Information Sciences 588 (2022) 174–195

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Explaining smartphone-based acoustic data in bipolar disorder: Semi-supervised fuzzy clustering and relative linguistic summaries

Katarzyna Kaczmarek-Majer [a,*], Gabriella Casalino [b], Giovanna Castellano [b], Olgierd Hryniewicz [a], Monika Dominiak [c]

[a] Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-147 Warsaw, Poland
[b] Computer Science Dept. University of Bari Aldo Moro, Bari, Italy
[c] Department of Pharmacology and Physiology of the Nervous System, Institute of Psychiatry and Neurology, Sobieskiego 9, 02-957 Warsaw, Poland

ARTICLE INFO

Article history:
Received 26 July 2021
Received in revised form 29 November 2021
Accepted 11 December 2021
Available online 18 December 2021

Keywords:
Linguistic summaries
Semi-supervised fuzzy clustering
Adaptive and evolving algorithms
Fuzzy linguistic descriptions
Explainable Artificial Intelligence
Acoustic markers
Smartphone monitoring
Bipolar disorder

ABSTRACT

Smartphones enable to collect large data streams about phone calls that, once combined with Computational Intelligence techniques, bring great potential for improving the monitoring of patients with mental illnesses. However, the acoustic data streams recorded in uncontrolled environments are dynamically changing due to various sources of uncertainty. In addition, such acoustic data are usually difficult to interpret by psychiatrists. Within this study, we propose an approach based on Linguistic Summaries with Fuzzy Clustering (LS-FC) aiming at the development of human-consistent and easily interpretable summaries about relations between acoustic data and mental state of a patient affected by Bipolar Disorder, e.g., *Most calls in the state of hypomania have low loudness compared to the state of euthymia* [T = 1]. To capture the dynamics of acoustic data streams, we apply a dynamic incremental semi-supervised fuzzy clustering that synthesizes data into clusters. These clusters are represented by prototypes which are used for the construction of the membership functions describing linguistic terms e.g., *low loudness,* and then, linguistic summaries. The main contribution of this paper is the incorporation of information about clusters' prototypes in the generation of linguistic summaries. The primary goal of this research is explainability. The semi-supervised learning algorithm is used mainly for deriving clusters and building improved linguistic summaries. Numerical results indicate that linguistic summaries provide intuitive and clear information about voice features in a patient's affective state and they are consistent with clinical observation. In particular, during most calls in hypomania/mania both the quality of the patient's voice and the dynamics of change in the spectrum signal reflected in spectral flux are low compared to euthymia. The proposed approach enables to summarize large data streams into meaningful descriptions that, although relatively simple, offer information granules that are very intuitive for clinicians and are promising to support the smartphone-based monitoring of bipolar disorder patients to inform about the potential change of mental state.

Abbreviations: XAI, Explainable artificial intelligence; LS, Linguistic summaries; BD, Bipolar disorder; DISSFCM, Dynamic Incremental Semi-Supervised Fuzzy C-Mean Algorithm.

* Corresponding author at: Department of Stochastic Methods at Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland.
E-mail address: k.kaczmarek@ibspan.waw.pl (K. Kaczmarek-Majer).

https://doi.org/10.1016/j.ins.2021.12.049
0020-0255/

1. Introduction

Artificial intelligence with voice analysis is being increasingly used in mental health care and has a great potential for improving the monitoring of patients with chronic recurrent illnesses [1]. In particular, voice data derived from everyday phone calls using smartphones are promising in monitoring the illness activity in bipolar disorder patients [2]. Bipolar disorder (BD) is a chronic, recurrent, highly-morbid illness affecting 2% of the population. In the course of the illness, there are fluctuations between mood states, ranging from euthymia (the healthy state) through depression and hypomanic/manic episodes, as well as mixed states. Hypomania is a state characterized with mild manic symptoms and mania is a state characterized with severe manic symptoms. The initial phases of BD appear to better respond to treatment, thus early intervention strategies could be vital for improving illness outcomes [3] by reducing conversion rates to full-blown illness and reducing symptoms severity. Although sensors mounted in smartphones are able to deliver large data streams, sensor data are only partially labeled, recorded in a naturalistic setting with various sources of uncertainty and hardly interpretable for psychiatrists. Thus, there is a clear need to develop not only technically robust and accurate systems, but also systems able to provide suitable explanations of the reasoning processes.

Currently, the potential of the data collected from sensors is only partially explored in healthcare. AI tools are not only capable to analyze the high quantity of daily-produced sensor data that is impossible to analyze manually, but also have the potential to exploit large quantities of data in order to uncover implicit markers of mental disorder risk, that can sometimes be difficult to notice by the medical expert.

Unfortunately, current AI diagnostic systems are not able to explain the processed data and the reason for their diagnosis. In particular, in the case of BD diagnosis, monitoring, and treatment, one of the main limitations is the lack of understanding of the sensor data and their relation with the affective states of the disease. Without the ability to interpret data, it becomes hard to determine if differences in severity of manic or depressive states relate to differences between characteristics of sensor data. The eXplainable AI (XAI) offers an answer to these problems, turning out to be of fundamental importance in healthcare [4] since it enables medical practitioners to effectively comprehend and validate the output of diagnostic models.

One of the main challenges of the XAI in the medical setting is providing explanations concerning the underlying data in terms of interpretation of features and an understanding of how the presence or absence of some features affects the model's performance. To cope with this issue, a variety of XAI methods have been proposed, broadly grouped in two categories: post hoc methods and ante-hoc methods. Post-hoc XAI methods are specifically designed for explainability and are applied after a model has been created. One of the most used post hoc XAI methods is Local Interpretable Model-agnostic Explanations (LIME) that was developed to explain the predictions of any classifier by calculating feature importance scores based on some assumptions [5]. Ante-hoc XAI methods achieve interpretability without an additional step, since they guarantee a certain level of interpretability during the construction of the model, hence they are immediately interpretable by a human. These include inherently white-box or gray-box models (such as Decision trees and Random Forests) that are typically used to achieve interpretability at the cost of lower performance scores as compared to complex machine learning models.

Among white-box models, fuzzy models offer the advantage of a linguistic knowledge representation which can be cast into a conventional mathematical framework based on fuzzy logic concepts [6]. This enables the development of knowledge-based models capable of both representing highly non-linear input–output relations, and at the same time offering an interpretable view of such relations through the use of linguistic IF-THEN rules. As such, fuzzy models turn out to be fully interpretable and have a great potential in XAI [7]. Moreover, fuzzy models have proved their efficiency in medical decision support and medical reasoning owing to the characteristics of fuzzy sets to describe and represent clinical vagueness and to handle the uncertainty related to medical knowledge, such as symptoms and diseases, that are fuzzy in nature [8].

Despite the wide application of fuzzy models in healthcare, so far mostly their predictive accuracy has been assessed while their potential to explain diagnostic predictions in medical domains has not been fully investigated. In particular, to the best of our knowledge, no attempt has been made to apply fuzzy models to explain through linguistic descriptions various states in BD. In this paper, we propose the use of fuzzy models to derive linguistic summaries useful to explain the relationship between acoustic features extracted from smartphone calls with patients and the state of the disease. This work is the first contribution toward this new research direction. Semi-supervised fuzzy clustering is particularly promising for explaining acoustic data because it enables to capture the information about the hidden structure of an evolving data stream which is sparsely labeled and subject to multiple sources of uncertainty, and this was the main motivation for adopting this approach in this work.

