Psychiatria Danubina, 2023; Vol. 35, No. 4, pp 489-499 https://doi.org/10.24869/psyd.2023.489 *Original paper*

A NEW ARTIFICIAL INTELLIGENCE-BASED CLINICAL DECISION SUPPORT SYSTEM FOR DIAGNOSIS OF MAJOR PSYCHIATRIC DISEASES BASED ON VOICE ANALYSIS

Neslihan Cansel[1], Ömer Faruk Alcin[2], Ömer Furkan Yılmaz[3], Ali Ari[4], Mustafa Akan[5] & İlknur Ucuz[6]

1 Assoc. Prof. Dr., Department of Psychiatry, Inonu University, Faculty of Medicine, Malatya, Turkey
2 Assoc. Prof. Dr., Department of Software Engineering, Inonu University, Faculty of Engineering, Malatya, Turkey
3 MD, Department of Psychiatry, Yeşilyurt Hasan Çalik Hospital, Malatya, Turkey
4 Assist. Prof. Dr., Department of Computer Engineering, Inonu University, Faculty of Engineering, Malatya, Turkey
5 Assist. Prof. Dr., Department of Psychiatry, Turgut Özal University, Faculty of Medicine, Malatya, Turkey
6 Assoc. Prof. Dr., Department of Child and Adolescent Psychiatry, Inonu University Faculty of Medicine Malatya, Turkey

received: 19. 04. 2023; revised: 24. 05. 2023; accepted: 31. 08. 2023

Summary

Background: *Speech features are essential components of psychiatric examinations, serving as important markers in the recognition and monitoring of mental illnesses. This study aims to develop a new clinical decision support system based on artificial intelligence, utilizing speech signals to distinguish between bipolar, depressive, anxiety and schizophrenia spectrum disorders.*

Subjects and methods: *A total of 79 patients, who were admitted to the psychiatry clinic between 2020-2021, including 15 with schizophrenia spectrum disorders, 24 with anxiety disorders, 25 with depressive disorders, and 15 with bipolar affective disorder, alongside with 25 healthy individuals were included in the study. The speech signal dataset was created by recording participants' readings of two texts determined by the Russell emotion model. The number of speech samples was increased by using random sampling in speech signals. The sample audio signals were decomposed into time-frequency coefficients using Wavelet Packet Transform (WPT). Feature extraction was performed using each coefficient obtained from both Mel-Frequency Cepstral Coefficients (MFCC) and Gammatone Cepstral Coefficient (GTCC) methods. The disorder classification was carried out using k-Nearest Neighbor (kNN) and Support Vector Machine (SVM) classifiers.*

Results: *The success rate of the developed model in distinguishing the disorders was 96.943%. While the kNN model exhibited the highest performance in diagnosing bipolar disorder, it performed the least effectively in detecting depressive disorders. Whereas, the SVM model demonstrated close and high performance in detecting anxiety and psychosis, but its performance was low in identifying bipolar disorder.*

The findings support the utilization of speech analysis for distinguishing major psychiatric disorders. In this regard, the future development of artificial intelligence-based systems has the potential to enhance the psychiatric diagnosis process.

Keywords: *Artificial intelligence, mental illness, psychiatry, speech signal, Russel emotion model.*

* * * * *

INTRODUCTION

Mental illnesses are common illnesses characterized by a clinically significant disturbance in an individual's cognition, emotional regulation, or behavior. It has been reported that approximately 970 million people worldwide are currently affected by one or more mental disorders, with an expected increase in the future (WHO, 2022). Psychiatric disorders are predominantly chronic and have a substantial impact on the global economy, leading to decreased work performance and high treatment costs (Jarman et al. 2016). Although these disorders can be treated and symptom improvement is possible with accurate diagnosis, the presence of biological and clinical heterogeneity, coupled with the absence of diagnostic biomarkers, makes the diagnostic process challenging (Bedi et al. 2015; Insel & Landis 2013).

The diagnosis of psychiatric disorders is still relies on self-reporting, information gathered from relatives, long-term interviews and scales (Regier et al. 2013). However, reasons such as avoiding social stigma, reluctance to interview, and retrospective recall bias may cause the data obtained to be far from objectivity (Yünden 2022, Low et al. 2020). Furthermore, the power of the scales used in assessment, management and scoring is limited and costly due to time consuming, serious training and multiple information requirements (Kobak et al. 2004). Despite advancements in neurobiological studies that enhance our understanding of the biological foundations of psychiatric disorders, they have not yielded sufficient biomarkers to enhance the objectivity of psychiatric

assessment (Insel & Landis 2013). Consequently, there is a need for new approaches. In this regard, significant advancements in computer technology have revolutionized the field of psychiatry, similar to other medical disciplines, by enabling the detection of disorder-specific features and facilitating the prediction of treatment response and prognosis (Siena et al. 2020, van der Sluis et al. 2011, Marmar et al. 2019, Hoque et al. 2009, Bzdok D & Meyer-Lindenberg A 2018).

Speech is a parameter that is frequently examined in this field. The reasons for this preference could be attributed to its advantages such as containing numerous clinical clues, difficulty in concealing verbal and non-verbal features of the person during speaking, direct expression of emotions and thoughts through language, and indirect reflection of neuromuscular modulation. Economic factors, availability, and low cost are also among the motivations for this preference (Bedi et al. 2015; Low et al. 2020, Yünden 2022).

In recent years, a significant number of studies have demonstrated that speech patterns and features collected through mobile devices and sensors can serve as biomarkers for early diagnosis and monitoring of mental disorders (van der Sluis et al. 2011, Cannizzaro et al. 2004, Pan et al. 2019, Hashim et al. 2017, Karam et al. 2014, Marmar et al. 2019, Hoque et al. 2009, de Boer et al. 2020, Siena et al. 2020). For instance, Hashim et al. suggested that changes in speech signals consisting of acoustic features which characterize specific spectral and timing properties can be used in monitoring severity of depressive symptoms and treatment response (Hashim et al. 2017). Faurholt-Jepsen et al. analyzed smartphone and self-monitored data over a period of 12 weeks, demonstrating the importance of voice in distinguishing affective fluctuations, depression, and manic symptoms (Faurholt-Jepsen et al. 2016). Mota et al successfully measured dysfunctional thought flow such as divergence and recurrence in the speech of a group of psychotic patients can be objectively measured by speech graph analysis (Mota et al. 2012).

