# Detection of Major Depressive Disorder, Bipolar Disorder, Schizophrenia and Generalized Anxiety Disorder Using Vocal Acoustic Analysis and Machine Learning

**Caroline Wanderley Espinola**
Universidade Federal de Pernambuco
**Juliana Carneiro Gomes**
Universidade de Pernambuco
**Jessiane Mônica Silva Pereira**
Universidade de Pernambuco
**Wellington Pinheiro dos Santos** ( wellington.santos@ufpe.br )
Universidade Federal de Pernambuco https://orcid.org/0000-0003-2558-6602

**Research Article**

**Keywords:** Major depressive disorder, bipolar disorder, schizophrenia, generalized anxiety disorder, diagnosis, voice, acoustic parameters, machine learning, support vector machines

**Posted Date:** March 28th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-648044/v1

# Abstract

### Purpose

Current diagnosis and treatment in psychiatry are heavily depends patients' reports and clinician judgement, thus are sensitive to memory and subjectivity biases. On the other hand, a growing effort in developing objective markers in psychiatry may be a way to face all these challenges. Vocal acoustic features have been recently studied as objective measures, with the advantages of being accessible, inexpensive, non-invasive and remotely performed.

### Method

The main objective of this work is to propose a methodology to support the diagnosis of major depressive disorder, bipolar disorder, schizophrenia, and generalized anxiety disorder using vocal acoustic analysis and machine learning. Seventy-eight individuals over 18 years old were recruited into five groups: 28 participants with major depressive disorder; 20 patients with schizophrenia; 15 patients with bipolar disorder; 4 patients with generalized anxiety disorder; and 12 healthy controls.

### Results

Recordings were submitted to a vocal feature extraction algorithm, and to experiments using different machine learning classification techniques. Random Forests with 300 trees achieved the greatest classification performance (75.27% for accuracy, 69.08% for kappa, 75.30% for sensitivity, and 93.80% for specificity) for the simultaneous detection of major depressive disorder, schizophrenia, bipolar disorder and generalized anxiety disorder.

### Conclusion

As changes in vocal patterns have been reported in several mental disorders and appear to correlate with illness severity, vocal acoustic features have shown to be promising markers, with the advantages of being abundant, inexpensive, non-invasive and remotely performed. The results provided by our proposed solution support the feasibility of a computational method based on vocal parameters for assisting clinicians in patient triage and diagnosis in psychiatry.

As the main contribution of this work, we present the intelligent system composed of the support vector machine and the time and frequency characteristics of the audio, which works as a digital biomarker to support the diagnosis of mental disorders in the context of psychiatric emergency.


# 1. Introduction

# a. Motivation and problem characterization

Unlike other medical fields, such as oncology [1] and nephrology [2], psychiatry still lacks objective and standardized measurement [3, 4]. Current psychiatric diagnosis is built upon diagnostic criteria from the Diagnostic and Statistical Manual, 5th Edition (DSM-5), and the International Classification of Diseases (ICD-10). However, theses manuals have been criticized due to their lack of neurobiological validity and their poor diagnostic stability [5, 6], leading to low clinical predictability and trial-and-error treatments [6, 7]. Current clinical practice in psychiatry depends on subjective accounts from patients and on clinician judgement [8], which makes diagnosis and therapeutic decision prone to memory and subjectivity biases [9]. Recent research efforts have focused on the development of objective biomarkers in psychiatry, although most of them are invasive, expensive or require special medical equipment [10]. Regardless of all efforts, instruments for evaluation of mental disorders remain a challenge [8].

Despite substantial advances in clinical psychiatry in the last century, diagnoses can be vague, prognoses are unpredictable, and pharmacotherapy is frequently ineffective [11]. Additionally, progress in neurosciences is not converted into improvements in nosology or clinical practice in psychiatry [7]. A hypothesis for this stagnation are study designs and solutions based on statistical framework: by targeting minimizing within-sample error, this approach lacks replication and reproducibility, with poor generalizability to new samples [11, 12].

Other equally relevant problems are difficult access to mental health services, lack of mental health professionals in those facilities, both in developed and in developing countries. Moreover, stigmatization and negative attitudes towards psychiatry frequently delay treatment onset and adherence, imposing prolonged suffering with poor treatment outcomes and worse prognosis [4].

One strategy to address these problems is computational psychiatry, which applies machine learning (ML) to optimize generalizability at an individual level and provide clinical applications and personalized treatments [6, 7]. Machine learning is a data-driven approach particularly well suited to deal with the complexity of multivariate relations and interdependencies related to brain connections [11]; it allows dealing with high-dimensional data improve classification diagnosis, treatment selection and prognosis. This new approach has already shown promising results in translational medicine, like multiple sclerosis diagnosis [13], Alzheimer's disease diagnosis based on neuroanatomical structures [14, 15, 16], and breast cancer diagnosis [17].

Voice characteristics are some of the most important nonverbal cues we use to infer other peoples' emotional states [18]. Since neurophysiological pathways underlying speech patterns are modulated by the affective state of the speaker [19], mood changes are expressed in a person's voice [10]. These changes in speech production can be measured using prosodic, articulatory and voice quality features [20]. In this context, computational analysis of speech can be used to assess and monitor mental states of people suffering from a variety of mental disorders, such as major depressive disorder (MDD), bipolar disorder (BD), post-traumatic stress disorder (PTSD), autism, schizophrenia, and even suicidal behavior [19, 20, 21, 22, 23, 24].

Major depressive disorder (depression) is the most common mental disorder, affecting more than 300 million people worldwide [25, 26]. It is also a leading cause of disability and economic burden [27], and is related to around half of all suicides globally [28]. Depressed patients present with sadness, 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 [29]. Early depressive symptoms such as psychomotor retardation and cognitive impairment are frequently related to abnormalities in speech patterns [8, 30]. In particular, the persistently 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 may allow detection of depression through analysis of vocal acoustics of depressed patients [9].

Schizophrenia is a group of severe psychotic disorders with heterogeneous etiologies, clinical presentations and responses to treatment [25], with a worldwide lifetime prevalence of 1.11% [31]. It is characterized by delusions, hallucinations, thought and behavior disorder, and 'negative symptoms' [29]. The latter are considered an absence of normal function and consist of blunted affect, poverty of speech, avolition, anhedonia, and asociality [32]. Since its initial descriptions, speech abnormalities have been a hallmark feature of schizophrenia, and are often associated with core negative symptoms and social impairment [12]. These speech-language alterations consist of poverty of speech, disorganized speech, derailment, tangentiality, neologism, incoherence, mutism, perseveration, echolalia, thought blocking [33], and aprosodia [34, 35]. Also known as flattened speech intonation, aprosodia consists of diminished vocal emphasis [36]; reduced inflection and fluency [37]; and prosody comprehension deficits, such as difficulties in recognizing intonation patterns [35]. These symptoms result from deficits in semantic memory, working memory and executive function and contribute to the frequent communication disorders seen in schizophrenia [33, 38].

Formerly known as manic-depressive disorder, bipolar disorder is one of the most severe and potentially disabling mental disorders, and is associated with intense psychological suffering and high suicide rates [39]. With a lifetime prevalence as high as 2.4% for bipolar spectrum [40], BD is considered a worldwide leading cause of disability, with a mean age of onset of 20–30 years [25, 41]. In addition, BD is strongly tied to premature death from medical illnesses, accidents and suicide [41, 42], with suicidal behavior present in more than one third of bipolar patients [43]. The clinical presentation of BD is characterized by cyclic episodes of persistent elevated mood (expansivity or irritability) and increased energy, associated with increased psychomotor activity, cognitive changes (distractibility, racing thoughts), and neurovegetative functions such as reduced need for sleep. These episodes are known as hypomania or, when symptoms are severe, mania. Most bipolar patients also experience depressive episodes [29]. Within bipolar patients, language-speech disorders such as increased speech activity [22], changes in vocal pitch, intensity/loudness and rhythm are frequently present and correlate to mood swings [44]. Consequently, changes in vocal patterns can be used as a potential marker of affective state change from euthymia to (hypo)mania [22].

Anxiety disorders share dysfunctional behavior responses related to excessive fear and anxiety [29]. They are the most prevalent mental disorders, with a lifetime prevalence of up to 33.7%, and are associated with significant disability and elevated healthcare costs [45]. Within this diagnostic group, generalized anxiety disorder (GAD) is a chronic and debilitating disorder characterized by persistent worrying, concentration problems, insomnia, muscular tension, irritability and restlessness [25, 46]. Since vocalization depends on the integration of central and autonomic nervous systems, stress can induce changes in speech patterns, with respiration playing a key role in this relation between speech and stress [47]. According to Almeida, Behlau, & Leite [48], abnormalities in several speech parameters are directly related to the speaker's anxiety severity, such as impairment of speech modulation and articulation, vocal resonance imbalance, and changes in facial expressions. In anxious states, an increased subglottic pressure leads to changes in vocal parameters, with decreased vocalization of vowels. These acoustic differences can be perceived by the listener and may be used to detect anxiety through vocal analysis [49].

The main objective of this work is to propose a methodology to support the diagnosis of major depressive disorder, bipolar disorder, schizophrenia, and generalized anxiety disorder using vocal acoustic analysis and machine learning.

This work is organized as follows: Section 2 briefly discusses studies related to the identification of major depressive disorder, bipolar disorder, schizophrenia and anxiety disorders using voice. Section 3 describes in detail the implementation of an unprecedent voice-based instrument for the classification of MDD, BD, schizophrenia and GAD. In Section 4 our results are presented and discussed. Section 5 states our conclusion and suggestions for future work.