This paper is a continuation of our previous works concerning the development of predictive models to support the monitoring of bipolar disorder starting from acoustic data streams collected by a dedicated phone application [9,10]. In this context continuous monitoring is necessary, but medical visits are performed with limited frequency, hence both labeled and unlabeled data streams are available for the analysis. For this reason, in [11] we have proposed the use of an incremental semi-supervised classification algorithm based on fuzzy C-Means clustering algorithm that is able to adapt the number of clusters as new data arrives [12]. The algorithm has shown its ability to adapt the learned clusters to the new data and capture abrupt concept drifts through a splitting mechanism, leading to relatively high classification performances with low labeling percentage. In these previous works, the dominant criterion for assessing the quality of models has been the accu-

racy of predictions expressed in terms of standard evaluation metrics. However, when predicting the disease state of a patient, the interpretability of the predictive model is of fundamental importance to understanding the decision of the model. While the priority remains on giving accurate predictions, medical experts need to be provided with an explanation of the reason why a given phone call is associated with a specific state of the disease. As the first step in this direction, in [13] we applied visualization techniques to display the outcomes of the semi-supervised clustering, to improve their interpretation, and to highlight patterns in data, that could be used to interpret the disease state of a patient.

In the present work, we take one more step towards explainability and we introduce Linguistic Summaries with Fuzzy Clustering (LS-FC) approach aiming at deriving linguistic summaries using clusters created by DISSFCM to explain acoustic data in BD episodes. While the proposed LS-FC framework makes use of previously existing methods [12,14], it also incorporates novelties in the following respects. Unlike [14], LS-FC is able to adapt to dynamically changing context and linguistic variables are constructed directly with information learned from the considered semi-supervised learning algorithm (DISSFCM). Unlike [12], LS-FC is able to explain the classification results obtained by DISSFCM.

Therefore, the major contributions of this work are as follows.

- LS-FC is a novel approach for linguistic summarization for partially-labeled data streams.
- To the best of authors' knowledge, this is the first work providing linguistic summaries as human-consistent information granules about partially labeled data streams, and the drifts in data streams are reflected in the construction of linguistic variables.
- Furthermore, to the best of our knowledge, this is the first work on linguistic summarization of large speech data streams in the applied context of bipolar disorder monitoring.
- The performance of the applied methods has been evaluated in terms of accuracy and time using out-of-time scenario and prequential-test-then-train validation.
- LS-FC delivers linguistic summaries that have been regarded as informative for physicians who confronted them with the clinical observations and the state-of-the-art.

The structure of the paper is as follows. Section 2 describes the related work about the considered application context of smartphone-based monitoring of bipolar disorder and the related work about the methodology including linguistic summarization and semi-supervised learning for partially labeled data streams. Section 3 describes the main characteristics of acoustic data streams considered in this research. Next, the recently proposed Dynamic Incremental Semi-Supervised Fuzzy C-Means algorithm is formally described in Section 4. The proposed Linguistic Summarization with Fuzzy Clustering algorithm (LS-FC) is presented in Section 5. Finally, experimental results with the discovered linguistic summaries are reported and discussed in Section 6 explaining the usefulness of the proposed approach for the BD application scenario. In Section 7, main conclusions are discussed and future work is outlined.

2. Related Work

2.1. Bipolar Disorder Monitoring

Speech has been already used to diagnose and manage several disorders including Parkinson's disease, Alzheimer's disease and major depression [15]. In particular, reduced speech activity, changes in specific voice features, and pause-related measures were found to be sensitive markers of depressive symptoms [16]. On the other hand, an increased speech activity turned out to predict the manic episodes [17]. The central features of BD are the abnormalities in psychomotor and social activities, typically with psychomotor retardation, social withdrawal, and paucity of speech during the depression and increased motor, social and speech activity during mania. The changes in the manner of speaking reflect the mood state very accurately and are used intuitively by psychiatrists in everyday practice. Considering the possibility of continuous speech data collection via a smartphone app, voice analysis has a great potential in monitoring the affective state in BD patients [18].

Nevertheless, apart from these encouraging results, there is still a great problem with the implementation of this knowledge into clinical practice. Up to date, no tool for clinicians to simply transform results of voice analysis into easily understandable information was created. Further research is needed to bridge the gap between research and implementation of given solutions into clinical settings [19].

2.2. Linguistic Summaries

Fuzzy sets have been proven successful in linguistically summarizing numerical data in various contexts. The main purpose of summarization is usually to improve comprehension of large datasets [20] or even prediction performance [9]. In [9], the authors apply the linguistically quantified sentences, the so-called linguistic summaries to formulate the prior model probabilities concluding that linguistic summaries resulted to be human-consistent information granules and relatively easy for interpretation for experts involved in the forecasting process, e.g., in the pharmaceutical market.

Within this research, linguistic summaries based on protoforms [21,22] are adapted. This methodology was selected due to the simple form of linguistic descriptions that capture the relative characteristics of phone calls. We will not deal here, due

to lack of space, with other approaches to linguistic data summarization exemplified by the mining of IF-THEN or gradual rules, summarization with type-2 fuzzy sets, complex protoforms, e.g., [23], or linguistic summarization with Natural Language Generation. Another example of an alternative approach to summarization was introduced in [24]. The authors aggregated information for multiple time series. Such summaries may be exemplified as follows: *Most days of year 2001, both series exhibit a local change with the same sign* [24]. In particular, this type of summaries seem promising to compare the euthymic and disease episodes at the same time.

The primary advantage of fuzzy linguistic summaries is their human consistency. In [25], Lesot et al. investigate the interpretability of groups of fuzzy linguistic summaries considering the consistency of the sentence set, non-redundancy, and information they provide. Already probabilistic linguistic term sets are confirmed to be useful in a decision-making context, see e.g., [26]. Probabilistic linguistic elements are simple constructions that consist of a term and the degree to which it is certain, e.g., *low height* [T = 0.5]. In this research, we consider more complex forms by combining several linguistic terms sets, however, similarly, the goal is to reflect the hesitation and uncertainty when providing easily interpretable insights about large datasets and relations between various objects. One of the main characteristics of the adapted linguistic summarization in the sense of [22] is that the results are very sensitive to the construction of the fuzzy sets for linguistic terms sets. To alleviate the potential problems related to inadequate construction of fuzzy sets, in [14] the authors introduce personalized linguistic summaries that may be exemplified as *Most outgoing calls in mania state (disease period) are short compared to the calls recorded in the euthymia state (healthy period)*. Within this paper, we improve the construction of the personalized linguistic summaries and dynamically incorporate in summarization the results from clustering. By doing this we take into account the evolutive nature of acoustic data characterizing a phone call and capture changes (concept drift) in the BD episodes of a patient. The main novelty of the proposed approach is to improve the linguistic summarization process by the inclusion of information about data streams derived with the use of fuzzy clustering.

2.3. Semi-supervised learning for partially-labeled data streams

Among the considerable research on data streams, relatively little deals with classification methods suitable for contexts where only some of the instances in the stream are labeled. Most state-of-the-art data-stream algorithms work under the supervised scenario of a fully-labeled stream, where deep learning approaches are demonstrated to be superior to conventional machine learning methods [27,28].