The fact that most of the studies are based on a single disease group may lead to a decrease in the general use of the obtained objective markers and consequently limit their reliability. To the best of our knowledge, there is no existing study in the literature that distinguishes between the main psychiatric disorder groups using voice analysis.

Therefore, in this study, it is aimed to develop an artificial intelligence-based clinical decision support system (CDSS) with high accuracy, sensitivity, specificity by using speech analysis which distinguishes patients with four main psychiatric disorders including schizophrenia spectrum, depressive, and bipolar affective disorders and healthy individuals.

SUBJECTS AND METHODS

This is a cross-sectional study that received prior approval from the local ethics committee (2020/25). The study included patients who admitted to the psychiatry outpatient clinic of Inonu University Faculty of Medicine between March 2020 and January 2021 and were diagnosed with Anxiety Disorders, Bipolar Disorder, Depressive Disorders, Schizophrenia Spectrum Disorders (SSD) and followed up according to DSM-5 diagnostic criteria (American Psychiatric Association 2013). Patients evaluated by two psychiatrists in accordance with the DSM-5 criteria and diagnoses which were confirmed by psychometric scales were invited to the study. Their voices were recorded using an android smartphone (Samsung Galaxy S8, Samsung Electronics, 2017, South Korea). Psychiatric symptoms were assessed using scales with proven validity and reliability in Turkey, including the Negative Syndrome Scale (SANS) (Erkoç et al. 1991a) and Positive Symptoms Rating Scale (SAPS) for schizophrenia spectrum disorders (Erkoç et al. 1991b), Hamilton Anxiety Rating Scale (HAM-A) for anxiety disorders (Yazıcı et al. 1998), Hamilton Depression Rating Scale (HAM-D) for depressive disorders (Akdemir et al. 1996), and Young Mania Rating Scale (YMRS) for bipolar affective disorder (Karadag et al. 2001). Demographic and clinical characteristics such as age, gender, marital status, duration of psychiatric illness, and use of psychotropic drugs were recorded.

A control group was selected, matching the patient groups in terms of age, and consisting of individuals who were evaluated by the same psychiatrists. These individuals underwent a semi-structured interview and were determined not to meet any psychiatric disorders criteria according to DSM-5. It was ensured that the healthy individuals had not received treatment for any mental illnesses previously. Furthermore, participants with conditions or history such as voice impairment or alteration due to diseases (reflux, pharyngitis, etc.) or surgeries, voice or diction training, speech disorders (stuttering, dysarthria), neurological diseases, intellectual disability causing cognitive impairment, or inability to speak Turkish were excluded from the study. Participation was voluntary, and written consent was obtained.

During the data collection period, the researcher responsible for the analysis did not have access to the collected data.

Data Collection

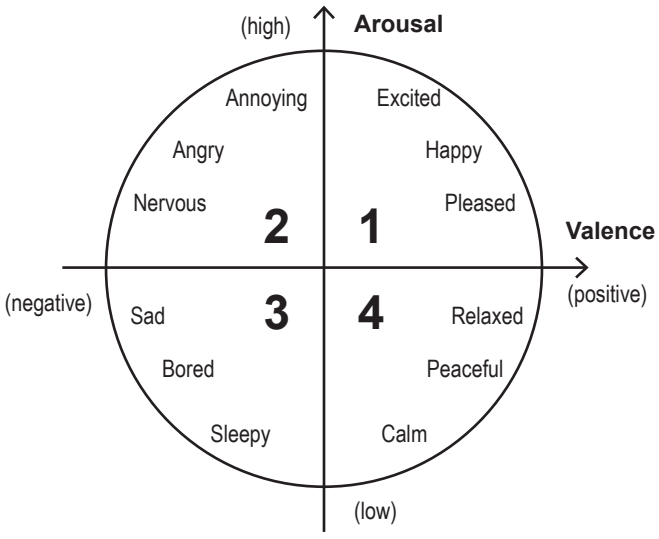
Data collection involved obtaining voice signals from the participants reading two texts provided to them. These texts were selected using The Russel Arousal-Valence emotion model in order to minimize any possible emotional impact (Figure 1). The model involves presenting a stimulus to the subjects first and followed by an self-evaluation of the emotion created by this stimulus with the Self-Assessment Manikin Questionnaire (SAM) (Russell 2003, Alakus et al. 2020). According to SAM, texts corresponding to zone 4th (relaxed, peaceful, calm) were considered neutral stimulus. Initially, 10 texts were chosen, and each text was read by the researchers. Two texts that corresponded to the 4th region were selected for the study (Figure 1).

The 2 texts determined using this method are as follows:

Text 1. "BUTTERFLY VALLEY: Fethiye is a paradise garden that can be accessed from the Dead Sea (Blue Lagoon) by boats. It brings many people together and has a unique magic. It is also famous for its waterfalls and the tiger-patterned butterflies found only in this region. "

Text 2. "THE DECLARATION OF REPUBLIC: With the acceptance of the constitutional amendment proposal prepared by Mustafa Kemal at Turkey's Grand National Assembly in its second period, 29 October 1923, it is determined that the form of government in Turkey is a republic.

Figure 1. Russel Arousal-Valence emotion model



Speech Signals Analyze and Classification Methods

Speech signals possess a complex structure that contains valuable information. However, in order to utilize these signals, they must undergo a series of preprocessing steps, such as enhancing signal quality, emphasizing relevant components, suppressing external noises, and determining appropriate sampling values. Following this preprocessing stage, the next step is to extract features from the speech signals. Feature extraction involves identifying characteristic values that describe the speakers and can be used for subsequent recognition. The features extracted from the audio signals can be classified into acoustic, linguistic, contextual, and hybrid features, which combine different sets of features. Acoustic features are often preferred in studies as they provide more objective insights into sound production and signal structure. Commonly used acoustic features include intonation, formant frequencies, speech rate, sound quality, and features based on Mel-Frequency Cepstral Coefficients (MFCC) (Özseven 2019, Eskidere & Ertaş 2009).