# b. Related Works

Speech production is the result of complex interactions between cognitive functions and musculoskeletal system, and slight physiological and cognitive changes due to intense fluctuations in affective states can yield noticeable acoustic changes [20]. Vocal and speech patterns in mental disorders have been reported in studies that date back to 1938, initially within affective disorders like depression and bipolar disorder [22, 50]. In the following decades, speech-language abnormalities have been reported within other mental disorders, including schizophrenia, autism, and anxiety disorders, such as social phobia and PTSD [18, 23, 24, 37, 51, 52, 53, 54].

In the context of depression, speech patterns consist of decreased pitch variability [55], reduced articulation rate [23], increased number and duration of pauses [27, 56], slower speech [22], reduced intensity/loudness and atypical voice quality [56]. For the recognition of changes in mood state, prosodic, phonetic and spectral components of voice are relevant, in particular pitch or fundamental frequency (F0), intensity, rhythm, speed, jitter, shimmer, energy distribution between formants, and cepstral features. Among these features, jitter is important for mood state recognition given its ability to detect quick momentary changes in voice [44]. Cepstral coefficients, particularly mel frequency cepstral coefficients

(MFCC), are also well suited for detecting depressed speech [57]. They consist of parametrical representation of the speech signal [58], and have been extensively studied for the identification of depression through vocal analysis [59].

In a sample of depressed patients, Cohn et al. [60] analyzed prosodic and facial expression cues 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.

In a study with adolescents, Ooi, Lech, & Brian Allen [61] 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% (sensibility: 79%; specificity: 67%). Using a larger sample of adolescents, Low, Maddage, Lech, Sheeber, & Allen [56] utilized the above parameters with the addition of cepstral features, and submitted to SVM and Gaussian Mixture Models (GMM) classifiers. They reported significant differences in classifier performances for detecting depression based on gender: 81–87% for males, and 72–79% for females.

Studying vocal patterns for the identification of depression, Hönig et al. [62] automatically selected 34 features: spectral (MFCCs, formants F1 to F4), prosodic (pitch, energy, duration, rhythm) voice quality or phonetic features (jitter, shimmer, raw jitter, raw shimmer, logarithm harmonics-to-noise ratio, spectral harmonicity and spectral tilt). In agreement with previous findings from Low et al. [56], they reported a slightly higher correlation in males than in females. This suggests that clinical depression can produce more significant changes in vocal features in men than in women.

Similarly, Jiang et al. [63] 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). Best results were achieved with SVM, with accuracy of 80.30% for males, and 75.96% for females.

Adversely, Higuchi et al. [10] analyzed pitch, spectral centroid and five MFCC parameters using polytomous logistic regression for the classification of depression, bipolar disorder and healthy controls. No difference between genders was found. An overall accuracy of 90.79% was reported, with accuracy of 93.33% for the binary classification between depression and healthy controls.

Another facet of study design that might influence the performance of automated classifiers is the type of speech task. Spontaneous speech (e.g., dialogues or interviews) is associated with higher classification performances than reading tasks. This finding suggests that spontaneous speech provides more acoustic variability, thereby improving the recognition of depression [63, 64]. Moreover, depressed individuals can probably suppress their affective state during reading tasks, because of the irrelevance of the content read or their concentration on reading, or even both [65].

Despite fewer publications, changes in vocal parameters have also been reported in bipolar disorder. For instance, pitch variations could help differentiate between bipolar patients and healthy individuals [44].

Higuchi et al. [10] reported different cepstral parameters (MFCCs), F0 envelope and spectral centroid in bipolar patients, depressed patients and healthy controls. In another study, Higuchi et al. [66] used polytomous logistic regression analysis of vocal features to distinguish between healthy and bipolar I (BD I) or II (BD II) disorder. Although their model could not easily distinguish between BD I and BD II, they reported overall accuracy of 66.7%, suggesting that vocal features could be used for the classification of bipolar patients.

In an attempt to monitor long-term mood states of bipolar patients, Karam et al. [19] recorded real-life cell phone conversations from six participants with BD for up to a year. Using SVM linear and radial basis function (RBF) kernels and feature selection, they reported an average AUC 0.63 ± 0.04 for detecting hypomania in three subjects, and 0.64 ± 0.16 for depression in four individuals. However, a critical limitation of this study is its relatively reduced sample size.

Faurholt-Jepsen et al. [22] also used phone calls to monitor illness activity in BD. They analyzed vocal features with and without phone data on social interaction (number of text messages and phone calls), motor activity (accelerometer data), and self-monitored data on mood of 28 bipolar outpatients for 12 weeks. They observed that increased speech activity may predict a mood switch to hypomania, while reduced speech activity and changes in pitch indicated prodromal symptoms of depression and response to treatment. Using random forest (RF) algorithms, they concluded that the combination of vocal features with smartphone and self-monitored data slightly improved the classification accuracy of affective states in BD to 73–77% for manic or mixed states, and 63–66% for depressive states in user-independent models.

Likewise, Maxhuni et al. [44] collected audio (prosodic and spectral features), accelerometer and self-assessment data from five bipolar patients over a period of twelve weeks during their routine activities. They evaluated the performance of several classifiers, and reported high confidence (≅ 85%) for the classification of relapses within those patients.

In a study with BD inpatients due to a mania episode, Ringeval et al. [67] used low-level descriptors (LLDs) of audio and video data to classify patients into mania, hypomania, and remission. Audio data included spectral, cepstral and voice quality features, while video data consisted of appearance and geometrical information. Features were analyzed using supervised, semi-supervised and unsupervised computational methods. A better performance for the unsupervised deep convolutional neural networks (CNN) was reported as opposed to supervised and semi-supervised ML algorithms. Their results emphasized the interest in unsupervised approaches such as deep learning for the representation of speech data in bipolar disorder.

As mentioned earlier, speech-language abnormalities are core features of schizophrenia. Patients with this disorder frequently present with slowed speech, reduced pitch variability, significantly increased number of pauses, and decreased variability in syllable timing than healthy individuals. These characteristics were observed by Martínez-sánchez et al. [68] in a semi-automatic analysis of pitch or fundamental frequency (F0) during an emotionally neutral reading task. A discrimination accuracy of

93.8% was reported between schizophrenic patients and healthy controls with audio signal processing algorithms. They also observed remarkable intergroup differences, with patients showing slower speech, with low volume and many pauses.

Vocal acoustic analysis can also measure the severity of negative symptoms such as aprosodia. Compton et al. [69] compared audio recordings of schizophrenic patients with aprosodia, schizophrenic patients without aprosodia, and healthy controls on variation in pitch (F0), first (F1) and second (F2) formants, and intensity/loudness. Their results showed significant differences for the group with aprosodia, with reduced variation in pitch, F2 and intensity/loudness than pacients without aprosodia and healthy controls.

Similarly, Covington et al. [70] analyzed F0, F1 and F2 of video-recorded interviews from schizophrenic patients. They investigated tongue movement as an indicator of negative symptom severity 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.

In a metanalysis of 46 articles of vocal patterns in schizophrenia, Alberto et al. [12] 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. All these abnormalities were 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 ones with constrained monologues.

Chakraborty, Yang, et al. [21] analyzed LLDs alone or in combination with body movements to predict negative symptoms of schizophrenia using SVM. They reported a classification accuracy of 79.49% using low-level speech signals alone, and of 86.36% for their combination with body movements.

Tahir et al. [71] also studied negative symptoms in schizophrenia using conversational and prosodic features. The former consisted of speaking duration, speaking turns, interruptions and interjections, while the latter corresponded to F0, formants F1, F2 and F3, MFCCs and amplitude (entropy, minimum, maximum and mean volume). The performance of four ML algorithms for the classification of patients and healthy controls was evaluated: SVM, Multilayer Perceptron (MLP), random forest, and ensemble (bagging). Better classification results were achieved using MLP, with accuracy of 81.3%, and speaking rate, frequency and volume entropy showing significant differences between groups.

With regards to vocal patterns of anxiety, there are several studies exploring vocal acoustic measures of stress or specific disorders, such as social anxiety disorder (SAD or social phobia) and PTSD. However, no previous study about vocal correlates of generalized anxiety disorder was found, even though it is a chronic, debilitating and common disorder with persistent anxiety symptoms.

In the context of SAD, Laukka et al. [18] compared acoustic parameters from audio samples of patients with social anxiety disorder before and after pharmacotherapy. They concluded that a reduction in experienced anxiety states following treatment was accompanied by decreases in these parameters (mean and maximum pitch, high-frequency components in energy spectrum, and proportion of silent pauses). A decrease in listeners' perceived level of anxiety was also reported.

From that standpoint, Weeks et al. [72] analyzed F0 (pitch) parameter of individuals diagnosed with SAD during a social task, and also found a positive correlation between increased mean F0 and SAD diagnostic severity. This finding was replicated in a subsequent study in a sample of men with social phobia [73].

Scherer et al. [23] investigated voice quality features as indicators of PTSD and depression using SVM RBF kernel classifier during interactions with a virtual human. Their classification results for PTSD depended on the emotional polarity, with accuracies of 52.38% for speech passages with negative affective polarity, and of 72.09% for passages with neutral polarity, with no significant difference between genders.

From the related works presented, an effort can be seen to relate the decisions of classifiers to support the diagnosis of various mental disorders. All of the cited works use characteristics that are already well established in speech therapy studies to represent the patients' audio samples. However, we believe that, despite the use of well-established attributes, classical attributes for audio representation in phonological studies may not be sufficient to support the diagnosis. In our study, we added statistical, temporal and time-frequency attributes commonly used in the analysis of other biomedical signals.

In addition, it is important to highlight that the works discussed present results obtained in a laboratory environment, with a high level of control over possible noise and interference on the patient's audio. In the present study, data were obtained in an emergency care environment at a hospital in the public health system in Brazil, under real conditions, with exposure to noise and interference typical of the context of care in a psychiatric emergency situation.