The case of a partially-labeled stream has not been thoroughly addressed by deep learning methods. In [29], deep learning techniques (i.e. Restricted Boltzmann Machines and Deep Belief Networks) are proposed to learn incrementally models from both labeled and unlabeled samples. In [30], the authors investigate the effectiveness of deep learning in medical image analysis with limited quantities of labeled training data. The underlying idea is to assign artificial labels to abundantly available unlabeled medical images and, through a process known as surrogate supervision, pre-train a deep neural network model for medical image analysis tasks lacking sufficient labeled training data. In [31], a strategy to reduce the expert supervision required for deep learning-based segmentation of histopathological images is proposed. The main limitation of such deep learning models is that they have a 'black-box' nature since they are implicitly represented in form of huge collections of numerical parameters that can not be interpreted due to their complex nature. This lack of model transparency is acceptable as soon as accuracy is the major objective in predictive modeling. But when it comes to medical applications, interpretability of results is a major concern.

Fuzzy models have recently received increased attention for data stream modeling since they are intelligible and can realize approximate reasoning to deal with imprecision and uncertainty in decision-making processes. These traits are suitable in healthcare where data streams to be analyzed are affected by uncertainty and should be interpreted qualitatively. One common approach to fuzzy modeling of data streams is through evolving fuzzy or neuro-fuzzy models [32]. In particular, very effective are the methods that are based on ensembles of evolving fuzzy models [33–35]. However, the primary goal of these methods is usually the accuracy of the delivered predictions, not the explainability. Ensembles do not provide a simple way to interpret information about the hidden structure of the dynamically changing data streams. Such information can be provided by another common approach to fuzzy modeling of data streams that is fuzzy clustering. Fuzzy clustering can capture concept drift through continuous membership functions and offers robustness of possibilistic models among others [36]. Several extensions of fuzzy clustering techniques to data streams have been proposed as an efficient and effective way to process data that evolve during time [37,38]. However, these methods are completely unsupervised and do not take into account the availability of labeled data. An attempt to introduce the concept of fuzziness in semi-supervised learning is given in [39,40] where the authors clarify the interrelationship between the fuzziness of a classifier and its generalization ability when using semi-supervised learning. They investigate that, for a given trained classifier whose output is a membership vector, it is possible to calculate the fuzziness of each input sample and use it to improve accuracy. However, the study is performed for stationary datasets and it does not address data stream scenarios.

To deal with partially-labeled data streams, semi-supervised methods based on neuro-fuzzy architectures have been used [41]. In [12], intending to combine advantages of fuzzy clustering and partial supervision for data stream processing, we proposed DISSFCM, a Dynamic Incremental Semi-Supervised FCM algorithm that processes data in form of chunks and evolves the clustering structure by adjusting the number of clusters through a mechanism of low-quality clusters splitting. In the present study, we resort DISSFCM as the basis to derive linguistic summaries that provide interpretable descriptions of acoustic features and their relations with states in Bipolar Disorder.

3. Smartphone-based Acoustic Data

Smartphones with a dedicated application are able to collect sequential data about physical characteristics of speech (voice), and they are called acoustic data streams about phone calls. One of the common libraries for extraction of acoustic data streams is OpenSMILE [42]. The voice signal is divided into smaller frames, e.g., of 20 ms, and it is assumed that within one frame the signal is approximately stationary. Within this research, the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) for voice research [43] was extracted from each frame of the patient's speech during phone calls. The eGeMAPS set comprises 86 features including time-domain descriptors (amplitude statistics, zero-crossing rate, etc.) and spectral features (e.g., spectral flux, mel-cepstral coefficients (MFCC), fundamental frequency (F0), and its harmonics).

Speech formation is regulated by the nervous system. It is considered that in BD, in depressive and manic states, all parts of this system might be altered, thus, acoustic features about speech seem very promising markers for bipolar disorder and are subject of ongoing research [44]. In particular, the changes in the melody and dynamics of speech (known as prosody) have been reported to vary in mood states [17,45]. To date, prosodic features were the most studied voice parameters in mood disorders as they appeared to best reflect the affect and emotions. The prosody is expressed through pitch, energy, and loudness, as well as speech rate and pauses. It has been shown that these parameters could discriminate depressed subjects from controls [15]. Moreover, spontaneous speech could also effectively differentiate manic state from euthymic state [46,47]. Among prosodic features, the most frequently examined feature was a pitch which is reflected in the fundamental frequency (F0) and first formants frequencies (F1, F2). Pitch has been shown to exhibit significant differences between different mood states in BD patients [45]. The most commonly reported finding in the literature in depressed patients is a decrease in fundamental frequency and first formants frequencies reflecting the monotonous speech often seen in depression [48,49]. On the other hand, increased pitch has been correlated with mania [17]. Other acoustic parameters commonly used in voice analysis are those related to voice quality, e.g., jitter. The jitter implies disturbances in the fundamental frequency and is mainly affected by the lack of control of vibration of the cords. Jitter variability tends to increase with depression severity [48,49].

Another extensive group of voice parameters includes a wide set of spectral features. The most studied among them were mel-frequency cepstral coefficients (MFCC). Other spectral parameters (spectral flux, spectral harmonicity) are severely understudied. In particular, spectral flux reflecting the manner and clarity of speech seems very rational to analyze in BD patients. A decrease of the spectral flux is expected in more slurred speech, and the opposite - the increase translates into clearer speech.

Within this research, four groups of acoustic features were manually selected based on the analysis of the related work and the clinical practice. These groups together with the related clinical observation are the following:

- **Loudness-related features** (loudness of speech signal and its energy)
  Patients in affective states are expected to state speak louder compared to euthymia.
- **Pitch-related features** (F0final, F0envelope).
  Patients in affective states are expected to speak with a higher or lower tone of voice (compared to euthymia).
- **Spectral-related features** (spectral flux, spectral harmonicity).
  Patients in affective states are expected to have lower dynamics of changes in the speech signal spectrum.
- **Voice quality-related features** (jitter, shimmer).
  Patients with depressive symptoms are expected to speak less clearly, less fluently, more monotonously (chanting less), the intensity of the voice fluctuates more, they have a more asthenic voice. Patients with manic symptoms speak less clearly, more fluently, chant less, the intensity of the voice fluctuates less.

An alternative approach would be to select the subset of predictive acoustic features with the use of machine learning tools, e.g., Recursive Feature Elimination (RFE). In [50], RFE enabled significant reduction of the number of acoustic features without lowering the accuracy of models. However, their interpretability has been little studied. The present work is one of the first attempts to focus on the interpretability of relation between acoustic features and the mental state of a patient.

4. Dynamic Incremental Semi-Supervised Fuzzy C-Means

Dynamic Incremental Semi-Supervised Fuzzy C-Means algorithm (DISSFCM) is an incremental clustering method that was originally proposed in [12] as an extension of the Semi-Supervised FCM (SSFCM) algorithm [51], which embeds partial supervision in the classical FCM clustering algorithm. SSFCM is designed to cluster static data, hence it assumes a fixed number of clusters. In the case of non-stationary data streams, a fixed number of clusters prevents the model to follow the distribution of classes, resulting in under-fitting, since the distribution of classes may evolve during the time, thus requiring reshaping the clusters. In [12], we therefore extended SSFCM by introducing the ability of evolving the clustering structure also in terms of the number of clusters. The resulting method, called Dynamical Incremental SSFCM (DISSFCM) adjusts the number of clusters by splitting low-quality clusters.