In this study, the speech signals also underwent a series of preprocessing steps, including screening and determination of sampling values. Subsequently, they were transformed into the time-frequency domain using the Wavelet Packet Transform (WPT) process. The speech features within the transformed signals were computed using the MFCC and Gammatone Cepstral Coefficient (GTCC) methods. To classify the extracted features according to disease groups, Support Vector Machines (SVM) and k-Nearest Neighbor (kNN) methods were employed. The following sections provide a brief explanation of these methods.

Wavelet Packet Transform

The wavelet transform is a useful tool for the short-time analysis of speech signals, particularly those that are quasi-stationary. The key aspect of the wavelet transform is to analyze a signal considering scale. This scaling approach enables both locality and spectral analysis, providing a time-frequency representation. There are various type of wavelet transforms such as discrete wavelet transform and WPT. Discrete wavelet transform is suitable for analyzing low-frequency signals, yet it exhibits relatively low resolution in the high frequency region. On the other hand, WPT can analyze a signal containing low, mid, and high-frequency components similarly to the speech signals (Burrus et al. 1998, Gao & Yan 2011). This feature made WPT to commonly used for detecting and distinguishing transients with high frequency characteristics.

Mel Frequencies Cepstral Coefficient

Mel Frequencies Cepstral Coefficient is a method used to detect and characterize speech-specific attributes in speaker verification systems. MFCC is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. By dividing the speech into small frames based on the number of input rows, window length, and overlap length, MFCC computes cepstral features for each frame. One of the most significant features of MFCC is its ability to mimic the frequency selectivity of the human ear, enabling the extraction of distinctive and highly performant values (Hossan et al. 2010, Dimitrov 2005).

Gammatone Cepstral Coefficient

Gammatone Cepstral Coefficients (GTCC) are a feature extraction technique widely used in speech and audio signal processing. Although, GTCC feature extraction is similar to MFCC, the Gammatone Cepstral Coefficient (GTCC) adopts a frequency scale that has been analytically devised, showcasing a more refined behavior compared to the Mel-scale. The impulse response of the Gammatone filter is derived from a combination of the Gamma distribution function and a sinusoidal tone with a specific frequency positioned at its center. Hence, by utilizing the Gammatone filter, the GTCC not only enhances the representation of auditory signals but also demonstrates the ability to capture the nuances of the human auditory system with greater fidelity than MFCC (Balli 2022, Valero & Alias 2012),

k-Nearest Neighbor

The k-NN algorithm, a non-parametric learning algorithm, is widely used among machine learning methods due to its simplicity and good performance (Cover & Hart 1967). The k-NN method determines the class to which a new observation belongs by utilizing the observation values in a sample set with predefined classes. In this algorithm, the distance between the new data and the recognized data is calculated for k points, and the new information is classified based on this distance (Ozkan 2016).

Support Vector Machines

SVM is an important and efficient supervised classification algorithm. The essence of the algorithm is based on detecting a hyperplane in a high-dimensional space that maximally separates the different classes. When a linearly separable dataset is provided to the system by binary classification, an infinite number of hyperplanes are formed that can separate the input set. SVM constructs a decision plane to maximize the distance between the two classes. This decision plane allows for the classification of objects with different class memberships, and SVM finds the best hyperplane with the maximum margin to separate them (Tekerek 2019, Siuly et al. 2020).

Statistical Analysis

SPSS 22 (Statistical Package for Social Sciences; SPSS Inc., Chicago, IL) package program was used to analyze the data of the participants. Among the descriptive data in the study, qualitative variables were shown as number and percentage (n and %), and quantitative variables as mean±standard deviation (Mean±SD) values. Independent sample t-test was used to compare the age distribution between the groups. A value of $p<0.05$ was considered statistically significant.

RESULTS

Sociodemographic Features

One hundred four participants were included in the study. Fifteen of the participants had a diagnosis of SSD, 24 had an anxiety disorder, 25 had major depressive disorder, 15 had a diagnosis of bipolar affective disorder. The control group was consisted of 25 healthy individuals. The mean age of the participants was 36.53±13.10 years and there was no statistical difference between the groups ($p=0.129$). It was found that, the mean SANS total score was 27.3±21.02 and the mean SAPS score was 19.46±20 in the SSD patients. The mean HAM-A score was 22.3±9.6 in the anxiety group, the mean HAM-D score was 15.3±5.7 in the depressive group, and the mean YMRS total score was 14.5±7.9 in the bipolar group. The sociodemographic characteristics of the participants were given in Table 1.

Design of Speech Analysis Clinical Decision Support System

The proposed method aims to detect bipolar, depressive, anxiety, and schizophrenia spectrum disorders from speech signals. Participants were instructed to read the two texts mentioned above once. Recordings were excluded from the analysis if there was unwanted background