Psychiatric diagnosis lacks biomarkers when comparing Psychiatry to other areas of Medicine and the health sciences in general. We believe that machine learning can be used to build digital biomarkers, capable of helping specialists in the investigative process through metrics with good reliability.

## 2. Materials And Methods

# Participant Selection

This study, entitled "Digital voice processing as a diagnostic aid tool for mental disorders", was approved by the Research Ethics Committee of the Hospital das Clínicas, Federal University of Pernambuco, Recife, Pernambuco, Brazil, under registration 19422619.2.0000.8807, report 3.565.104. The research only involves individuals over 18 years of age and, therefore, parental or guardian consent is not required.

For this study 78 volunteers from both genders aged over 18 years old were recruited into one of the five following diagnostic groups:

- Control group: 12 healthy participants (7 males) were selected through the Self-Reporting Questionnaire (SRQ-20), a screening for common mental disorders [74];
- Depression group: 28 patients with major depressive disorder (17 males), in conformity with Hamilton Depression Rating Scale – HAM-D 17 [75];
- Schizophrenia group: 21 patients with diagnosis of schizophrenia (12 males) were assessed through the Brief Psychiatric Rating Scale (BPRS), one of the most widely used instruments for evaluation of symptom severity within schizophrenia [76, 77];
- Bipolar disorder group: 14 patients with bipolar disorder, in current mania or hypomania episode diagnosed by an independent clinician, with symptom severity assessed by the Portuguese version of Young Mania Rating Scale (YMRS) [78, 79];
- Generalized anxiety disorder (GAD) group: four patients diagnosed with GAD, with symptoms assessed by GAD-7 Scale [80, 81].

All individuals from the disease groups fulfilled DSM-5 diagnostic criteria for their respective disorders and were diagnosed by an independent clinician prior to this study. Data for these groups were collected in an outpatient setting and in psychiatric wards 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. The use of validated psychometric scales aimed to verify previous diagnostic consistency and evaluate symptom severity. The diagnoses were considered ground-truth due to the large clinical experience of the specialists involved in this research, but there is still an uncertainty due to the diagnosis by a human specialist that cannot be measured. This study was conducted only after approval of a local Research Ethical Board, and all participants have given written consent. Table 1 summarizes mean age and mean scale scores for all groups.

For control group the SRQ-20 cutoff score was 6/7 [82], and for depression group the eligibility criterion was HAM-D 17 score above 7. Consequently, patients suffering from mild to severe depression were selected, and a mean HAM-D 17 score of 19.32 indicates moderate depression [83]. For schizophrenia group, participants with prior diagnosis were included, irrespective of their score in the BPRS; an average score of 45.16 found in this group corresponds to moderate illness severity [76]. Likewise, patients from BD group were selected regardless of their YMRS score; their average YMRS score was 23.00, which corresponds to ‘severely ill’ [84]. Unfortunately, during the data acquisition phase we were able to recruit only four patients with GAD; their mean GAD-7 score was 13.75, which corresponds to moderate disease [81]. Table 1 provides a summary of mean age and scale scores for each group.

Table 1
Mean age and rating scale scores for all groups

| Group | Age (years) (SD) | Rating scale | Avg. score (points) (SD) |
|---|---|---|---|
| Control | 29.2 (± 12.4) | SRQ-20 | 3.00 (± 1.86) |
| Major depressive disorder | 42.0 (± 12.4) | HAM-D 17 | 19.32 (± 7.36) |
| Schizophrenia | 36.0 (± 11.3) | BPRS | 45.16 (± 11.25) |
| Bipolar disorder | 40.5 (± 8.0) | YMRS | 23.00 (± 11.95) |
| Generalized anxiety disorder | 25.8 (± 8.5) | GAD-7 | 13.75 (± 2.22) |

Notes: BPRS: Brief Psychiatric Rating Scale; YMRS: Young Mania Rating Scale; GAD-7: Generalized Anxiety Disorder-7 Scale; HAM-D 17: Hamilton Depression Scale; SD: standard deviation; SRQ-20: Self-Reporting Questionnaire.

# Acquisition Of Voice Samples

Audio recordings were made with a Tascam™ 16-bit linear PCM recorder, at 44.1KHz sampling rate, in WAV format, without compressions, using environment noise cancellation. No time limit was set for any recording. Voice acquisitions were made during an interview with a psychiatrist in naturalistic settings, i.e., patients from the disease groups were recorded during a routine medical evaluation in an outpatient office or a hospital ward. After each interview, the interviewer applied an appropriate rating scale to assess symptom severity. Exceptions were GAD and control groups, since GAD-7 scale and SRQ-20 are self-applied; consequently, participants from these groups were required to answer it after the interview. Recordings for the control group were acquired in different environments, specifically an office, a classroom, and a gym, but using the same recorder's noise cancellation function. Therefore, the acquired signals in all environments are similar regarding environment noise cancellation residuals. As conversations were thoroughly recorded, clinician's and potential third parties' speech were also acquired and needed to be further removed. We obtained one record per subject. The total duration of recordings for all groups was 980.3 minutes (16.3 hours).

It is important to notice that, in order to acquire data in accordance with the clinical practice and differently from the works of the state-of-the-art, all audio data was recorded in a naturalistic environment: the psychiatric emergency care services of Hospital Ulysses Pernambucano and Hospital das Clínicas. All data were recorded under similar real conditions, including interferences and noise. We considered the background noise in these non-clinical environments slightly similar to the one present at the psychiatric emergency health service, with the sounds of corridors (human steps and voices, for instance) and air-conditioning noise as the most common noise present at the recordings. However, these types of noise were minimized by the use of a professional audio recorder with environment noise cancellation hardware provided by the device arrangement based on a pair of unidirectional microphones positioned in opposition, with a 90 degrees angle between them. This arrangement is able to cancel or, at least, minimize the environment noise, common to both audio sources, while the main information of the two audio sources slightly out of phase is emphasized. Therefore, no additional audio preprocessing by software was performed to minimize this background noise. The set of all participants with a positive diagnosis for a mental disorder seen in the emergency condition is approximately equally distributed between the two health services. This was done so that there would be no classification bias due to different signal acquisition environments, despite the minimization of ambient noise performed by the digital audio recorder.

# Audio Editing

After voice acquisition, Audacity™ audio software was used to remove recorded audio from the interviewer and any potential companion. The edition process was manually made, and yielded 591 minutes (9.85 hours) of recorded audio from participants as follows: 100.7 minutes for control group; 222.6 minutes for MDD group; 125.7 minutes for schizophrenia group, 102 minutes for BD group, and 40 minutes for GAD group. Thus, all voices other than the voices of volunteers were removed. In this way, for each volunteer, the recordings contain only his/her voice. Table 2 provides detailed information about recording duration for all groups.

Table 2
Recording duration after audio editing

| Group | Number of participants | Total recording duration | Avg. recording duration (SD) |
|---|---|---|---|
| Control | 12 | 6039s (100.7 min) | 503.3s (8.4 min) ± 159.0s |
| Major depressive disorder | 28 | 13355s (222.6 min) | 477.0s (≅ 8.0 min) ± 203.0s |
| Schizophrenia | 20 | 7541s (125.7 min) | 377.1s (6.3 min) ± 270.4s |
| Bipolar disorder | 14 | 6122s (102.0 min) | 437.3s (7.3 min) ± 253.9s |
| Generalized anxiety disorder | 4 | 2401 s (40.0 min) | 600.3s (10.0 min) ± 194.8s |

# Feature Extraction

Edited audio recordings were submitted to vocal feature extraction on GNU Octave™, a free open-source signal-processing software. Windowing was made with rectangular windows and frame length of 10s with 50% overlap. As raw audio data was used, no filtering process was applied. Consequently, background noise was captured as well. However, with the selected attributes, we believe this noise is not able to interfere significantly due to its homogeneous spectral behavior. At this phase, the following 33 features were extracted: 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. Table 3 presents the detailed mathematical description of the 33 characteristics applied to the description of the audio windows.

The duration of the 10s-windows was defined by the research team empirically, considering a scenario for the use of a future application to support the clinical diagnosis of mental disorders in a psychiatric emergency context. A window of only 10s would allow the specialist to obtain diagnostic indications

during the patient's interview, and then make the decision based on the most frequent indication, also considering the discourse analysis and the anamnesis process as a whole.

The choice of the above features relies on their accurate representation of input signals to computational models, because decision making process in machine learning does not depend on human interpretation. Furthermore, these features have already been successfully used for representing other signal types, such as electroencephalography in rectangular windows, which comprises much more spectral complexity than audio signals.