DISSFCM works on a data stream defined as a continuous sequence containing T ($T \to \infty$) data items: $D = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, where $\mathbf{x}_t$ is a data instance observed at time t and described by n numerical features, namely $\mathbf{x}_t \in \mathbb{R}^n$. The data stream struc-

ture is granulated as a sequence of chunks, i.e. $D = X_1, X_2, \ldots, X_t, \ldots$ being X_t a chunk of N_t data instances. Moreover, assuming the semi-supervised scenario, the data stream is composed of labeled and unlabeled instances, hence a true class label may or may not be available for $\mathbf{x}_t$. Each sample belongs to a class in $S_C = \{1, \ldots, C\}$ through a classification function $g : X \mapsto S_C$. Hence, according to the semi-supervised hypothesis, the true classification function is generally known only for some samples. We formalize this hypothesis through a function $b : X \mapsto \{0, 1\}$ such that $b(\mathbf{x}_j) = 1$ iff $\mathbf{x}_j$ is pre-labeled, i.e. its class label $g(\mathbf{x})$ is known, $b(\mathbf{x}_j) = 0$ otherwise.

The clustering process starts with a number of clusters equal to the number of classes (i.e. one cluster for each class). Then, during the incremental application of the method to the subsequent chunks, the number of clusters can grow (due to the splitting mechanism) to become greater than the number of classes (i.e. many clusters are associated with the same class). This requires a mechanism to keep track of the association of each cluster to its class label. To achieve this, DISSFCM integrates a cluster-class mapping mechanism that ensures that each cluster is associated with a class. At the beginning of clustering, a number $K = C$ of pre-labeled samples are randomly selected from the first chunk. These samples will define the initial values of the cluster centers (class distribution is preserved). Being these samples pre-labeled, each cluster center is tagged with a class label that is preserved throughout the optimization process, so that it is always possible to compute the value of f_{jk}.

Given the sequence of chunks, DISSFCM applies the SSFCM semi-supervised clustering to each chunk by minimizing the following objective function [51]:

$$J = \sum_{k=1}^{K}\sum_{j=1}^{N_t} u_{jk}^2 d_{jk}^2 + \alpha \sum_{k=1}^{K}\sum_{j=1}^{N_t} (u_{jk} - b_j f_{jk})^2 d_{jk}^2 \tag{1}$$

where $K \geqslant C$ is the number of clusters, $N_t = |X_t|$ is the cardinality of the t-th chunk in the data stream, $u_{jk} \in [0, 1]$ is the membership degree of a sample $\mathbf{x}_j$ in the k-th cluster, d_{jk} is the Euclidean distance between sample $\mathbf{x}_j$ and center $\mathbf{c}_k$ of the k-th cluster, $\alpha \geqslant 0$ is a regularization parameter for the second part of the objective function that exploits class information, $b_j = b(\mathbf{x}_j)$ and $f_{jk} = 1$ iff the j-th sample has the same class label of the k-th cluster ($f_{jk} = 0$ otherwise). The choice of α depends on the relative number of labeled and unlabeled data [51]. Since the number of unlabeled data is much higher than the number of labeled data, the value of α should be set to equally weigh the two additive components of J; this suggests that α be proportional to the rate N/M where N is the number of data and M denotes the number of labeled data. In this work we fixed $\alpha = 1.0$.

The objective function (1) is minimized subject to the constraint

$$\forall j : \sum_{k=1}^{K} u_{jk} = 1$$

which leads to the following update formulas:

$$u_{jk} = \frac{1}{1+\alpha}\left[\frac{1 + \alpha(1 - b_j \sum_{h=1}^{K} f_{jh})}{\sum_{h=1}^{K} d_{jk}^2 / d_{jh}^2}\right] + \alpha b_j f_{jk} \tag{2}$$

and

$$\mathbf{c}_k = \frac{\sum_{j=1}^{N_t} u_{jk}^2 \mathbf{x}_j}{\sum_{j=1}^{N_t} u_{jk}^2} \tag{3}$$

that are iteratively applied in an alternating optimization scheme, as in [51] to update both cluster centers and the partition matrix.

For each cluster center $\mathbf{c}_k$, a medoid $\mathbf{p}_k$ is derived. Such medoids are regarded as prototypes representatives of data locally processed at each chunk. At each step prototypes are used to reconstruct the data and the ability of the model in reconstructing the original data points is evaluated through the so-called *reconstruction error*, which is defined as follows for the k-th cluster C_k:

$$V_k = \frac{1}{q} \sum_{\mathbf{x}_j \in C_k} \|\mathbf{x}_j - \widehat{\mathbf{x}}_j\|^2 \tag{4}$$

where $q = \sqrt{\sum_{i=1}^{n}\sum_{l=1}^{N_t} |x_{il}|^2}$ and the reconstructed data $\widehat{\mathbf{x}}_j$ is computed as:

$$\hat{\mathbf{x}}_j = \frac{\sum_{k=1}^{K} u_{jk}^2 \mathbf{p}_k}{\sum_{k=1}^{K} u_{jk}^2} \tag{5}$$

If the maximum reconstruction error increases concerning the previous chunk, then the cluster with the highest reconstruction error is split into two new clusters and the resulting prototypes are stored. The splitting method is described in Section 4.1.

The motivation behind using the reconstruction error is the granulation-degranulation strategy [52] whose rationale is to perform firstly a data granulation step (we use prototypes to represent each data point by computing the corresponding membership degrees) and then a degranulation step (we reconstruct the data point using prototypes and membership degrees). Ideally, for a good granulation of data we would expect that the result of the degranulation step should return the original data points, hence the distance (i.e. the error) between the original data point $\mathbf{x}_j$ and it is degranulated (reconstructed) version $\widehat{\mathbf{x}}_j$ as a good measure of the quality of the obtained granulation. In practice, using fuzzy clustering we look for a synthetic representation of data in the form of prototypes and membership values; such a synthetic representation is then used to reconstruct the data to evaluate the quality of the representation. This mechanism has some affinity with the idea of autoencoders that can learn efficient codings (representations) of unlabeled data by a two-step process: the encoding is firstly validated and then refined by attempting to reconstruct the input from the encoding. However, there are some fundamental differences between the granulation-degranulation strategy underlying DISSFCM and the autoencoders. DISSFCM transforms numeric data into fuzzy granules (fuzzy clusters) that are a rich form of data compression since they exploit the concept of partial membership of data to clusters, and the reconstructed data point is obtained as a weighted sum of the membership degrees and the prototypes. The membership degrees of a datum express the extent to which the prototypes should be involved in the reconstruction of each datum, so all prototypes contribute to the decoding process. Indeed, the concept of partial membership is missing in autoencoders. Another main difference is that autoencoders work on unlabeled data (unsupervised learning) while DISSFCM works on partially labeled data (semi-supervised learning).

The labeled prototypes are used to classify data according to a best-matching mechanism. Namely, each data sample is matched against all prototypes (using Euclidean distance) and assigned to the class label of the best-matching prototype. At each time t only the current chunk X_t is analyzed, and the algorithm does not need to store old data. Indeed in a streaming scenario, where data could be infinite, there is the need to process the current data in almost real-time, while old data are discarded. However, to assure transfer knowledge from one chunk to another, the labeled prototypes derived from the previous chunk are incorporated into the next chunk before applying the semi-supervised clustering. The steps of DISSFCM are detailed in Algorithm 1 [12].