Table 1. Sociodemographic and clinical features of the participants

Variable		SSD group	Anxiety group	Depressive group	Bipolar group	Healthy controls
Age (mean±SD)		36.5±13.1	35.2±11.3	37.5±11.0	35.6±11.0	29.8±8.1
Psychiatric onset age (mean±SD)		24.3±6.9	30.4±9.6	30.9±10.1	24.3±7.2	-
n (%)						
Gender	Female	6 (40)	20 (83.3)	22 (88)	13 (86.7)	17 (68)
	Male	9 (60)	4 (16.7)	3 (12)	2 (13.3)	8 (32)
Place of birth	Province	11 (73.3)	18 (75)	14 (56)	8 (53.3)	20 (80)
	District	3 (20)	5 (20.8)	7 (28)	6 (40)	3 (12)
	Village	1 (6.7)	1 (4.2)	4 (16)	1 (6.7)	2 (8)
Education level	Primary	4 (26.7)	7 (29.2)	13 (52)	5 (33.3)	0
	High school	6 (40)	8 (33.3)	4 (16)	5 (33.3)	5 (20)
	University	5 (33.3)	9 (37.5)	8 (32)	5 (33.3)	20 (80)
Profession	Unemployed	10 (66.7)	17 (70.8)	19 (76)	11 (73.3)	12 (48)
	Worker	0	1 (4.2)	4 (16)	1 (6.7)	3 (12)
	Officer	5 (33.3)	4 (16.7)	2 (8)	2 (13.3)	10 (40)
	Other	0	2 (8.3)	0	1 (6.7)	0
Marital status	Single	11 (73.3)	7 (29.2)	6 (24)	5 (33.3)	14 (56)
	Married	3 (20)	15 (62.5)	18 (72)	7 (46.7)	11 (44)
	Widow/Divorced	1 (6.7)	2 (8.3)	1 (4)	3 (20)	0
Economic Level	0-2000	8 (53.3)	10 (41.7)	16 (64)	12 (80)	11 (44)
	2000-5000	1 (6.7)	0	2 (8)	0	0
	>5000	6 (40)	14 (58.3)	7 (28)	3 (20)	14 (56)
Living place	Province	15 (100)	21 (87.5)	23 (92)	10 (66.7)	20 (80)
	District	0	1 (4.2)	1 (4)	3 (20)	3 (12)
	Village	0	2 (8.3)	1 (4)	2 (13.3)	2 (8)
Chronic internal disease history	No	13 (86.7)	16 (66.7)	22 (88)	14 (93.3)	20 (80)
	Yes	2 (13.3)	8 (33.3)	3 (12)	1 (6.7)	5 (20)
Alcohol--substance use history	No	14 (93.3)	23 (95.8)	25 (100)	13 (86.7)	25 (100)
	Yes	1 (6.7)	1 (4.2)	0	2 (13.3)	0
History of smoking	No	9 (60)	16 (66.7)	15 (60)	9 (60)	22 (88)
	Yes	6 (40)	8 (33.3)	10 (40)	6 (40)	3 (12)
Current complaints start time	> 1 month	2 (13.3)	4 (16.7)	0	3 (20)	
	1-6 months	1 (6.7)	11 (45.8)	9 (36)	8 (53.3)	
	6 months-1 year	4 (26.7)	2 (8.3)	7 (28)	2 (13.3)	-
	> 1 year	5 (33.3)	7 (29.2)	9 (36)	2 (13.3)	
	Remission	3 (20)	0	0	0	
Drug currently used	None	1 (6.7)	3 (12.5)	6 (24)	0	
	Only AP	12 (80)	1 (4.2)	1 (4)	1 (6.7)	
	Only SRI/SNRI	0	19 (79.2)	13 (52)	0	-
	Only MSD	0	0	0	3 (20)	
	SSRI+AP	1 (6.7)	1 (4.2)	4 (16)	0	
	AP+MSD	1 (6.7)	0	1 (4)	11 (73.3)	

AP: Antipsychotic, SSRI: Selective Serotonin Reuptake Inhibitor; SNRI: Serotonin-Noradrenaline Reuptake inhibitor, MSD: Mood-Stabilizing Drugs, SSD: Schizophrenia Spectrum Disorders

noise or if the speech was not clearly understandable. As a result, a dataset consisting of 193 speech signals was created, involving 104 participants. The speech signals were recorded at a sampling frequency of 44,100Hz, and the overall speech record, which lasted an average of 30 seconds, contained approximately one million samples.

In the experiments, the number of speech signals was increased by a random subsampling approach. through data augmentation, 100,000 samples were randomly selected from each speech signal, and this process was repeated ten times. Consequently, an augmented dataset of 1,930 sampled data points was obtained from the original data. The augmented data set was transformed into time-frequency coefficients using WPT with four levels. Thus, time-frequency coefficients with distinctive speech signal clips were obtained. The parameters of the WPT method, including Shannon entropy and the db3 wavelet family, were chosen heuristically. Sixteen ($2^{\text{-level}}$) time-frequency coefficients are obtained with four-level WPT. Distinctive features were calculated from the WPT coefficients of the speech signals using the MFCC and GTCC methods. A 1x28 feature vector, consisting of fourteen MFCCs and fourteen GTCCs, was formed from each WPT coefficient Considering all the WPT coefficients, a 1,930x448-dimensional feature vector was ultimately obtained. Subsequently, k-NN and SVM classifiers were employed to detect disorders based on the obtained features. For kNN, the parameters were determined experimentally, taking into account the lowest error rate, the Euclidean distance metric, and k=3.

The linear kernel in SVM kernel functions are used because of providing the best results compared to other kernel functions. All experiments were carried out in MATLAB environment on a computer equipped with a 2.70 GHz CPU processor and 32 GB RAM. The algorithm of the proposed method were given in Figure 2.

Experimental Results

To evaluate the performance of the proposed method, accuracy (Acc.), sensitivity (Sens.), specificity (Spec.), precision (Prec.), F1-score, MCC, and kappa metrics,