TABLE 3. Detailed mathematical description of the 33 characteristics applied to the description of the audio windows

| Attribute | Mathematical expression | Attribute | |
|---|---|---|---|
| Mean (μ) | $\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$ | Zero crossings | $ZC = \sum_{n=1}^{N-1}[sgn(x_n \times x_{n+1}) \cap \lvert x_n - x_{n+1}\rvert \geq threshold]$ <br> $sgn(x) = \begin{cases} 1, if\ x \geq threshold \\ 0, \quad otherwise \end{cases}$ |
| Variance | $var = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \mu)^2$ | Slope Sign Change | $SSC = \sum_{n=1}^{N-1}[f(x_n - x_{n-1}) \times (x_n - x_{n+1})]]$ <br> $f(x) = \begin{cases} 1, if\ x \geq threshold \\ 0, \quad otherwise \end{cases}$ |
| Standard deviation (σ) | $\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\lvert x_n - \mu\rvert^2}$ | Hjorth parameter activity | $Hjorth_{activity} = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \mu)^2$ |
| Root mean square | $RMS = \sqrt{\frac{\sum_{n=1}^{N}(x_n)^2}{N}}$ | Hjorth parameter mobility | $Hjorth_{mobility} = \sqrt{\frac{var\left(\frac{d\,x(t)}{dt}\right)}{var(x(t))}}$ |
| Average Amplitude Change | $AAC = \frac{1}{N}\left(\sum_{n=1}^{N}\left\lvert\frac{d\,x(t)}{dt}\right\rvert\right)$ | Hjorth parameter complexity | $Hjorth_{complexity} = \frac{Hjorth_{mobility}\left(\frac{d\,x(t)}{dt}\right)}{Hjorth_{mobility}(x(t))}$ |
| Difference Absolute Deviation | $DASDV = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{d\,x(t)}{dt}\right)^2}$ | Mean frequency | $MNF = \frac{\sum_{j=1}^{M} f_j P_j}{\sum_{j=1}^{M} P_j}$ <br> Where $f_j$,$P_j$ are the frequencies and power of the spectrum, respectively, and $M$ is the length of the frequencies |
| Integrated Absolute Value | $\mathrm{IAV} = \sum_{n=1}^{N} x_n$ | Median frequency | $MDF = \frac{1}{2}\sum_{j=1}^{M} P_j$ |
| Logarithm Detector | $LOGD = e^{(\frac{1}{N}\sum_{n=1}^{N}\log(\lvert x_n\rvert))}$ | Mean power | $MNP = \sum_{j=1}^{M}\frac{P_j}{M}$ |
| Simple Square Integral | $\mathrm{SSI} = \sum_{n=1}^{N} {x_n}^2$ | Peak frequency | $PKF = \max(P_j)$ |
| Mean Absolute Value | $\mathrm{MAV} = \frac{1}{N}\sum_{n=1}^{N}\lvert x_n\rvert$ | Power Spectrum ratio | $PSR = \frac{PKF}{\sum_{j=1}^{M} P_j}$ |
| Mean Logarithm Kernel | $\mathrm{MLOGK} = \frac{1}{N}\left\lvert\sum_{n=1}^{N} x_n\right\rvert$ | Total Power | $TP = \sum_{j=1}^{M} P_j$ |
| Skewness (s) | $s = \frac{\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^3}{\sigma^3}$ | First Spectral Moment | $SM1 = \sum_{j=1}^{M} f_j P_j$ |
| Kurtosis | $\mathrm{kurt} = \frac{\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^4}{\sigma^4}$ | Second Spectral Moment | $SM2 = \sum_{j=1}^{M} {f_j}^2 P_j$ |
| Maximum Amplitude | $MAX = \max(x_n)$ | Third Spectral Moment | $SM3 = \sum_{j=1}^{M} {f_j}^3 P_j$ |
| Third Moment | $M3 = \left\lvert\frac{1}{N}\sum_{n=1}^{N}(x_n)^3\right\rvert$ | Variance of Central Frequency | $VCF = \frac{SM2}{TP} - \left(\frac{SM1}{TP}\right)^2$ |
| Fourth Moment | $M4 = \left\lvert\frac{1}{N}\sum_{n=1}^{N}(x_n)^4\right\rvert$ | Waveform length | $WL = \sum_{n=1}^{N-1}\lvert x_{n+1} - x_n\rvert$ |
| Fifth Moment | $M5 = \left\lvert\frac{1}{N}\sum_{n=1}^{N}(x_n)^5\right\rvert$ | Shannon Entropy | $S = \sum_{i} s_i^2 \log(s_i^2)$ |

# Classification

To find the best classification model, we used an approach based on a training/testing step and a validation step. The training/test set was built from a random sample of 10% of the database instances, so that, in each class and for each attribute, the statistical behavior was the same as in the original data set. This was ensured by comparing the histograms of each attribute in each class and the mean and

standard deviation descriptors. Sampling was considered sufficient when, qualitatively, the histograms of the attributes in each sample class approached or coincided with those of the original data set, considered as the statistical population. Similarly, sampling was considered sufficient when, quantitatively, the confidence intervals for each attribute in each class in the sample coincided with those in the original dataset. This training/test set, designed as an approximation of the original dataset, was used to investigate what would be the best classification model for the problem. The training/test set was balanced by the SMOTE algorithm of insertion of synthetic instances and subjected to tests using cross validation. Once the best classification model was found, this classifier was trained with the training/test set and tested with the validation set. As with the training/testing set, the validation set is designed to be an approximation of the original dataset, considered, for all intents and purposes, as the statistical population. However, the same number of instances per class (300 instances) was selected for the validation set, to keep this set balanced. These instances were selected in such a way that the validation and training and testing sets are totally disjoint with respect to the instances: there are no repeated instances between the sets. In addition, there are also no repeated volunteers in the two sets. Still, both the training/test set and the validation set have the same statistical behavior as the original dataset, as expected from a statistical learning approach.

To investigate the best machine learning model, all classes were balanced through the addition of synthetic instances by using the algorithm SMOTE, Synthetic Minority Oversampling Technique [85, 86]. SMOTE is an oversampling technique where random synthetic feature vectors are generated for the minority class, overcoming the overfitting problem [85, 86]. Using the k-nearest neighbors algorithm, for a randomly selected instance in a given class, the k nearest neighbors are selected. Then, a synthetic instance is generated by the interpolation among these k selected instances [85, 86]. The original dataset was composed by 1175 windows instances for control, 2640 instances for major depressive disorder, 1483 instances for schizophrenia, 1206 instances for bipolar disorder, and 475 instances for generalized anxiety disorder. The SMOTE algorithm was applied to an unbalanced dataset selected as a 10% sample of the 6979 instances original dataset, generating a dataset with 1320 instances, 264 instances for each class. This sample was obtained preserving the statistical behavior of the original dataset, evaluated by the similarities between feature statistics (mean and standard deviation) and histograms in each class compared to its previous descriptive statistics in the original dataset.

Class balancing was crucial to prevent computational biases towards the classes with more representativeness, in this case depression and schizophrenia classes. Experiments were performed using 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). Investigation experiments were performed using 30-run 10-fold cross-validation. In the distribution of instances in folds for cross validation, it is guaranteed that there is no repetition of instances in the training and test sets, given that the distribution in folds is without instance replacement. Figure 1 summarizes the steps of data collection and our proposed solution.

Afterwards, considering the best machine learning model found with the oversampled balanced dataset, we created two balanced, disjoint and statistically similar sets, to be used as training and test set at the validation stage. In each set, we selected 300 instances for each class, trying to preserve the statistical behavior for each feature in each class regarding the original dataset.

## 3. Results

In this section, we present the results for the three stages of our research: (a) investigation of the best machine learning model based on a 10% sample of the dataset. This sample was balanced by oversampling, inserting synthetic instances generated by the SMOTE method, in order to optimize the training conditions for all evaluated models. (b) feature selection by evolutionary computing and bioinspired meta-heuristic optimization methods, showing that just one feature is redundant. (c) validation: the best machine learning model found in stage (a) is evaluated considering balanced training and test sets constructed by undersampling, by randomly selecting 300 instances in each class, for each set, trying to get two statistically similar sets. These results are presented as quality metrics and a confusion matrix corresponding to a one-shot learning.

# a. Best Model Investigation

Experiments were initially performed under default settings on Weka™. Subsequently, different setups for all algorithms with adjustable settings were tested (MLP; polynomial kernel and normalized polynomial kernel SVM, SVM PUK kernel, and random forest). Table 3 describes in detail our best results for each machine learning model. These results are highlighted. We tested the following configurations:

- Decision tree J48;
- Random Tree;
- Random Forests, for 50, 100, 200, and 300 trees;
- Bayesian Network;
- Naïve Bayes' classifier;
- Multi-layer Perceptron (MLP): one hidden layer with 20, 50, 100, 150, and 200 neurons;
- Support Vector Machine (SVM): linear kernel; 2- and 3-degree polynomial kernels (p = 2, 3); Pearson Universal VII (PUK) kernel; Radial Basis Function (RBF) kernel, (G = 0.01, 0.25 and 0.50). All SVM configurations were evaluated for C = 0.1, 1.0 and 10.0.

The results above show that classification performances varied significantly according to which machine learning algorithm was used, with Bayesian Network, Naïve Bayes and Random Forest (300 trees) achieving the highest discrimination accuracy of 91.29% ± 2.38, 92.12% ± 2.12 and 90.01% ± 2.42 (sensitivity: 0.8566 ± 0.0694, 0.8777 ± 0.0607 and 0.8630 ± 0.0627; specificity: 0.9616 ± 0.0185, 0.9754 ±

0.0138 and 0.9561 ± 0.0193), and kappa coefficient of 0.8911 ± 0.0298, 0.9014 ± 0.0265 and 0.8751 ± 0.0303, respectively. The best Multi-layer Perceptron configuration employed 200 neurons in the hidden layer, obtaining the following metrics for accuracy, kappa, sensitivity, and specificity: 79.81% ± 3.58, 0.7476 ± 0.0448, 0.6939 ± 0.1020, and 0.9223 ± 0.0348, in this order. The best SVM configuration was the RBF kernel, with G = 0.5 and C = 10.0, achieving the following results for accuracy, kappa, sensitivity, and specificity: 84.50% ± 3.11, 0.8062 ± 0.0388, 0.7719 ± 0.0772, and 0.9423 ± 0.0240, respectively. Since the best results, i.e. Bayesian Network, Naïve Bayes and 300-tree Random Forest, are statistically similar for the three best models, we have chosen the 300-tree Random Forest. This decision was supported by the fact that Random Forests do not assume normality over the data. Furthermore, since Random Forests are based on committees of decision trees, they tend to be more robust to noise and other outliers. Random Forest also tend to present larger generalization capacity. The confusion matrix for a random run of this model is shown in Table 4 below.