Algorithm 1: DISSFCM [12]

Require: Data stream of chunks $X_1, \ldots, X_t, \ldots$ containing some labeled data belonging to C classes
Require: Initial set P_0 of C labeled prototypes
Ensure: set P of K prototypes ($K \geqslant C$) labeled with classes from S_C;
$t \leftarrow 1$
$P \leftarrow P_0$
while $\exists$ nonempty chunk X_t **do**
  $P \leftarrow SSFCM(X_t, P)$
  Compute the reconstruction error $V_{max}^{(t)}$
  while $V_{max}^{(t)} > V_{max}^{(t-1)}$ **do**
    $P \leftarrow split(P, X_t)$
    Compute the reconstruction error $V_{max}^{(t)}$ according to (4)
  end while
  Classify data in X_t using labeled prototypes in P
  $t \leftarrow t + 1$
end while

Summarizing, DISSFCM has the following workflow. When a new chunk is available, the SSFCM clustering is applied to generate labeled prototypes. Prototypes are used to reconstruct the data and the reconstruction error is evaluated. If the reconstruction error increases concerning the previous chunk, then a splitting is applied and the resulting prototypes are stored. The prototypes are then used to classify data in the incoming chunk.

In this work, the labeled prototypes derived by DISSFCM are not only used for data classification (as in [12]) but also for deriving explanations of the acoustic data useful to support monitoring of BD episodes. Indeed, prototypes are the primary information used to create linguistic summaries, as described in Section 5.

4.1. Splitting

When the reconstruction error on the current chunk exceeds the reconstruction error computed on the previous chunk, we deduce that the current number of clusters is not enough to effectively represent the data, hence the number of clusters should be augmented. This is done by splitting one cluster into two parts, to form two new clusters. The cluster has the highest value of the reconstruction error, i.e. the cluster with the lowest reconstruction ability is selected as a candidate for splitting. We denote by k^* the selected cluster. Its splitting is performed using the *Conditional Fuzzy Clustering* [53] applied to the collection of data samples belonging to the cluster to create two novel clusters. Given the set C_{k^*} of data samples belonging to the selected cluster k^*, the splitting process is performed by the minimization of the following objective function

$$J_c = \sum_{k=1}^{2} \sum_{\mathbf{x}_j \in C_{k^*}} v_{jk}^2 \|\mathbf{x}_j - \mathbf{z}_k\|^2 \tag{6}$$

under the constraint

$$\sum_{k=1}^{2} v_{jk} = u_{jk^*}$$

where v_{jk} is the membership degree of $\mathbf{x}_j$ to the cluster $k, k = 1, 2$. The objective function (6) is minimized by iteratively computing the membership values v_{jk} and the prototypes $\mathbf{z}_k$ according to:

$$v_{jk} = \frac{u_{jk^*}}{\sum_{c=1}^{2} \left(\frac{\|\mathbf{x}_j - \mathbf{z}_k\|}{\|\mathbf{x}_j - \mathbf{z}_c\|} \right)^2} \tag{7}$$

and

$$\mathbf{z}_k = \frac{\sum_{\mathbf{x}_j \in C_{k*}} v_{jk}^2 \mathbf{x}_j}{\sum_{\mathbf{x}_j \in C_{k*}} v_{jk}^2}, k = 1, 2 \tag{8}$$

Iterative application of (7) and (8) stops when there is no significant decrease of objective function J_c. Once the new two clusters are generated, they inherit the class label of the original cluster (which is no more used), and membership degrees are re-computed according to (2). The splitting procedure is repeated until the reconstruction error drops below the reconstruction error of the previous chunk.

4.2. Illustrative example

To show the dynamic behaviour of DISSFCM, Fig. 1 plots the evolution of prototypes derived from synthetic chunks extracted from IRIS data. The first chunk contains only data belonging to the class *Setosa* and the algorithm consistently creates only one cluster. In the second chunk both *Setosa* and *Versicolor* classes are present. Data in these classes are well separated, thus the algorithm consistently creates one cluster prototype for each class. The third chunk contains samples belonging to the classes *Versicolor* and *Virginica*, thus a new class appeared and a previously seen class disappeared. In this chunk, data are not completely separated, thus the splitting mechanisms of DISSFCM is activated to better represent the data. Finally, in the fourth chunk, *Versicolor* class disappears, thus the model is adapted to represent *Virginica* data only. This example shows that when new classes appear or some classes disappear as data evolve, the DISSFCM algorithm can add or remove the corresponding prototypes. Moreover, it can be seen that, if needed, more than one cluster prototype is created to represent data belonging to one class.

5. Linguistic Summarization with Fuzzy Clustering (LS-FC)

Linguistic summaries [22] describe with natural language the general facts about the evolution of numerical datasets and are adapted for this study as human-consistent features about acoustic data streams. Let $O = \{o_1, o_2 \ldots, o_N\}$ be a set of objects in a data stream of a considered domain (e.g., phone calls). The properties of objects are measured by a set of attributes $\mathscr{A} = \{a_1, a_2 \ldots, a_r\}$ (e.g., the loudness of speech). Next, linguistic terms set $lst_a = \{l_1^a, \ldots, l_k^a\}$ (e.g., *low*, *high*) are established for each attribute from $\mathscr{A}$. In this work, to take into account the uncertainty of acoustic features, we propose the use of fuzzy sets to describe acoustic features and derive linguistic summaries. The following steps of the proposed Linguistic Summarization with Fuzzy Clustering (LS-FC) approach are implemented.

1. Fuzzification of acoustic features using clusters' prototypes learned by DISSFCM algorithm.
2. Granulation of the acoustic features.