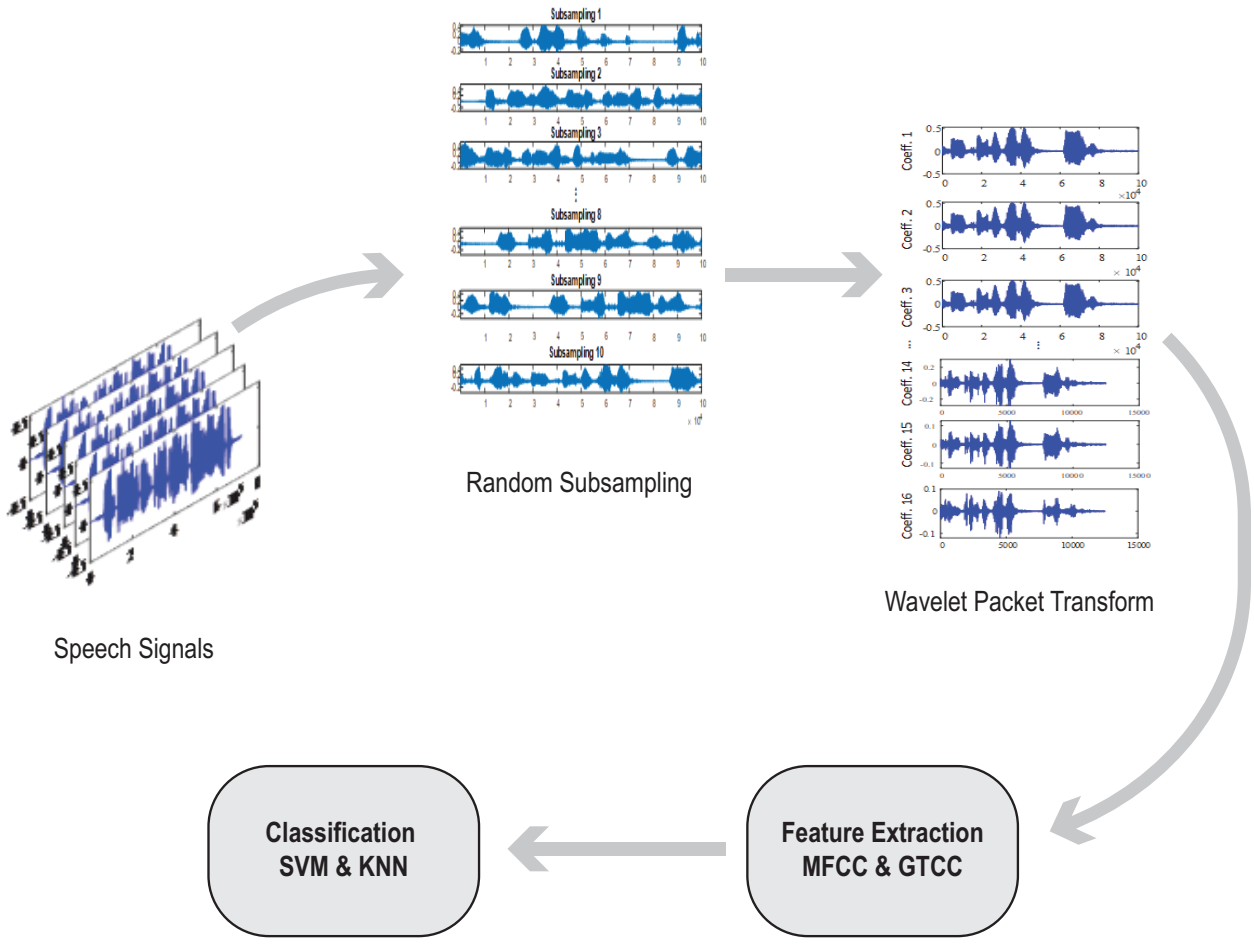


Figure 2. The proposed framework for detection psychiatric disorders from speech signal

which are widely used in biomedical applications were used. Five-fold cross-validation was used to validate the proposed method. Accuracy represents the ability to give the same value during repeated measurements of physical size. Sensitivity indicates the ability of a classification model to correctly identify positive instances from all the actual positive instances. Conversely, specificity refers to the ability of a model to find actual healthy individuals among undiagnosed individuals. Precision is the proportion of correctly predicted positive instances out of all instances predicted as positive. F1-score value is the harmonic mean of precision and recall values. Kappa parameter indicates the agreement between observed and expected values and ranges from -1 to +1. A high kappa value implies that the evaluated model is performing well and shows a strong agreement with the reference. Henceforward, kappa values close to 1 were used for our model.

The performance parameters of the SVM classifier were calculated as 96.477% accuracy, 96.051% sensitivity, 99.110% specificity, 96.338% precision, F1 score of 0.962, MCC value of 0.953, and kappa value of 0.890. The performance parameters of the kNN classifier were calculated as 96.943% accuracy, 96.930% sensitivity, 99.228% specificity, 96.802% precision, F1 score of 0.969, MCC value of 0.961 and kappa value of 0.904. These results of the proposed system are showed in Table 2.

Although the performance parameters of the two classifiers seem to be close to each other, kNN performed slightly better than SVM. Experimental analysis was proceeded by drawing the Receiver Operating Characteristic (ROC) curves. The ROC curve helps to evaluate the overall classifier performance. The ROC curves were shown in Figure 3 for the kNN, and in Figure 4 for the SVM classifier.

Table 2. Performance results of the proposed methods

Method	Acc. (%)	Sens. (%)	Spec. (%)	Prec.(%)	F1-score	MCC	Kappa
SVM	96.477	96.051	99.110	96.338	0.962	0.953	0.890
kNN	96.943	96.930	99.228	96.802	0.969	0.961	0.904

