentais*. Artmed (5.). Porto Alegre: Artmed. <https://doi.org/10.1176/9780890425596>.
- Arjmandi MK, Pooyan M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomed. Signal Process. Control*. 2012;7(2012):3–19. <https://doi.org/10.1016/j.bspc.2011.03.010>.
- 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/npjscz.2015.30>.
- Bzdok, D, Meyer-lindenberg, A (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3, 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>.
- Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. *Brain Cogn*. 2004;56:30–5. <https://doi.org/10.1016/j.bandc.2004.05.003>.
- Cohn, J. F, Krueez, T. S, Matthews, I, Yang, Y, Nguyen, M. H, Padilla, M. T, ... De La Torre, F. (2009). Detecting depression from facial actions and vocal prosody. *Proceedings - 2009 3rd International conference on affective computing and intelligent interaction and workshops, ACII 2009*, (October). <https://doi.org/10.1109/ACII.2009.5349358>.
- 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.
- Cordeiro FR, Lima SM, Silva-Filho AG, Santos WP. Segmentation of mammography by applying extreme learning machine in tumor detection, In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 92–100). Berlin Heidelberg: Springer; 2012.
- Cordeiro FR, Santos WP, Silva-Filho AG. A semi-supervised fuzzy GrowCut algorithm to segment and classify regions of interest of mammographic images. *Expert Syst Appl*. 2016;65:116–26.
- Cummins N, Epps J, Sethu V, Krajewski J. Variability compensation in small data: oversampled extraction of i-vectors for the classification of depressed speech. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - proceedings*; 2014. p. 970–4. <https://doi.org/10.1109/ICASSP.2014.6853741>.

- Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Comm.* 2015;71(April):10–49. <https://doi.org/10.1016/j.specom.2015.03.004>.
- da Silva Junior M, de Freitas RC, dos Santos WP, da Silva WWA, Rodrigues MCA, Conde EFQ. Exploratory study of the effect of binaural beat stimulation on the EEG activity pattern in resting state using artificial neural networks. *Cogn Syst Res.* 2019;54:1–20.
- Darby JK, Hollien H. Vocal and speech patterns of depressive patients. *Folia Phoniatr.* 1977;29:279–91.
- de Freitas RC, Alves R, da Silva Filho AG, de Souza RE, Bezerra BL, dos Santos WP. Electromyography-controlled car: a proof of concept based on surface electromyography, extreme learning machines and low-cost open hardware. *Comput Electr Eng.* 2019;73:167–79.
- 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.
- dos Santos, W. P, de Souza, R. E, dos Santos Filho, P. B (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).
- dos Santos, W. P, de Assis, F. M, de Souza, R. E, Santos, D, Filho, P. B (2008). Evaluation of Alzheimer's disease by analysis of MR images using objective dialectical classifiers as an alternative to ADC maps. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5506–5509).
- 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.
- Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry.* 2016;6(7):e856–8. <https://doi.org/10.1038/tp.2016.123>.
- 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>.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56–62.
- Hasan R, Jamil M, Rabbani G, Rahman S. Speaker identification using Mel frequency cepstral coefficients. In: *3rd International Conference on Electrical & Computer Engineering ICECE 2004*, (December); 2004. p. 565–8.
- Hashim NW, Wilkes M, Salomon R, Meggs J, France DJ. Evaluation of voice acoustics as predictors of clinical depression scores. *J Voice.* 2016;31:256.e1–6. <https://doi.org/10.1016/j.jvoice.2016.06.006>.
- Higuchi, M, Tokuno, S, Nakamura, M, Shinohara, S (2018). Classification of bipolar disorder, major depressive disorder, and healthy state using voice. *Asian Journal of Pharmaceutical and Clinical Research*, 11(3), 89–93. <https://doi.org/10.22159/ajpcr.2018.v11i3.30042>
- Hönig, F, Batliner, A, Nöth, E, Schnieder, S, Krajewski, J. (2014). Automatic modelling of depressed speech: relevant features and relevance of gender. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, (444), 1248–1252.
- Iwabuchi SJ, Liddle PF, Palaniyappan L. Clinical utility of machine-learning approaches in schizophrenia: improving diagnostic confidence for translational neuroimaging. *Front. Psychol.* 2013;4(August):1–9. <https://doi.org/10.3389/fpsy.2013.00095>.
- Jiang H, Hu B, Liu Z, Yan L, Wang T, Liu F, et al. Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Comm.* 2017;90:39–46. <https://doi.org/10.1016/j.specom.2017.04.001>.
- 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;2018:2018–9. <https://doi.org/10.1155/2018/6508319>.
- Joshi J, Goecke R, Alghowinem S, Dhall A, Wagner M, Epps J, et al. Multimodal assistive technologies for depression diagnosis and monitoring. *Journal on Multimodal User Interfaces.* 2013;7(3): 217–28. <https://doi.org/10.1007/s12193-013-0123-2>.
- Liu, Z, Hu, B, Yan, L, Wang, T., Liu, F., Li, X., Kang, H. (2015). Detection of depression in speech. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 743–747).
- Low LSA, Maddage NC, Lech M, Sheeber LB, Allen NB. Detection of clinical depression in adolescents' speech during family interactions. *IEEE Trans Biomed Eng.* 2011;58(3 PART 1):574–86. <https://doi.org/10.1109/TBME.2010.2091640>.
- Maxhuni A, Muñoz-meléndez A, Osmani V, Perez H, Mayora O, Morales EF. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive Mob. Comput.* 2016;31(1):50–66. <https://doi.org/10.1016/j.pmcj.2016.01.008>.
- Mcginnis EW, Anderau SP, Hruschak J, Gurchiek RD, Lopez-duran NL, Fitzgerald K, et al. Giving voice to vulnerable children: machine learning analysis of speech detects anxiety and depression in early childhood. *IEEE.* 2019;23:1–8. <https://doi.org/10.1109/JBHL.2019.2913590>.
- Miguel EC, Gentil V, Gattaz WF. Clínica Psiquiátrica. Barueri: Manole; 2011.
- Mitra, V., Shriberg, E (2015). Effects of feature type, Learning Algorithm and Speaking Style for Depression Detection from Speech *IEEE*, 4774–4778.
- Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralt 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>.
- Ooi KEB, Lech M, Brian Allen N. Multichannel weighted speech classification system for prediction of major depression in adolescents. *IEEE Trans Biomed Eng.* 2013;60(2):497–506. <https://doi.org/10.1016/j.bspc.2014.08.006>.
- Ozdas A, Shiavi RG, Silverman SE, Silverman MK, Wilkes DM. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Trans Biomed Eng.* 2004;51(9):1530–40. <https://doi.org/10.1109/TBME.2004.827544>.
- Rodrigues AL, de Santana MA, Azevedo WW, Bezerra RS, Barbosa VA, de Lima RC, et al. Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. *Res Bio Eng.* 2019;35(3):213–22.
- Sadock B, Sadock V, Ruiz P. *Compêndio de Psiquiatria: Ciência do Comportamento e Psiquiatria Clínica* (11.). Porto Alegre: Artmed; 2017.
- 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.
- Scherer, S., Stratou, G., Gratch, J., Morency, L. P. (2013). Investigating voice quality as a speaker-independent indicator of depression and PTSD. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, (August), 847–851.

- Sellam V, Jagadeesan J. Classification of normal and pathological voice using SVM and RBFNN. *Journal of Signal and Information Processing*. 2014;5:1–7. <https://doi.org/10.5120/9675-4102>.
- Vanello, N., Guidi, A., Gentili, C., Werner, S., Bertschy, G., Valenza, G., ... Scilingo, E. P. (2012). Speech analysis for mood state characterization in bipolar patients. In *34th Annual International Conference of the IEEE EMBS* (pp. 2104–2107).
- Wang X, Zhang J, Yan Y. Discrimination between pathological and normal voices using GMM-SVM approach. *J Voice*. 2011;25(1): 38–43. <https://doi.org/10.1016/j.jvoice.2009.08.002>.
- Weinberger AH, Gbedemah M, Martinez AM, Nash D, Galea S, Goodwin RD. Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychol Med*. 2017;48:1–10. <https://doi.org/10.1017/S0033291717002781>.
- World Health Organization. Depression and other common mental disorders global health estimates. Geneva: Switzerland; 2017.
- World Health Organization. (2018). Depression. Retrieved November 11, 2019, from <https://www.who.int/en/news-room/fact-sheets/detail/depression>
- Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton depression rating scale. *J Affect Disord*. 2013;150(2):384–8. <https://doi.org/10.1016/j.jad.2013.04.028>.

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