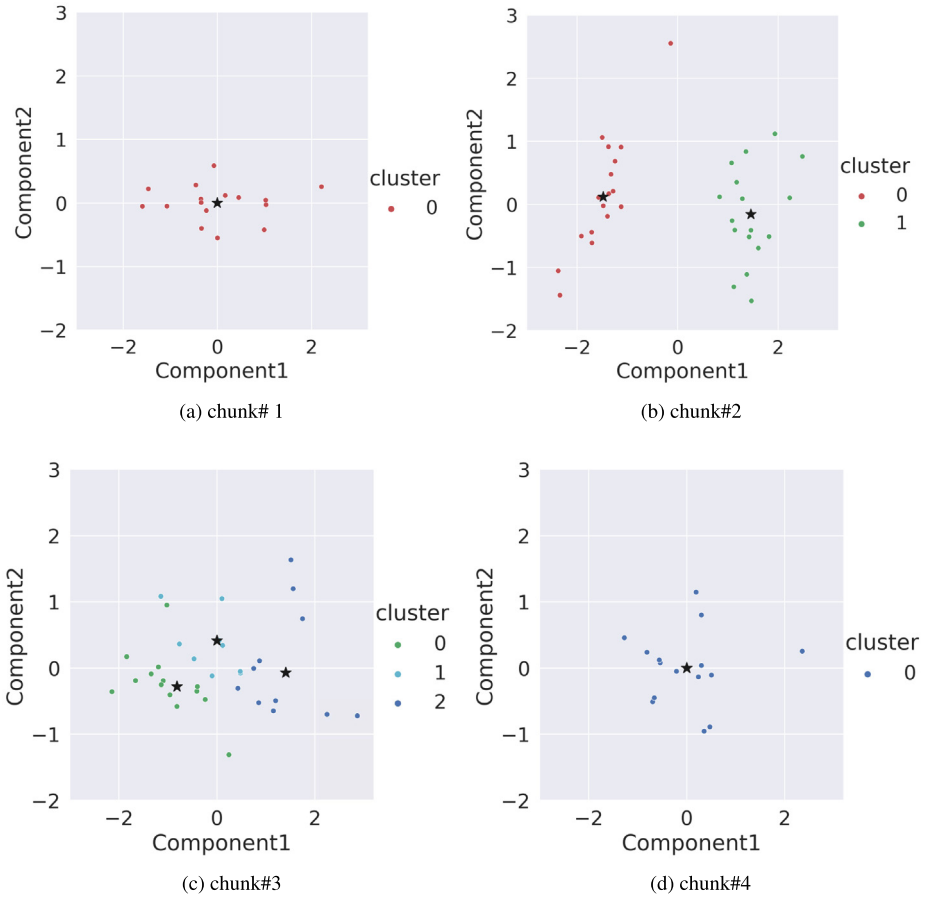


Fig. 1. PCA visualization of four chunks obtained from IRIS data and prototypes (denoted by a ☆) resulting from DISSFCM..

3. Derivation of linguistic summaries.

In the following, details for each step are given.

5.1. Fuzzification of acoustic features using clusters' prototypes from DISSFCM

To provide a qualitative interpretation of acoustic features through linguistic summaries, we firstly represent them as fuzzy variables, whose values are fuzzy sets associated with linguistic terms such as *low/medium/high* (e.g., *low loudness*). Fuzzy sets represent vague descriptions of objects and are mathematically formulated as an extension of boolean sets [6]. Formally, a fuzzy set A is represented by a membership function defined on a universe of discourse X as:

$$\mu_A : X \rightarrow [0, 1]$$

where A is the fuzzy label or linguistic (value) term describing the variable $x \in X$. As an extension to boolean logic, $\mu_A(x)$ represents the grade of membership of x belonging to the fuzzy set A. A fuzzy variable can be described by many different adjectives each with its own fuzzy set. It is clear that the definition of fuzzy sets is non-unique for the nature of language, but it is very context-dependent and user-specific. On specifying a membership function $\mu_A(x)$ in its present context the vague fuzzy label A is precisely defined. Hence fuzzy sets can be thought of as measuring the inherent vagueness of language precisely. In this work, we exploit these properties of fuzzy sets that play an important role to achieve explainability through linguistic summaries.

According to [22], linguistic summaries in Yager's sense are an intuitively appealing and powerful tool for obtaining relatively easy to use, even for novice users, the results of data analyses and mining. Their generation is not trivial and various methods can be employed. In [14], it is shown that static linguistic variables constructed for all patients in different affective states usually result ineffective.

To alleviate the problem of inadequately defined linguistic variables, in this work, cluster prototypes resulting from DISSFCM are used to construct the membership functions of fuzzy sets (linguistic terms) describing the acoustic features.

DISSFCM algorithm enables to assign the predicted class to the dynamically created clusters. Prototypes of clusters predicted as euthymia are considered for calculating membership functions of the linguistic terms, and this is the primary novelty of the proposed LS-FC approach. Formally, to get insight into the healthy state we collect the cluster prototypes extracted from chunks of the acoustic data stream and select the subset P_e of prototypes that are assigned to the euthymia (healthy) class according to the DISSFCM algorithm. Next, for each acoustic feature $x_i, i \in \{1, 2, \ldots, n\}$ type-I fuzzy sets represented by trapezoidal membership functions are constructed to reflect the linguistic terms. To calculate the parameters of membership functions, the following three quartiles are computed using prototypes in P_e: the first quartile (0.25) denoted by q_i, the second quartile (0.5) denoted by b_i and the third quartile (0.75) denoted by c_i. The quartiles q_i, b_i and c_i are used to define membership functions of the three linguistic terms (fuzzy sets) *low, medium, high* as follows:

- Membership functions for *low* terms are z-shape fuzzy numbers and are characterized by the two parameters q_i and b_i.
- Membership functions for *medium* terms are triangular fuzzy numbers that are characterized by the three parameters q_i, b_i, and c_i.
- Membership functions for *high* terms are s-shape fuzzy numbers that are characterized by the two parameters b_i and c_i.

Fig. 2 presents membership functions for an exemplary feature - level of loudness.

5.2. Granulation of acoustic features into attributes

Once acoustic features are represented by fuzzy sets, we granulate them into high-level concepts representing attributes of a call. Specifically, starting from the acoustic features characterizing a call, the following high-level attributes are derived for each call:

1. **loudness**, that granulates loudness-related features (i.e., level of loudness and logarithm of the signal energy).
2. **pitch**, that granulates pitch-related features (i.e., F0, F0final, F0envelope).
3. **spectrum**, that granulates spectral-related features (i.e., spectral flux, spectral harmonicity, spectral centroid)
4. **quality_of_voice**, that granulates quality of voice-related features (i.e., jitter, shimmer, loghnr)

These high-level attributes are derived by applying a t-norm operator to fuzzy sets describing features in each group. In experiments, minimum and Łukasiewicz t-norms are applied and compared. Table 1 presents an illustrative example for calculation of one of the high-level attribute: loudness that groups two acoustic features: energy and the level of loudness.

Values of *loudness low* are calculated using a minimum t-norm in this context, e.g., the loudness of call#2 is calculated as min(1,0.21)=0.21.

5.3. Derivation of linguistic summaries

Starting from attributes and their description through fuzzy terms, we specify the underlying structure of linguistic summaries (LS), also called fuzzy quantified sentences, using short and extended protoforms [21].

Specifically, we construct the linguistic summaries based on the short and extended protoforms. Short linguistic summaries explain general patterns in data. For example, *In the state of mania, most calls have low loudness compared to the state of euthymia [T = 0.8]*. These summaries, although relatively simple are very intuitive for medical experts and easy to compare to the state of the art in this field. The linguistic summaries based on the extended protoforms enable to capture relations within and between the groups of parameters. For example, *In the state of mania, most calls have low loudness have low tone compared to the state of euthymia [T = 1]*. Following [54] we implement the summarization by the tree search algorithm with tree nodes corresponding to linguistic terms sets lst_a.

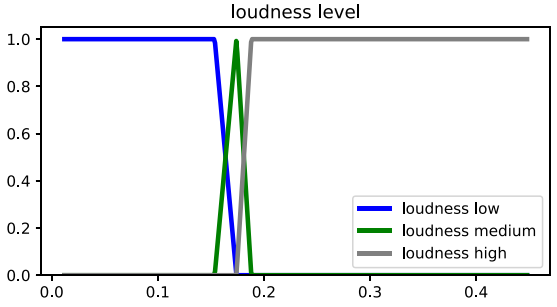


Fig. 2. Linguistic variable *loudness* of speech signal and fuzzy numbers describing its linguistic terms.

Table 1
Example of high-level acoustic feature *loudness* being *low* about phone calls that granulates energy and the level of loudness. Minimum is applied as t-norm.