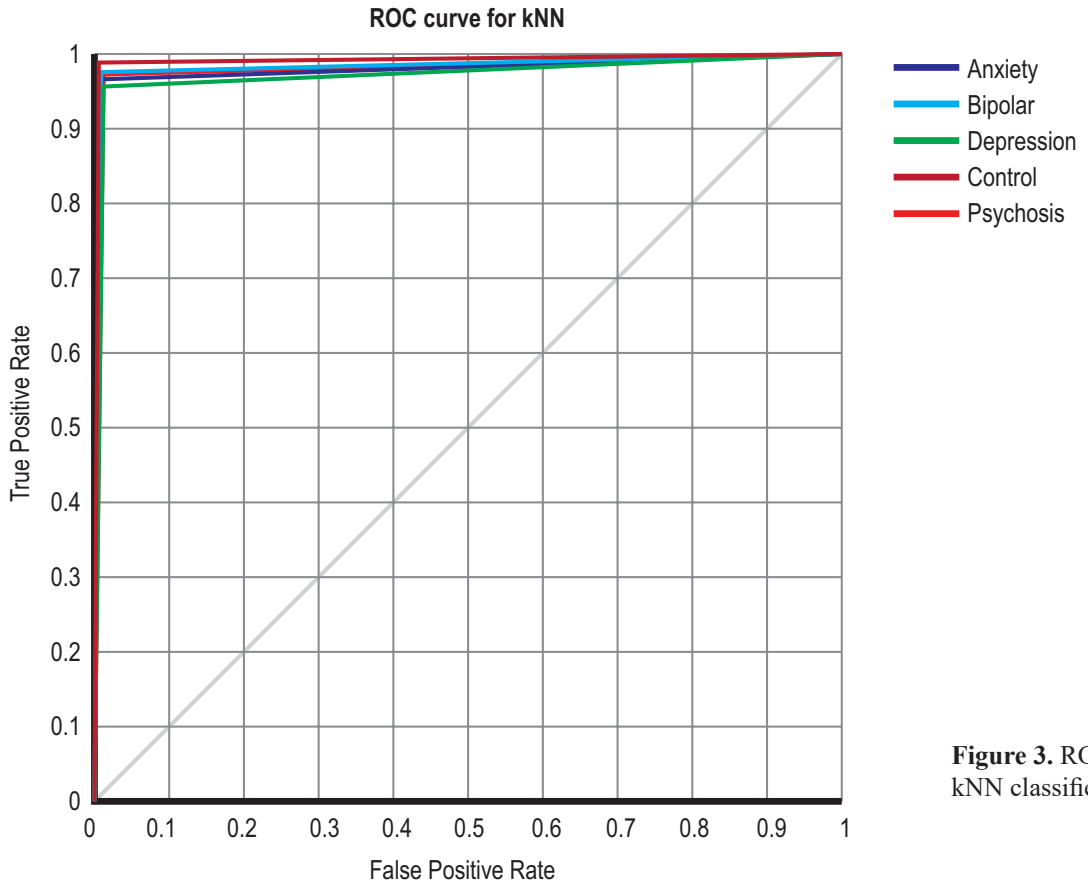


Figure 3. ROC curve for kNN classifier

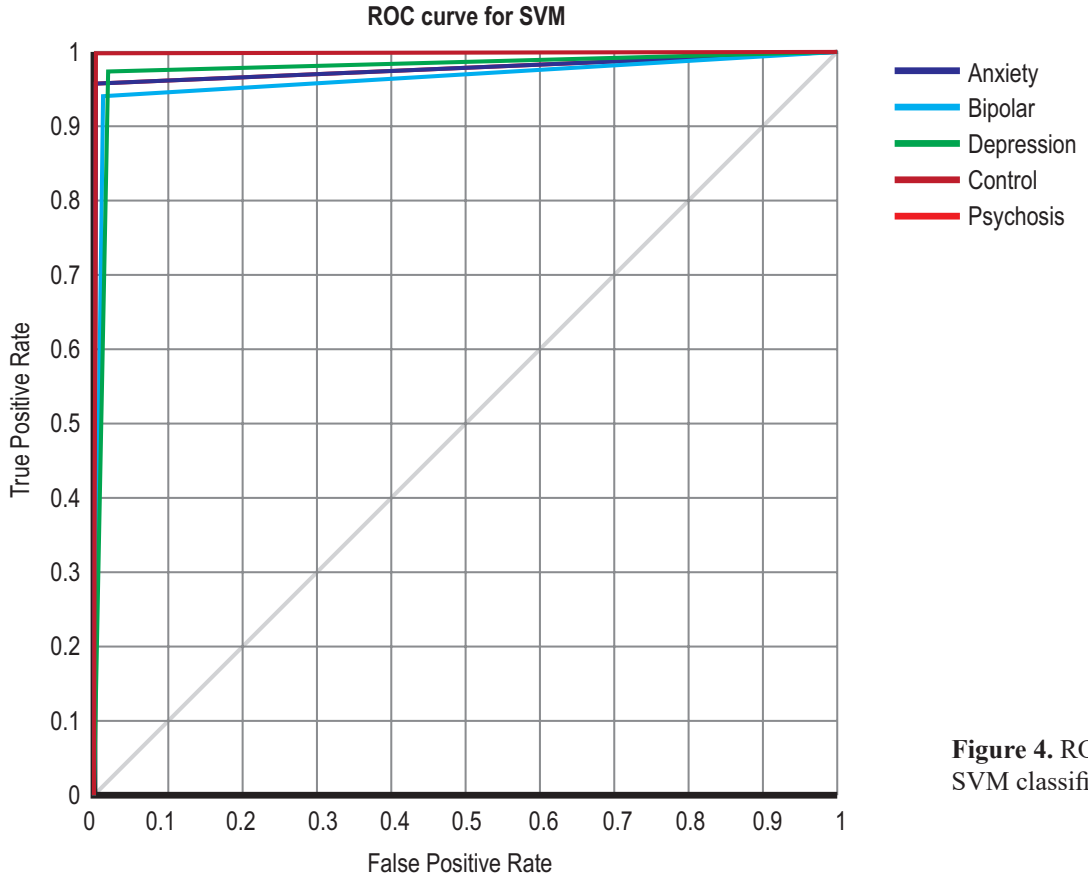


Figure 4. ROC curve for SVM classifier

When examining the ROC curve in Figure 3, kNN was the most successful method in distinguishing the control group from the patient group. However, the kNN model performed lower than other methods in detecting depression. On the other hand, the SVM model distinguishes the control group better, as well as the kNN model. While SVM showed a close performance resemblance in detecting anxiety and psychosis, it displayed lower performance in bipolar detection (Figure 4).

DISCUSSION

In this study, an artificial intelligence-based clinical decision support system was developed with a performance of 96.943% to detect bipolar disorder, depressive, anxiety, and schizophrenia spectrum disorders from sound signals and distinguish them from the healthy control group.

Various studies in the literature have shown that voice analysis can be used for clinical classification in different patient groups. For example, Tahir et al. classified schizophrenia patients and healthy controls with 81.3% accuracy using non-verbal language measures (Tahir et al. 2019). Martínez-Sánchez et al., were able to distinguish schizophrenic patients from healthy controls with over 90% sensitivity, specificity and overall accuracy, thanks to the method they developed using the prosodic and phonetic sound features determined while reading a text (Martínez-Sánchez et al. 2015). By examining the sounds obtained from patients' videotapes, Cannizzaro et al. obtained data indicating that motor timing parameters reflecting speech production (for example, speech rate and pause time) and frequency modulation (for example, pitch variability) were significantly associated with the severity of depression (Cannizzaro et al. 2004). Similarly, Maxhuni et al. in their study followed bipolar patients with a smartphone-based system and classified their mood transitions at a rate higher than 80% by using the speech, accelerometer and self-assessment-related data obtained from daily phone calls (Maxhuni et al. 2016).

In our study, unlike these studies that focused on a single disorder sample, we created a CDSS that covers four main disorders groups and obtained 448 different features from the audio signals to develop this system. This allowed for more sophisticated separations and promising clinical use of the model. It should be noted

that most previous studies used data from daily conversations and interviews, which could involve possible emotional manipulations. In our study, we provided a neutral stimulus using the Russell Arousal-Valence emotion model, resulting in a more objective dataset. With random sampling, the number of samples was increased, so that the developed artificial-intelligence model learned from more samples and its classification performance was increased (Jiao et al. 2018, Aggarwal 2019). By decomposing the augmented dataset into four levels using WPT and extracting distinctive features using MFCC and GTCC methods, we developed a method with a success rate of 96.943%. This result indicated a considerably higher performance compared to studies that partially included a small number of patients, mostly used only one type of data, and did not consider more than two disorders at the same time. Furthermore, we tested existing classifiers and determined which disorders the machine learning models were more specific for. As an illustration, the kNN model exhibited superior performance in identifying bipolar disorder, but it demonstrated the least effective performance in detecting depression. On the other hand, the SVM model demonstrated a strong and comparable performance in distinguishing between anxiety and SSD, although its performance in detecting bipolar disorders was comparatively lower than that of the other groups.