TABLE 3. Classification performance for machine learning algorithms for all groups

| Classifier | Accuracy (%) | Kappa | Sensitivity | Specificity |
|---|---|---|---|---|
| Decision Tree J48 | 78.37 ± 3.26 | 0.7296 ± 0.0408 | 0.6767 ± 0.0912 | 0.9166 ± 0.0275 |
| Random Tree | 71.20 ± 4.11 | 0.6400 ± 0.0514 | 0.6286 ± 0.1020 | 0.8980 ± 0.0301 |
| Bayes Net | 91.29 ± 2.38 | 0.8911 ± 0.0298 | 0.8566 ± 0.0694 | 0.9616 ± 0.0185 |
| Naïve Bayes | 92.12 ± 2.12 | 0.9014 ± 0.0265 | 0.8777 ± 0.0607 | 0.9754 ± 0.0138 |
| Random Forest, 50 trees | 88.91 ± 2.63 | 0.8613 ± 0.0329 | 0.8422 ± 0.0690 | 0.9524 ± 0.0203 |
| Random Forest, 100 trees | 89.51 ± 2.60 | 0.8688 ± 0.0325 | 0.8514 ± 0.0668 | 0.9539 ± 0.0198 |
| Random Forest, 200 trees | 89.90 ± 2.48 | 0.8738 ± 0.0310 | 0.8619 ± 0.0636 | 0.9547 ± 0.0202 |
| Random Forest, 300 trees | 90.01 ± 2.42 | 0.8751 ± 0.0303 | 0.8630 ± 0.0627 | 0.9561 ± 0.0193 |
| MLP, 20 neurons | 77.55 ± 3.51 | 0.7193 ± 0.0439 | 0.6481 ± 0.1113 | 0.9139 ± 0.0369 |
| MLP, 50 neurons | 78.89 ± 3.58 | 0.7361 ± 0.0447 | 0.6744 ± 0.1082 | 0.9210 ± 0.0361 |
| MLP, 100 neurons | 79.38 ± 3.48 | 0.7422 ± 0.0435 | 0.6905 ± 0.1013 | 0.9211 ± 0.0338 |
| MLP, 150 neurons | 79.81 ± 3.47 | 0.7476 ± 0.0434 | 0.6970 ± 0.0992 | 0.9237 ± 0.0346 |
| MLP, 200 neurons | 79.81 ± 3.58 | 0.7476 ± 0.0448 | 0.6939 ± 0.1020 | 0.9223 ± 0.0348 |
| SVM, linear, C = 0.1 | 68.83 ± 3.50 | 0.6103 ± 0.0437 | 0.4796 ± 0.0935 | 0.9158 ± 0.0236 |
| SVM, linear, C = 1.0 | 80.02 ± 3.27 | 0.7502 ± 0.0408 | 0.6866 ± 0.0818 | 0.9287 ± 0.0252 |
| SVM, linear, C = 10.0 | 82.06 ± 3.36 | 0.7757 ± 0.0420 | 0.7166 ± 0.0865 | 0.9327 ± 0.0258 |
| SVM, polynomial, p = 2, C = 0.1 | 76.71 ± 3.42 | 0.7089 ± 0.0428 | 0.6551 ± 0.0866 | 0.9146 ± 0.0279 |
| SVM, polynomial, p = 2, C = 1.0 | 82.51 ± 3.24 | 0.7814 ± 0.0405 | 0.7166 ± 0.0838 | 0.9384 ± 0.0236 |
| SVM, polynomial, p = 2, C = 10.0 | 83.97 ± 3.31 | 0.7997 ± 0.0413 | 0.7638 ± 0.0771 | 0.9418 ± 0.0246 |
| SVM, polynomial, p = 3, C = 0.1 | 80.04 ± 3.34 | 0.7505 ± 0.0417 | 0.6937 ± 0.0808 | 0.9269 ± 0.0254 |
| SVM, polynomial, p = 3, C = 1.0 | 83.53 ± 3.24 | 0.7941 ± 0.0405 | 0.7549 ± 0.0798 | 0.9364 ± 0.0248 |
| SVM, polynomial, p = 3, C = 10.0 | 82.83 ± 3.26 | 0.7854 ± 0.0408 | 0.7721 ± 0.0767 | 0.9366 ± 0.0257 |
| SVM, PUK, C = 0.1 | 76.78 ± 3.39 | 0.7097 ± 0.0424 | 0.5676 ± 0.0916 | 0.9408 ± 0.0229 |
| SVM, PUK, C = 1.0 | 83.28 ± 3.20 | 0.7910 ± 0.0400 | 0.7369 ± 0.0837 | 0.9384 ± 0.0240 |
| SVM, PUK, C = 10.0 | 83.79 ± 3.15 | 0.7973 ± 0.0394 | 0.7720 ± 0.0814 | 0.9388 ± 0.0232 |
| SVM, RBF, G = 0.01, C = 0.1 | 33.96 ± 8.31 | 0.1776 ± 0.1035 | 0.3622 ± 0.3710 | 0.7859 ± 0.2773 |
| SVM, RBF, G = 0.01, C = 1.0 | 45.62 ± 5.09 | 0.3222 ± 0.0634 | 0.4040 ± 0.4240 | 0.7516 ± 0.2924 |
| SVM, RBF, G = 0.01, C = 10.0 | 73.15 ± 3.65 | 0.6643 ± 0.0456 | 0.5877 ± 0.0878 | 0.9147 ± 0.0262 |
| SVM, RBF, G = 0.25, C = 0.1 | 60.25 ± 4.46 | 0.5039 ± 0.0554 | 0.4396 ± 0.3101 | 0.8504 ± 0.1335 |
| SVM, RBF, G = 0.25, C = 1.0 | 79.04 ± 3.29 | 0.7380 ± 0.0411 | 0.6610 ± 0.0849 | 0.9292 ± 0.0243 |
| SVM, RBF, G = 0.25, C = 10.0 | 83.17 ± 3.23 | 0.7964 ± 0.0404 | 0.7296 ± 0.0808 | 0.9430 ± 0.0234 |
| SVM, RBF, G = 0.5, C = 0.1 | 69.66 ± 3.60 | 0.6207 ± 0.0450 | 0.4817 ± 0.0929 | 0.9123 ± 0.0245 |
| SVM, RBF, G = 0.5, C = 1.0 | 81.30 ± 3.34 | 0.7662 ± 0.0417 | 0.6960 ± 0.0862 | 0.9357 ± 0.0229 |
| SVM, RBF, G = 0.5, C = 10.0 | 84.50 ± 3.11 | 0.8062 ± 0.0388 | 0.7719 ± 0.0772 | 0.9423 ± 0.0240 |

Notes: MLP: Multilayer Perceptron; PUK: Pearson Universal VII Kernel; RBF: Radial Basis Function; SVM: Support Vector Machines.

Table 4
Confusion matrix for the model with the highest performance, Random Forest with 300 trees, considering the balanced oversampled dataset

| | Classified as control | Classified as major depressive disorder | Classified as schizophrenia | Classified as bipolar disorder | Classified as generalized anxiety disorder |
|---|---|---|---|---|---|
| Control | **2361 (89.43%)** | 25 (0.95%) | 74 (2.80%) | 0 (0.00%) | 180 (6.82%) |
| Major depressive disorder | 58 (2.20%) | **2342 (88.71%)** | 222 (8.41%) | 0 (0.00%) | 18 (0.68%) |
| Schizophrenia | 146 (5.53%) | 67 (2.54%) | **2362 (89.47%)** | 1 (0.04%) | 64 (2.42%) |
| Bipolar disorder | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | **2640 (100%)** | 0 (0.00%) |
| Generalized anxiety disorder | 138 (5.23%) | 2 (0.08%) | 45 (1.70%) | 0 (0.00%) | **2455 (92.99%)** |

Our Random Forest best model correctly classified 88.71% of time-windows related to depressed patients, and 89.43% of time-windows of healthy controls; discrimination rates between schizophrenia and GAD groups were high, with accuracies of 89.47% and 92.99%, respectively. The highest performance was achieved in BD group, with 100% classification accuracy. Higher confusion rates were observed between depression and schizophrenia, with 8.41% time-samples of depressed patients classified as schizophrenia; and between GAD and control, with 5.23% GAD time-samples classified as control and 6.82% control time-samples classified as GAD.

# b. Feature Relevance Investigation

To evaluate the relevance of the features selected to represent the 10s-window samples, we used meta-heuristic optimization methods. Artificial populations of 50 individuals were used, evolving in 50 generations. As an objective function, we used a decision tree, trained and tested using 10-fold cross-validation. Each individual represents the attributes used in the classification by means of a binary vector, where “1” models the presence of that attribute, whilst “0” represents its opposite. We employed meta-heuristic libraries developed in Java for Weka data mining platform [87]. We adopted the following feature selection methods:

1. Evolutionary Search, with crossover probability of 0.6, mutation probability of 0.1, bit-flip mutation, random initialization, generational replacement operator, report frequency of 20, survivor selection by tournament [88, 89];

2. Particle Swarm Optimization, individual weight of 0.34, inertia weight of 0.33, mutation probability of 0.01, report frequency of 20, social weight of 0.33 [90, 91, 92];

3. Ant Colony Search, with chaotic coefficient of 4.0, chaotic type of logistic map, evaporation of 0.9, heuristic of 0.7, bit-flip mutation, mutation probability of 0.01 [93, 94].

# C. Generalization Evaluation

In this stage, we evaluated the generalization capacity of the best model found in the model investigation stage. In Table 5 we present the validation results, considering training and test sets with a total of 1500 instances, i.e. 300 instances for each class. We reached a validation accuracy of 75.2667% and kappa index of 0.6908. Although this result is inferior to the one obtained in the context of the oversampled dataset, it is realistic and can be considered clinically interesting for our scope. For the control group, we obtained 0.713 for sensitivity, 0.925 for specificity, and 0.940 for area under ROC curve. For major depressive disorder, we obtained 0.600 for sensitivity, 0.957 for specificity, and 0.905 for area under ROC curve. For schizophrenia, we got 0.700 for sensitivity, 0.913 for specificity, and 0.929 for area under ROC curve. For bipolar disorder, we have 0.830 for sensitivity, 0.952 for specificity, and 0.966 for area under ROC curve. Taking into account the generalized anxiety disorder, we reached 0.920 for sensitivity, 0.943 for specificity, and 0.985 for area under ROC curve. Considering the control group and all mental disorders investigated in this work, we have weighted sensitivity of 0.753, weighted specificity of 0.938, and weighted area under ROC curve of 0.945.