Call	energy	level	*energy_low*	*level_low*	*loudness_low*
1	0.01	0.2	0	0	0
2	0.005	0.16	1	0.21	0.21

To capture the difference between data collected in the two states healthy and disease (denoted by S_H and S_D, respectively), we leverage relative summaries. A relative linguistic summary based on the short protoform LS_{rs} takes the form:

'Among all objects in state S_D, Q have P compared to the state S_H [T]'.

A relative linguistic summary based on the extended protoform LS_{re} takes the form:

'Among R objects in state S_D, Q have P compared to the state S_H [T]'

where Q is the quantifier (the amount determination, e.g., *most*, *minority*); R is a qualifier (attribute together with an imprecise label) about objects $o \in O$ in state S_D and P is the summarizer (attribute together with an imprecise label, e.g., calls having *low loudness*) about objects $o \in O$; and $T \in [0, 1]$ measures the quality of the summary (level of confidence).

Let $\mu_R, \mu_P, \mu_Q : Re \rightarrow [0, 1]$ be the membership functions of fuzzy sets representing the qualifier R, summarizer P and quantifier Q, respectively, $\wedge$ is a t-norm. To compute the degree of truth T we consider the Zadeh's degree of truth (validity) (T) of the summary defined as follows:

$$T_{rs} = \mu_Q\left(\frac{1}{N}\sum_{i=1}^{N}\mu_P(y_i)\right) \tag{9}$$

$$T_{re} = \mu_Q\left(\frac{\sum_{i=1}^{N}\left(\mu_R(y_i) \wedge \mu_P(y_i)\right)}{\sum_{i=1}^{N}\mu_R(y_i)}\right) \tag{10}$$

We chose this definition of degree of truth since, according to [55], it is used in most of the theoretical studies on linguistic summarization.

6. Experimental Results

6.1. About BIPOLAR Data Streams

Experiments are performed on acoustic data streams BIPOLAR collected in a prospective study with bipolar disorder patients[1]. BIPOLAR dataset contains 10 attributes describing the acoustic features related to loudness, pitch, spectrum and quality of voice (see Section 3). BIPOLAR refers to a 3-class classification problem, and classes correspond to the following three BD states:

- euthymia that is the healthy state,
- hypomania that is a state characterized with mild manic symptoms,
- and finally mania that is a state characterized with severe manic symptoms.

The smartphone-based data collection was performed during the everyday life of a patient. However, the number of labeled data is reduced only to days around the psychiatric assessments that were performed occasionally. The ground-truth for the analysis was assumed to be 7 days before the clinical visit and 2 days after the visit, similarly as the in the work of [56]. Datasets considered in the experiments are completely labeled and the lack of labels is simulated by systematically removing labels for every h samples (e.g., $h = 2$ for 50% labeling).

Acoustic parameters were calculated for frames of 20 ms and a sequence of frames forms a data stream. We can distinguish phone calls of different lengths in it. The interlocutor speech was removed at the data preprocessing stage according to the study protocol. As described in Table 2, the number of available frames extracted from the patient speech varies from 83.7 K for the state of hypomania (3 days of recordings available) to 673 K of frames for the state of mania (9 days of recordings available in the considered ground-truth).

Validation of the proposed semi-supervised algorithm is performed in two ways. First, prequential-test-then-train validation was performed for the considered data stream on selected benchmarks to show its performance in terms of accuracy and time. Secondly, we perform the out-of-time validation reflecting the real-life context of predicting a change in bipolar disorder.

Table 2
About datasets with acoustic data streams. *SP* informs whether speech problems were clinically assessed according to the Young Mania Rating Scale (YMRS).

Dataset	SP	days	all data	patient speaking data
Mania (M)	yes	9	2 571.2	673 K
Euthymia (E)	no	6	164.5 K	680.4 K
Hypomania (HM)	no	3	370.7 K	83.7 K

6.2. Evaluation of DISSFCM towards baseline semi-supervised approaches on benchmark streaming datasets

To assess the effectiveness of DISSFCM, we compared it against WeScatterNet, that is a recent semi-supervised algorithm for data stream mining proposed in [35] as the semi-supervised and distributed improvement of Scalable PANFIS [57] for big data stream problems. Besides BIPOLAR dataset, we considered six popular big data stream problems, also used in [35] for numerical comparison, namely Higgs, Hepmass, RLCPS, Susy, KDDCup, and PokerHand. Train-and-test prequential evaluation was performed. In this particular bipolar disorder monitoring context, the number of labeled data is low because psychiatric assessments confirming the mental state of a patient occur every few weeks or months depending on the patient's need. For this reason, we studied the influence of the labeling amount by considering four percentages of labeled data, namely $25\%, 50\%, 75\%$, and 100%. To this aim, only labeled data points were considered and the lack of labeling was simulated. Table 3 shows the comparative results in terms of time and average accuracy per batch (chunk).

It should be noted that the comparison in terms of training and testing time is not completely fair, since WeScatterNet was operating under distributed computing platform of Apache Spark with model fusion strategy, while DISSFCM was run on Google Colaboratory[2] platform. However, it can be seen that DISSFCM is faster than WeScatternet at test time. This is thanks to the very light structure of the classification model learned by DISSFCM, which is represented only by a few labeled prototypes. As concerns accuracy, compared to WeScatterNet, DISSFCM achieves comparable values for Hepmass and PokerHand datasets, lower accuracy values for RLCPS dataset, and higher values for the remaining datasets. It is worth notice that as for WeScatterNet, the labeling percentage does not influence the average accuracy per batch, except for KDDCup dataset where a 20% decrease is observed with low labeling percentages (i.e. 50% and 25%), and a slight decrease in performance is observed for BIPOLAR data from the completely supervised scenario (100% labeling) to the semi-supervised ones.

We conclude that DISSFCM provides competitive results in terms of accuracy and testing time, and there is potential for improvement in terms of training time. In further analysis of the BIPOLAR data streams, we consider the DISSFCM algorithm since it provides satisfactory accuracy and additional information about dynamically created cluster prototypes that are used to support linguistic summarization.

6.3. Out-of-time validation of DISSFCM

In this section, we evaluate the accuracy of the DISSFCM approach in the out-of-time validation. Data streams are grouped according to calendar days and create chunks of data streams collected from 2 consecutive calendar days, as shown in Fig. 3 that illustrates the classes contained in each chunk. It can be seen that there is a class change at chunk 5 (from mania to euthymia) and chunk 8 (from euthymia to hypomania).

Given the sequence of chunks, we apply DISSFCM to learn the hidden and evolving structure of the acoustic data streams in form of cluster prototypes and dynamically assign class labels to the cluster prototypes. The four previously described labeling percentages have been considered. DISSFCM was applied to the sequence of chunks for each labeling percentage.