While this study contributes valuable insights, it is important to address certain limitations that should be taken into consideration. First, participation in the research was dependent on voluntary, and as a result, the sample size remained relatively low. Additionally, all of the participants resided in the same region and had similar cultural and linguistic characteristics. Therefore, although the developed method has shown an effective performance for this data set, it is unknown whether the obtained results can be generalized to a new sample with different ages, geography, socioeconomic levels, and registration types. Second, the small sample size prevented the differentiation of variables that could impact the voice structure, such as gender (length of vocal cords and speech patterns, as well as acoustic characteristics like duration, intensity, and frequency of voice and speech signals, which may differ between genders), smoking (which can change the acoustic properties of the voice by causing histological changes on the vocal fold epithelium), and the type of psychotropic treatment (which may cause impairment in speech-related motor functions). Third, subgroups (delusional disorder, schizoaffective disorder, obsessive compulsive disorder, phobias, panic disorder, dysthymia etc) were not analyzed, and patient groups were grouped based on the core symptoms. Consequently, it remains unclear whether the model would be effective in distinguishing subgroups. Fourth, although neuropsychological deficiencies that may affect acoustic or lexical expression were assessed based on history, neuropsychological tests were not employed to detect these conditions. Finally, considering the depth and obscurity of the human psyche, it may not be ethically appropriate to rely solely on sound features for a "definitive" diagnosis. In the future, multicenter studies that would be conducted in large populations and that evaluate more markers together could provide more reliable results.

Despite all these limitations, our study is the first to utilize machine learning methods distinguishing four different main psychiatric disorders and healthy person and may lead to future studies. Considering the performance parameters, the developed method can be employed by experts as an effective and reliable tool to assist in diagnosing mental illnesses, thereby contributing to efficiency in terms of speed and time. The CDSS may be beneficial for patients to be evaluated more objectively, independent of factors such as the evaluator's experience, attention, and mood, as well as the partially low-reliability of self-report measures. Furthermore, it can also be applied to assess specific populations, such as children or the elderly, who may encounter difficulties in expressing themselves. Additionally, it can be employed for monitoring patients undergoing extreme circumstances, such as disasters, wars, or pandemics, where regular check-ups may be hindered.

CONCLUSION

In this study, a new artificial intelligence-based method was presented, which achieves a performance parameter above 96.943% and can automatically and accurately classify psychiatric disorders from healthy controls. The kNN model demonstrated high performance, particularly in diagnosing bipolar disorder, while the SVM model showed comparable high performance in distinguishing anxiety and SSD. The developed method holds the potential to assist psychiatrists in efficiently and reliably distinguishing their patients.

Acknowledgements: None.

Ethical Considerations: Does this study include human subjects? YES

Authors confirmed the compliance with all relevant ethical regulations.

Conflict of interest: No conflict of interest

Funding sources: The authors received no funding from an external source.

Authors contributions: Neslihan cansel & İlknur Ucuz: study design, statistical analysis, first draft. Ömer Furkan Yılmaz & Mustafa Akan: data collection. Ömer Faruk Alçin & Ali Arı: data analysis and development of clinic support system. All authors approval of the final version.

References

1. *Aggarwal SLP: Data augmentation in dermatology image recognition using machine learning. Skin Research and Technology 2019; 25: 815-820.*
2. *Akdemir A, Örsel DS, Dağ İ, Türkçapar MH, İşcan N, Özbay H: Hamilton depresyon derecelendirme ölçeği (HD-DÖ)'nin geçerliliği-güvenirliliği ve klinikte kullanımı. Psikiyatri Psikoloji Psikofarmakoloji Dergisi 1996; 4: 251-259.*
3. *Alakus TB, Gonen M, Turkoglu I: Database for an emotion recognition system based on EEG signals and various computer games–GAMEEMO. Biomed Signal Process Control 2020; 60: 101951.*
4. *American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders 5. Washington, D.C., 2013.*
5. *Balli O: Vücut Seslerinden Bölge Tanımlanması için İdeal Kayıt Süresinin Belirlenmesinde MFCC ve GTCC Özniteliklerinin Etkisinin Karşılaştırılması." Avrupa Bilim ve Teknoloji Dergisi 2022;43:36-40.*
6. *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. NPJ Schizophr 2015; 1: 15030.*
7. *Burrus CS, Gopinath RA, Guo H: Introduction to Wavelets and Wavelet Transforms, A Primer. Upper Saddle River, Nj, Prentice-Hall, 1998.*
8. *Bzdok D & Meyer-Lindenberg A: Machine Learning for Precision Psychiatry: Opportunities and Challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018; 3: 223-230.*
9. *Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ: Voice acoustical measurement of the severity of major depression. Brain Cogn 2004; 56: 30-35.*
10. *Cover T & Hart P: Nearest neighbor pattern classification. IEEE transactions on information theory 1967; 13: 21-27.*
11. *de Boer JN, Brederoo SG, Voppel AE, Sommer IEC: Anomalies in language as a biomarker for schizophrenia. Curr Opin Psychiatry 2020; 33: 212-218.*
12. *Dimitrov Ganchev T: Speaker recognition. Unpublished doctoral dissertation. Dept of Electrical and Computer Engineering, University of Patras, Greece, 2005.*
13. *Erkoç S, Arkonaç O, Ataklı C, Özmen E: Pozitif Semptomları Değerlendirme Ölçeğinin güvenilirliğini ve geçerliliği. Düşünen Adam Dergisi, 1991a; 4: 20-24.*
14. *Erkoç S, Arkonaç O, Ataklı C, Özmen E: Negatif Semptomları Değerlendirme Ölçeğinin güvenilirliğini ve geçerliliği. Düşünen Adam Dergisi, 1991b; 4: 14-15.*
15. *Eskidere Ö & Ertaş F: MEL Frekansı Kepstrum Katsayılarındaki Değişimlerin Konuşmacı Tanımaya Etkisi. Uludağ Üniversitesi Mühendislik Fakültesi Dergisi. 2009; 14(2): 93-110*
16. *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. Translational Psychiatry, 2016; 6(7): e856*
17. *Gao, R.X., Yan, R. (2011). Wavelet Packet Transform. In: Wavelets. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1545-0_5*
18. *Hashim NW, Wilkes M, Salomon R, Meggs J, France DJ: Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores. J Voice 2017; 31: 256.e1-256.e6.*
19. *Hoque ME, Lane JK, El Kaliouby R, Goodwin M, Picard RW: Exploring speech therapy games with children on the autism spectrum. Proc Annu Conf Int Speech Commun Assoc, INTERSPEECH, 2009.*
20. *Hossan MA, Memon S, Gregory MA: A novel approach for MFCC feature extraction. 4th Int Conf Signal Process Commun Syst IEEE, 2010; 1-5.*
21. *Insel TR & Landis SC: Twenty-five years of progress: the view from NIMH and NINDS. Neuron 2013; 80: 561-567.*
22. *Jarman L, Martin A, Venn A, Otahal P, Blizzard L, Teale B & Sanderson K: Workplace Health Promotion and Mental Health: Three-Year Findings from Partnering Healthy@Work. PloS 2016;1(8):e0156791.*
23. *Jiao Y, Tu M, Berisha V, Liss J: Simulating dysarthric speech for training data augmentation in clinical speech applications. In IEEE international conference on acoustics, speech and signal processing (ICASSP) 2018; 6009-6013.*
24. *Karadağ F, Oral ET, Aran Yalçın F, Erten E: Young mani derecelendirme ölçeğinin Türkiye'de geçerlik ve güvenilirliği. Türk Psikiyatri Dergisi 2001; 13: 107-14.*
25. *Karam ZN, Provost EM, Singh S, Montgomery J, Archer C, Harrington G, et al.: Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. Proc IEEE Int Conf Acoust Speech Signal Process 2014; 2014: 4858-4862.*
26. *Kobak KA, Engelhardt N, Williams JB, Lipsitz JD: Rater training in multicenter clinical trials: issues and recommendations. J Clin Psychopharmacol 2004; 24: 113-117.*
27. *Low DM, Bentley KH & Ghosh SS: Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investigative Otolaryngology 2020: 5(1); 96–116*