Table 5
Confusion matrix for the Random Forest with 300 trees, considering the balanced undersampled validation dataset

| | Classified as control | Classified as major depressive disorder | Classified as schizophrenia | Classified as bipolar disorder | Classified as generalized anxiety disorder |
|---|---|---|---|---|---|
| Control | **214 (71.33%)** | 16 (5.33%) | 32 (10.67%) | 9 (3.00%) | 29 (9.67%) |
| Major depressive disorder | 42 (14.00%) | **180 (60.00%)** | 42 (14.00%) | 14 (4.67%) | 22 (7.33%) |
| Schizophrenia | 29 (9.67%) | 23 (7.67%) | **210 (70.00%)** | 30 (10.00%) | 8 (2.67%) |
| Bipolar disorder | 6 (2.00%) | 10 (3.33%) | 26 (8.67%) | **249 (83.00%)** | 9 (3.00%) |
| Generalized anxiety disorder | 13 (4.33%) | 3 (1.00%) | 4 (1.33%) | 4 (1.33%) | **276 (92.00%)** |

# 4. Discussion

Through analysis of the confusion matrix of Table 5, we notice that classification accuracies for depression, schizophrenia, and control groups varied, but can be considered good for a clinical diagnosis (accuracies of 60% and 70% only considering audio features). However, discrimination rates for GAD and BD were considerably higher. We hypothesize this is because patients from this groups have severe symptoms and probably more intense changes in vocal patterns, whereas patients from other groups are moderately ill, and therefore less intense vocal abnormalities. Another hypothesis is that changes in acoustic parameters in bipolar patients, such as increased pitch variability and increased intensity/volume, may be ‘unique’ to this group, and consequently more easily distinguishable. In contrast, some vocal characteristics are shared by other diagnostic groups; for example, reduced pitch range is found both in depression and schizophrenia, and may be a confounding factor for automated classifiers.

Greater confusion between depression and control classes was also seen, and we believe this misclassification is due to the inclusion of individuals with mild depressive disorder in the depression group. As mild depressive symptoms may produce only subtle changes in vocal parameters in comparison to healthy individuals, their recognition becomes more difficult. Additionally, slightly lesser confusion rates were seen among GAD and control samples, and our hypothesis is that generalized anxiety symptoms may not be as impactful on our selected vocal features as other disorders. However, the small number of participants in this group is an important limitation to our analysis.

Other important limitations to this study are the lack of demographic control among groups, once variables such as age and education may influence acoustic properties. Smoking history, which is known to affect a person's voice, and pharmacotherapy were not controlled either and may also limit our findings.

We assume that a larger number of categories increases data complexity, which tends to decrease classification performances. Even in this scenario, our results outperform those from most previous studies, which were carried out using simple binary classification. This shows the strength of this proposed solution for the identification of several mental disorders at the same time. To the best of our knowledge, the relevance of this work is unprecedent, as no previous study has used automated classifiers for the simultaneous classification of several mental disorders, and no other study has encompassed this number of disorders before. This high discrimination power supports the prospective use of machine learning models as diagnostic (and even screening) tools for mental disorders.

Regarding feature selection using evolutionary search, ant colony search and particle swarm optimization, these three meta-heuristics based feature selection methods agreed that 32 of the employed 33 features were statistically relevant for the classification: the Average Amplitude Change (AAC) returned a probability of relevance of 80%; to the Integrated Absolute Value (IAV) was associated a relevance of 90%; the Simple Square Integral (SSI), however, was considered absolutely irrelevant: 0%; all other features returned a relevance of 100%. After removing SSI, classification results remained the same. This indicates that not just SSI is not relevant to the classification, but using additional features not usually adopted as audio features contributes to get good classification results.

# 5. Conclusion And Future Works

Current psychiatric diagnosis still lacks objective biomarkers, and relies mostly on diagnostic criteria. These show no correlation with neurobiology and etiopathogenesis of mental disorders, leading to trial-and-error treatments and prolonging patient suffering and disability. Despite the prevalence and relevance of mental disorders, access to mental health services is problematic due to lack of professionals, difficulty in seeking help, and the stigma of mental disorders. These issues can be addressed with the development of accessible objective biomarkers, which can be achieved with the use of vocal acoustic features. As changes in vocal patterns have been reported in several mental disorders and appear to correlate with illness severity, vocal acoustic features have shown to be promising markers, with the advantages of being abundant, inexpensive, non-invasive and remotely performed. In this study, we performed classification experiments in an unprecedent manner using machine learning algorithms with four mental disorders and a healthy control group. The results provided by our proposed solution are very promising and outperform most from previous studies with binary classes. These evidences support the feasibility of a computational method based on vocal parameters for assisting clinicians in patient triage and diagnosis in psychiatry. In future studies we intend to perform the same experiments in a larger sample and with a greater number of mental disorders.

# Declarations


### Acknowledgements

The authors are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq-Brazil, for the partial support of this research.


### Compliance with Ethical Standards


This study was funded by the Brazilian research agency CNPq (grant 314896/2018-0).


### Ethical Approval

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. The study was approved by the Research Ethics Committee of the Hospital das Clínicas, Federal University of Pernambuco, Recife, Pernambuco, Brazil, under registration 19422619.2.0000.8807, report 3.565.104.

### Conflict of Interest

All authors declare they have no conflicts of interest.

# References


[1] Schalper, K. A., Brown, J., Carvajal-Hausdorf, D., McLaughlin, J., Velcheti, V., Syrigos, K. N., … Rimm, D. L. (2015). Objective Measurement and Clinical Significance of TILs in Non–Small Cell Lung Cancer. *JNCI: Journal of the National Cancer Institute*, *107*(3), dju435.

[2] Zhao, Y., Zhu, L., Liu, L., Shi, S., Lv, J., & Zhang, H. (2016). Measures of Urinary Protein and Albumin in the Prediction of Progression of IgA Nephropathy. *CJASN*, *11*(6), 947–955.

[3] Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., … Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *Nature Partner Journals*. https://doi.org/10.1038/npjschz.2015.30

[4] Hirschtritt, M., & Insel, T. (2018). Digital Technologies in Psychiatry: Present and Future. *Focus*, *16*(3), 251–258. https://doi.org/10.1176/appi.focus.20180001

[5] Baca-Garcia, E., Perez-Rodriguez, M. M., Basurte-Villamor, I., Fernandez Del Moral, A. L., Jimenez-Arriero, M. A., Gonzalez De Rivera, J. L., … Oquendo, M. A. (2007). Diagnostic stability of psychiatric disorders in clinical practice. *British Journal of Psychiatry*, *190*(MAR.), 210–216. https://doi.org/10.1192/bjp.bp.106.024026

[6] 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

[7] Petzschner, F. H., Weber, L. A. E., Gard, T., & Stephan, K. E. (2017). Review Computational Psychosomatics and Computational Psychiatry : Toward a Joint Framework for Differential Diagnosis. *Biological Psychiatry*, 1–10. https://doi.org/10.1016/j.biopsych.2017.05.012

[8] Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics*, *20*, 50–64. https://doi.org/10.1016/j.jneuroling.2006.04.001

[9] Jiang, H., Hu, B., Liu, Z., Wang, G., Zhang, L., Li, X., & Kang, H. (2018). Detecting Depression Using an Ensemble Logistic Regression Model Based on Multiple Speech Features. *Computational and Mathematical Methods in Medicine*, *2018*. https://doi.org/10.1155/2018/6508319

[10] 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.v11s3.30042

[11] Dwyer, D., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu. Rev. Clin. Psychol.*, *14*(January), 1–28.

[12] Alberto, P., Arndis, S., Vibeke, B., & Riccardo, F. (2019). Voice Patterns in Schizophrenia: A systematic Review and Bayesian Meta-Analysis. *VOICE IN SCHIZOPHRENIA: REVIEW AND META-ANALYSIS*, (1–40).

[13] Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., S., & M., ... & Kerbrat, A. (2018). Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Scientific Reports*, *8*(1), 1–17.

[14] dos Santos, W. P., De Assis, F. M., De Souza, R. E., Mendes, P. B., & De Souza Monteiro, H. S., Alves, H. D. (2009). A Dialectical Method to Classify Alzheimer's Magnetic Resonance Images. *Evolutionary Computation*, 473.

[15] 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).

[16] 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).

[17] Azevedo, W. W., Lima, S. M., Fernandes, I. M., Rocha, A. D., Cordeiro, F. R., da Silva-Filho, A. G., & dos Santos, W. P. (2015). Fuzzy morphological extreme learning machines to detect and classify masses in mammograms. In *2015 IEEE international conference on fuzzy systems (fuzz-IEEE)* (pp. 1–8).

[18] Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., … Furmark, T. (2008). In a Nervous Voice: Acoustic Analysis and Perception of Anxiety in Social Phobics' Speech. *J Nonverbal Behav*, *32*, 195–214. https://doi.org/10.1007/s10919-008-0055-9

[19] Karam, Z. N., Provost, E. M., Singh, S., Montgomery, J., Archer, C., Harrington, G., & Mcinnis, M. G. (2014). ECOLOGICALLY VALID LONG-TERM MOOD MONITORING OF INDIVIDUALS WITH BIPOLAR DISORDER USING SPEECH. *2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 4858–4862.