28. Marmar CR, Brown AD, Qian M, Laska E, Siegel C, Li M, et al.: Speech-based markers for posttraumatic stress disorder in US veterans. Depress Anxiety 2019; 36: 607-616.
29. Martínez-Sánchez F, Muela-Martínez JA, Cortés-Soto P, García Meilán JJ os., Vera Ferrándiz JA ntoni., Egea Caparrós A, et al.: Can the Acoustic Analysis of Expressive Prosody Discriminate Schizophrenia? Span J Psychol 2015; 18: E86.
30. 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: 50-66.
31. Mota NB, Vasconcelos NA, Lemos N, Pieretti AC, Kinouchi O, Cecchi GA et al.: Speech graphs provide a quantitative measure of thought disorder in psychosis. PLoS 2012;7(4):e34928.
32. Ozkan H: A comparison of classification methods for telediagnosis of Parkinson's disease. Entropy 2016:18(4);(2016):115
33. Özseven T: Konuşma Tabanlı Duygu Tanımada Ön İşleme ve Öznitelik Seçim Yöntemlerinin Etkisi. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi 2019; 10(1): 99-112
34. Pan W, Flint J, Shenhav L, Liu T, Liu M, Hu B, et al.: Re-examining the robustness of voice features in predicting depression: Compared with baseline of confounders. PLoS One 2019; 14: e0218172.
35. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, et al.: DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. Am J Psychiatry 2013; 170: 59-70.
36. Russell JA: Core affect and the psychological construction of emotion. Psychol Rev 2003; 110: 145-172.
37. Siena FL, Vernon M, Watts P, Byrom B, Crundall D, Breedon P: Proof-of-Concept Study: a Mobile Application to Derive Clinical Outcome Measures from Expression and Speech for Mental Health Status Evaluation. J Med Syst 2020; 44: 209.
38. Siuly S, Alcin OF, Kabir E, Sengur A, Wang H, Zhang Y, et al.: A New Framework for Automatic Detection of Patients With Mild Cognitive Impairment Using Resting-State EEG Signals. IEEE Trans Neural Syst Rehabil Eng 2020; 28: 1966-1976.
39. Tahir Y, Yang Z, Chakraborty D, Thalmann N, Thalmann D, Maniam Y, et al.: Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. PLoS One 2019; 14: e0214314.
40. Tekerek A: Support Vector Machine Based Spam SMS Detection. Politeknik Dergisi 2019;22 (3):779-784
41. Valero X & Alias F: Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification," in IEEE Transactions on Multimedia. IEEE transactions on multimedia 2012; 14(6): 1684-1689.
42. van der Sluis F, van den Broek E, Dijkstra T: Towards an artificial therapy assistant: Measuring excessive stress from speech. In Traver V, Fred A, Filipe J, Gamboa H (eds): Proceedings of the International Conference on Health Informatics, HealthInf, 357-363. INSTICC PRESS, Portugal, 2011.
43. WHO:Mental Disorders. Available from: https://www.who.int/news-room/fact-sheets/detail/mental-disorders. (accessed June 15, 2023)
44. Yazıcı MK, Demir B, Tanrıverdi N, Karaoğlu E, Yolaç P: Hamilton Anksiyete Değerlendirme Ölçeği, Değerlendiriciler Arası Güvenirlik ve Geçerlik Çalışması. Türk Psikiyatr Derg 1998; 9: 114- 117.
45. Young RC, Biggs JT, Ziegler VE, Meyer DA: A rating scale for mania: Reliability, validity and sensitivity. Br J Psychiatry 1978; 133: 429-435.
46. Yünden S: Psikiyatrik Hastalıklarda Ses Analizi. Current Research and Reviews in Psychology and Psychiatry 2022:2(2);201-216

Correspondence:
İlknur Ucuz
Assoc. Prof. Dr.,
Department of Child and Adolescent Psychiatry,
Inonu University Faculty of Medicine Malatya, Turkey,
ilknur_27@yahoo.com, +00905072383095,
ORCID id: 0000-0003-1986-4688.