[20] Larsen, M. E., Cummins, N., Boonstra, T. W., O'Dea, B., Tighe, J., Nicholas, J., … Christensen, H. (2015). The use of technology in Suicide Prevention. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (pp. 7316–7319). https://doi.org/10.1109/EMBC.2015.7320081

[21] Chakraborty, D., Yang, Z., Tahir, Y., Maszczyk, T., Dauwels, J., Thalmann, N., … Lee, J. (2018). PREDICTION OF NEGATIVE SYMPTOMS OF SCHIZOPHRENIA FROM EMOTION RELATED LOW-LEVEL SPEECH SIGNALS. *IEEE*, 6024–6028.

[22] Faurholt-Jepsen, M., Busk, J., Frost, M., Vinberg, M., Christensen, E. M., Winther, O., … Kessing, L. V. (2016). Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry*, *6*(7), e856-8. https://doi.org/10.1038/tp.2016.123

[23] 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.

[24] Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., … Singh, N. C. (2010). Sounds of melody—Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45. https://doi.org/https://doi.org/10.1016/j.neulet.2010.04.066

[25] Sadock, B., Sadock, V., & Ruiz, P. (2017). *Compêndio de Psiquiatria: Ciência do Comportamento e Psiquiatria Clínica* (11.). Porto Alegre: Artmed.

[26] World Health Organization. (2018). Depression. Retrieved November 11, 2019, from https://www.who.int/en/news-room/fact-sheets/detail/depression

[27] Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015.Vocal

[28] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*(April), 10–49. https://doi.org/10.1016/j.specom.2015.03.004

[29] American Psychiatric Association. (2013). *DSM-5 - Manual Diagnóstico e Estatístico de Transtornos Mentais*. *Artmed* (5.). Porto Alegre: Artmed. https://doi.org/1011769780890425596

[30] Hashim, N. W., Wilkes, M., Salomon, R., Meggs, J., & France, D. J. (2016). Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2016.06.006

[31] Simeone, J. C., Ward, A. J., Rotella, P., Collins, J., & Windisch, R. (2015). An evaluation of variation in published estimates of schizophrenia prevalence from 1990-2013: A systematic literature review. *BMC Psychiatry*, *15*(193), 1–14. https://doi.org/10.1186/s12888-015-0578-7

[32] Foussias, G., & Remington, G. (2010). Negative Symptoms in Schizophrenia: Avolition and Occam's Razor. *Schizophrenia Bulletin*, *36*(2), 359–369. https://doi.org/10.1093/schbul/sbn094

[33] Mac-Kay, A., Jerez, I., & Pesenti, P. (2018). Speech-language intervention in schizophrenia: an integrative review. *Rev. CEFAC*, *20*(2), 238–246. https://doi.org/10.1590/1982-0216201820219317

[34] Chakraborty, D., Xu, S., Yang, Z., Han, Y., Chua, V., Tahir, Y., … Lee, J. (2018). Prediction of Negative Symptoms of Schizophrenia from Objective Linguistic, Acoustic and Non-verbal Conversational Cues. *IEEE 2018 International Conference on Cyberworlds Prediction*, 280–283. https://doi.org/10.1109/CW.2018.00057

[35] Elite, A., Pedrão, L. J., Zamberlan-Amorim, N. E., Carvalho, A. M. P., & Bárbaro, A. M. (2014). Comportamento comunicativo de indivíduos com esquizofrenia. *Rev. CEFAC*, *16*(4), 1283–1293.

[36] Alpert, M., & Anderson, L. T. (1977). Imagery mediation of vocal emphasis in flat affect. *Archives of General Psychiatry*, *34*(2), 208–212.

[37] Alpert, Murray, Rosenberg, S. D., Pouget, E. R., & Shaw, R. J. (2000). Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Research*, *97*, 107–118.

[38] Kuperberg, G. R. (2010). Language in schizophrenia Part 1: an Introduction Gina. *Lang Linguist Compass*, *4*(8), 576–589. https://doi.org/10.1111/j.1749-818X.2010.00216.x.Language

[39] Sanches, M., Bauer, I. E., Galvez, J. F., Zunta-Soares, G. B., & Soares, J. C. (2015). The Management of Cognitive Impairment in Bipolar Disorder. *American Journal of Therapeutics*, *22*(6), 477–486. https://doi.org/10.1097/mjt.0000000000000120

[40] Merikangas, K. R., Jin, R., He, J. P., Kessler, R. C., Lee, S., Sampson, N. A., … Zarkov, Z. (2011). Prevalence and correlates of bipolar spectrum disorder in the World Mental Health Survey Initiative.

*Archives of General Psychiatry*, *68*(3), 241–251. https://doi.org/10.1001/archgenpsychiatry.2011.12

[41] Rowland, T. A., & Marwaha, S. (2018). Epidemiology and risk factors for bipolar disorder. *Therapeutic Advances in Psychopharmacology*, *8*(9), 251–269. https://doi.org/10.1177/https

[42] Hayes, J. F., Miles, J., Walters, K., King, M., & Osborn, D. P. J. (2015). A systematic review and meta-analysis of premature mortality in bipolar affective disorder. *Acta Psychiatrica Scandinavica*, *131*, 417–425. https://doi.org/10.1111/acps.12408

[43] Novick, D. M., Swartz, H. A., & Frank, E. (2010). Suicide attempts in bipolar I and bipolar II disorder: a review and meta-analysis of the evidence. *Bipolar Disord.*, *12*(1), 1–9. https://doi.org/10.1016/j.physbeh.2017.03.040

[44] Maxhuni, A., Muñoz-meléndez, A., Osmani, V., Perez, H., Mayora, O., & Morales, E. F. (2016). Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive and Mobile Computing*, *31*(1), 50–66. https://doi.org/10.1016/j.pmcj.2016.01.008

[45] Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience*, *17*(3), 327–335.

[46] Hans-Ulrich Wittchen. (2002). GENERALIZED ANXIETY DISORDER: PREVALENCE, BURDEN, AND COST TO SOCIETY. *DEPRESSION AND ANXIETY*, *16*, 162–171. https://doi.org/10.1002/da.10065

[47] Van Puyvelde, M., Neyt, X., McGlone, F., & Pattyn, N. (2018). Voice Stress Analysis: A New Framework for Voice and Effort in Human Performance. *Frontiers in Psychology*, *9*(NOV), 1–25. https://doi.org/10.3389/fpsyg.2018.01994

[48] Almeida, A., Behlau, M., & Leite, R. (2011). Correlação entre ansiedade e performance comunicativa. *Rev. Soc. Bras. Fonoaudiol.*, *16*(4), 384–389.

[49] Özseven, T., Dügenci, M., Doruk, A., & Kahraman, H. I. (2018). Voice Traces of Anxiety: Acoustic Parameters Affected by Anxiety Disorder. *Archives of Acoustics*, *43*(4), 625–636. https://doi.org/10.24425/aoa.2018.125156

[50] Newman, S., & Mather, V. G. (1938). Analysis of spoken language of patients with affective disorders. *American Journal of Psychiatry*, *94*, 913–942.

[51] Chevrie-Muller, C., Seguier, N., Spira, A., & Dordain, M. (1978). Recognition of Psychiatric Disorders From Voice Quality. *Language and Speech*, *21*(1), 87–111. https://doi.org/https://doi.org/10.1177/002383097802100106

[52] Goldfarb, W., Braunstein, P., & Lorge, I. (1956). Childhood schizophrenia: Symposium, 1955: 5. A study of speech patterns in a group of schizophrenic children. *American Journal of Orthopsychiatry*, *26*(3), 544–555. https://doi.org/https://doi.org/10.1111/j.1939-0025.1956.tb06201.x

[53] Gottschalk, L. A., Gleser, G. C., Magliocco, E. B., & D'Zmura, T. L. (1961). Further studies on the speech patterns of schizophrenic patients. *Journal of Nervous and Mental Disease*, *132*, 101–113. https://doi.org/https://doi.org/10.1097/00005053-196113220-00001

[54] Smith, G. A. (1977). Voice analysis for the measurement of anxiety. *British Journal of Medical Psychology*, *50*(4), 367–373. https://doi.org/https://doi.org/10.1111/j.2044-8341.1977.tb02435.x

[55] 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).

[56] Low, L. S. A., Maddage, N. C., Lech, M., Sheeber, L. B., & Allen, N. B. (2011). Detection of clinical depression in adolescents' speech during family interactions. *IEEE Transactions on Biomedical Engineering*, *58*(3 PART 1), 574–586. https://doi.org/10.1109/TBME.2010.2091640

[57] Cummins, N., Epps, J., Breakspear, M., & Goecke, R. (2011). An Investigation of Depressed Speech Detection: Features and Normalization.

[58] Hasan, R., Jamil, M., Rabbani, G., & Rahman, S. (2004). Speaker Identification Using Mel Frequency Cepstral Coefficients. *3rd International Conference on Electrical & Computer Engineering ICECE 2004*, (December), 565–568.

[59] Cummins, N., Epps, J., Sethu, V., & Krajewski, J. (2014). Variability compensation in small data: Oversampled extraction of i-vectors for the classification of depressed speech. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 970–974. https://doi.org/10.1109/ICASSP.2014.6853741

[60] Cohn, J. F., Kruez, 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

[61] Ooi, K. E. B., Lech, M., & Brian Allen, N. (2013). Multichannel Weighted Speech Classification System for Prediction of Major Depression in Adolescents. *IEEE Transactions on Biomedical Engineering*, *60*(2), 497–506. https://doi.org/10.1016/j.bspc.2014.08.006

[62] 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.

[63] Jiang, H., Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., … Li, X. (2017). Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Communication*, *90*, 39–46. https://doi.org/10.1016/j.specom.2017.04.001

[64] Alghowinem, S., Goecke, R., Wagner, M., & Epps, J. (2013). Detecting Depression: A Comparison Between Spontaneous and Read Speech. *IEEE*, 7547–7551.

[65] Mitra, V., & Shriberg, E. (2015). Effects of Feature Type, Learning Algorithm and Speaking Style for Depression Detection from Speech. *IEEE*, 4774–4778.

[66] Higuchi, M., Nakamura, M., Shinohara, S., Omiya, Y., Takano, T., Toda, H., … Tokuno, S. (2019). Discrimination of Bipolar Disorders Using Voice. *MindCare*, *1*, 199–207. https://doi.org/10.1007/978-3-030-25872-6

[67] Ringeval, F., Valstar, M., Cowie, R., Schmitt, M., Cummins, N., Lalanne, D., … Salah, A. A. (2018). AVEC 2018 Workshop and Challenge: Bipolar Disorder and Cross-Cultural Affect Recognition. *AVEC'18*, 3–13.

[68] Martínez-sánchez, F., Muela-martínez, J. A., Cortés-soto, P., José, J., Meilán, G., Antonio, J., … Valverde, P. (2015). Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? *The Spanish Journal of Psychology*, *18*(86), 1–9. https://doi.org/10.1017/sjp.2015.85

[69] Compton, M. T., Lunden, A., Cleary, S. D., Pauselli, L., Alolayan, Y., Halpern, B., … Covington, M. A. (2018). The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech. *Schizophrenia Research*, 1–8. https://doi.org/10.1016/j.schres.2018.01.007

[70] Covington, M. A., Lunden, S. L. A., Cristofaro, S. L., Wan, C. R., Bailey, C. T., Broussard, B., … Compton, M. T. (2012). Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. *Schizophrenia Research*, *142*, 93–95.

[71] Tahir, Y., Yang, Z., Id, D. C., Thalmann, N., Thalmann, D., Maniam, Y., … Dauwels, J. (2019). Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. *PLOS ONE*, 1–17. https://doi.org/10.1371/journal.pone.0214314

[72] Weeks, J. W., Lee, C., Reilly, A. R., Howell, A. N., France, C., Kowalsky, J. M., & Bush, A. (2012). Journal of Anxiety Disorders " The Sound of Fear ": Assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. *Journal of Anxiety Disorders*, *26*(8), 811–822. https://doi.org/10.1016/j.janxdis.2012.07.005

[73] Weeks, J. W., Srivastav, A., Howell, A. N., & Menatti, A. R. (2016). "Speaking More than Words": Classifying Men with Social Anxiety Disorder via Vocal Acoustic Analyses of Diagnostic Interviews. *J Psychopathol Behav Assess*, *38*, 30–41. https://doi.org/10.1007/s10862-015-9495-9

[74] Gonçalves, D. M., Stein, A. T., & Kapczinski, F. (2008). Avaliação de desempenho do Self-Reporting Questionnaire como instrumento de rastreamento psiquiátrico: Um estudo comparativo com o Structured Clinical Interview for DSM-IV-TR. *Cadernos de Saude Publica*, *24*(2), 380–390. https://doi.org/10.1590/S0102-311X2008000200017

[75] Hamilton, M. (1960). A RATING SCALE FOR DEPRESSION. *J. Neurol. Neurosurg. Psychiat.*, *23*, 56–62.

[76] Leucht, S., Kane, J. M., Kissling, W., Hamann, J., Etschel, E., & Engel, R. (2005). Clinical implications of Brief Psychiatric Rating Scale scores. *British Journal of Psychiatry*, *187*(2), 366–371. https://doi.org/10.1016/j.physbeh.2017.03.040

[77] Overall, J. E., & Gorham, D. R. (1962). THE BRIEF PSYCHIATRIC RATING SCALE. *Psychological Reports*, *10*, 799–812.

[78] Vilela, J. A. A., Crippa, J. A. S., Del-Ben, C. M., & Loureiro, S. R. (2005). Reliability and validity of a Portuguese version of the Young Mania Rating Scale. *Brazilian Journal of Medical and Biological Research*, *38*(9), 1429–1439. https://doi.org/10.1590/S0100-879X2005000900019

[79] Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A Rating Scale for Mania. *British Journal of Psychiatry*, *133*, 429–435. https://doi.org/10.1192/bjp.133.5.429

[80] Jordan, P., Shedden-Mora, M. C., & Löwe, B. (2017). Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. *PLoS ONE*, *12*(8), 1–14. https://doi.org/10.1371/journal.pone.0182162

[81] Spitzer RL, Kroenke K, Williams JW, & Löwe B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder. *Archives of Internal Medicine*, *166*(10), 1092–1097.

[82] Santos, K. O. B., Araújo, T. M., Pinho, P. S., & Silva, A. C. C. (2010). Avaliação de um Instrumento de Mensuração de Morbidade Psíquica. *Revista Baiana de Saúde Pública*, *34*(3), 544–560.

[83] Zimmerman, M., Martinez, J. H., Young, D., Chelminski, I., & Dalrymple, K. (2013). Severity classification on the Hamilton depression rating scale. *Journal of Affective Disorders*, *150*(2), 384–388. https://doi.org/10.1016/j.jad.2013.04.028

[84] LUKASIEWICZ, M., GERARD, S., BESNARD, A., FALISSARD, B., PERRIN, E., SAPIN, H., … GROUP, T. E. S. (2013). Young Mania Rating Scale: how to interpret the numbers?Determination of a severity threshold and of the minimal clinically significant difference in the EMBLEM cohort. *International Journal of Methods in Psychiatric Research*, *22*(1), 46–58. https://doi.org/10.1002/mpr

[85] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321-357.

[86] Han, H., Wang, W. Y., & Mao, B. H. (2005, August). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In *International Conference on Intelligent Computing* (pp. 878-887). Springer, Berlin, Heidelberg.

[87] Gnanambal, S., Thangaraj, M., Meenatchi, V. T., & Gayathri, V. (2018). Classification algorithms with attribute selection: an evaluation study using WEKA. *International Journal of Advanced Networking and*

*Applications*, 9(6), 3640-3644.

[88] Sivanandam, S. N., & Deepa, S. N. (2008). Genetic algorithms. In *Introduction to Genetic Algorithms* (pp. 15-37). Springer, Berlin, Heidelberg.

[89] Holland, J. H. (1992). Genetic algorithms. *Scientific American*, 267(1), 66-73.

[90] Kennedy, J., & Eberhart, R. (1995, November). Particle swarm optimization. In *Proceedings of ICNN'95 -International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE.

[91] Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization. *Swarm Intelligence*, 1(1), 33-57.

[92] Bratton, D., & Kennedy, J. (2007, April). Defining a standard for particle swarm optimization. In *2007 IEEE Swarm Intelligence Symposium* (pp. 120-127). IEEE.

[93] Dorigo, M., & Di Caro, G. (1999, July). Ant colony optimization: a new meta-heuristic. In *Proceedings of The 1999 Congress on Evolutionary Computation - CEC99* (Cat. No. 99TH8406) (Vol. 2, pp. 1470-1477). IEEE.

[94] Sun, Y., Dong, W., & Chen, Y. (2017). An improved routing algorithm based on ant colony optimization in wireless sensor networks. *IEEE Communications Letters*, 21(6), 1317-1320.

# Figures

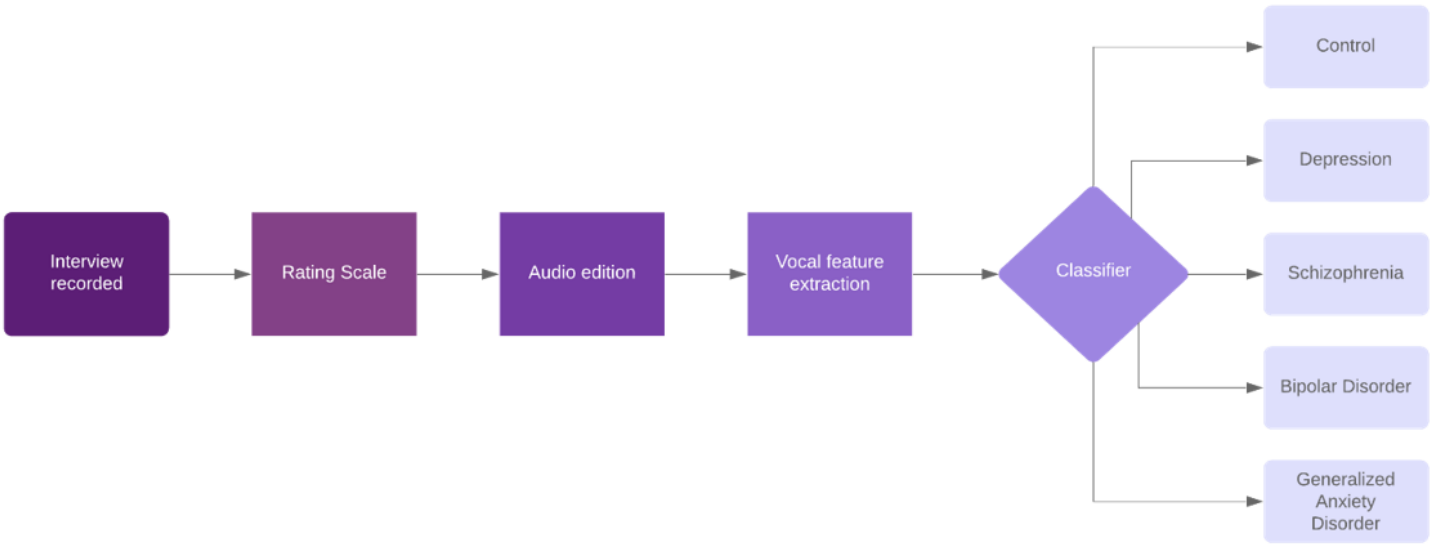


**Figure 1**

Block diagram of data collection and proposed solution