For each chunk, 70% of data were used as a training set and the remaining 30% as a test set. Results in terms of standard classification measures and visualization techniques are shown only for chunk#5 and chunk#8 that consisted of phone calls taken in both states (healthy and disease), namely they contain two classes. Classification measures on the other chunks are not shown since they include data from just one class. Table 4 and Table 5 show the classification results for chunk#5 and chunk#8. It can be seen that the labeling percentage strongly affects the results for chunk#8 whilst low differences are observed for chunk#5. This was due to strong class unbalancing in chunk#8. Indeed, the hypomania class is highly over-represented concerning the minority class euthymia, as shown in Fig. 3. Moreover, from Table 5 we can observe that when labels are available (75% and 100% labeling), the disease classes mania or hypomania are better recognized (high recall values) than the healthy class euthymia. On the contrary, when the number of available labels decreases, the healthy condition becomes the easiest to recognize. Maybe this is due to the different geometrical structures of the classes. When labels are available, the algorithm uses them to derive the model. On the contrary, when labels are not available, or they are only partially available, DISSFCM relies on the geometrical structure of data. These results suggest that frames belonging to the euthymic state share some patterns that are captured by the algorithm. Very low precision values are observed for the healthy class in chunk#8, suggesting a high number of false-positives (also due to the low number of frames belonging to

[1] The study was conducted in the Department of Affective Disorders, Institute of Psychiatry and Neurology in Warsaw, Poland within the project entitled "Smartphone-based diagnostics of phase changes in the course of bipolar disorder". which included patients diagnosed with bipolar disorder (according to ICD-10 classification).

[2] Google Colaboratory:https://colab.research.google.com/

Table 3
Performance evaluation of DISSFCM and WeScatterNet for BIPOLAR and benchmark data streams.

Dataset	#classes	#obs	#batch	Algorithm	Labeling per batch (%)	Average accuracy (%)	Training time per batch (s)	Testing time per batch (s)
Bipolar	3	921 K	25	DISSFCM	100	83.75	188.42	0.06
					75	79.17	121.53	0.04
					50	79.17	114.17	0.04
					25	79.17	121.22	0.04
				WeScatterNet	100	71.95	19.81	1.72
					75	71.95	7.23	1.63
					50	71.95	4.27	1.61
					25	71.95	2.91	4.65
Higgs	2	11,500 K	198	DISSFCM	100	96.48	128.52	0.10
					75	96.43	161.82	0.45
					50	96.40	162.79	0.10
					25	96.38	175.62	0.10
				WeScatterNet	100	63.62	9.92	5.25
					75	63.59	10.2	6.14
					50	63.47	8.69	5.67
					25	63.26	6.41	5.1
Hepmass	2	11,000 K	189	DISSFCM	100	98.86	130.34	0.10
					75	98.77	153.07	0.10
					50	98.73	159.98	0.10
					25	98.68	206.79	0.10
				WeScatterNet	100	83.54	12.12	5.19
					75	83.49	12.68	5.79
					50	83.48	11.28	2.91
					25	83.45	9.21	2.74
RLCPS	2	5,000 K	90	DISSFCM	100	47.09	267.62	0.07
					75	47.06	225.89	0.08
					50	47.06	1269.68	0.35
					25	47.05	1147.14	0.08
				WeScatterNet	100	99.64	11.16	-
					75	99.64	10.41	1.92
					50	99.64	10.23	2.07
					25	99.64	9.11	1.89
Susy	2	5,000 K	90	DISSFCM	100	96.76	109.71	0.08
					75	96.68	129.56	0.08
					50	96.65	141.37	0.08
					25	96.58	196.39	0.07
				WeScatterNet	100	75.05	10.03	4.33
					75	75.24	10.16	4.63
					50	75.29	9.31	2.39
					25	75.7	6.96	2.25
KDDCup	2	4,898 K	87	DISSFCM	100	97.63	182.69	0.94
					75	88.23	267.68	0.12
					50	66.67	336.63	0.12
					25	66.67	496.79	0.12
				WeScatterNet	100	99.55	17.57	3.51
					75	99.58	16.39	3.55
					50	99.50	15.38	3.44
					25	99.41	13.86	3.68
PokerHand	10	1,025 K	18	DISSFCM	100	56.00	452.68	0.09
					75	55.95	312.42	0.08
					50	55.89	222.05	0.08
					25	55.87	183.17	0.08
				WeScatterNet	100	50.11	11.75	3.41
					75	50.11	9.38	3.4
					50	50.13	7.49	3.48
					25	50.11	9.33	2.99

that class). However, it should be noted that acoustic data for bipolar disorder are very complex and labels assigned by the experts have an intrinsic uncertainty. Previous works on fully labeled datasets, analyzed in batch mode, have shown average accuracy values of about 0.6.

6.4. Prototypes about acoustic data learned from DISSFCM

Visualization techniques have been also used to evaluate the algorithm capability to adapt the model to changes in data. To this purpose Principal Component Analysis has been used to represent data in 2-dimensional space. Fig. 4 shows the original distribution of the classes for chunk#4, #5, #6 and #8, with 100% labeling (first column) and the cluster prototypes

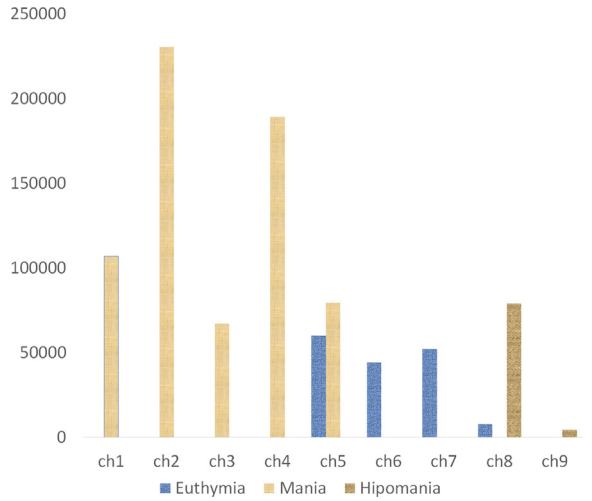


Fig. 3. Classes distribution through chunks.

Table 4
Accuracy values on test sets for chunks #5 and #8, varying the labeling percentages 25%, 50%, 75%, 100%.

Chunk	**Labeling 25%**	**Labeling 50%**	Labeling **75%**	Labeling **100%**
#5	0.45	0.55	0.58	0.58
#8	0.40	0.44	0.77	0.79

Table 5
Recall and *Precision* values on test sets for disease (D) and healthy (H) conditions, for chunks #5 and #8, varying the labeling percentages 25%, 50%, 75%, 100%. Note that for chunk #5: D is mania, whilst for chunk #8 D is hypomania.

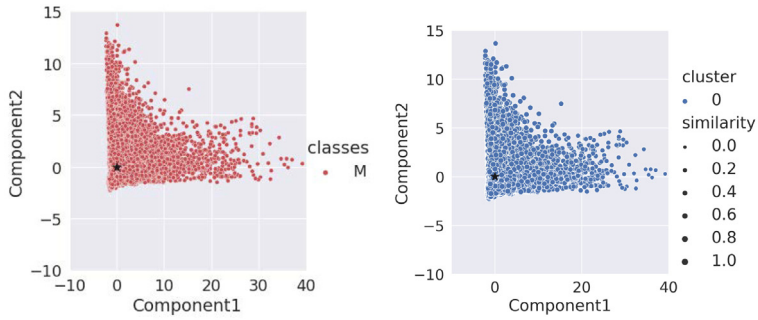
Recall

	Labeling **25%**		Labeling **50%**		Labeling **75%**		Labeling **100%**	
Chunk	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*
#5	0.64	0.30	0.68	0.44	0.45	0.66	0.40	0.71
#8	0.64	0.37	0.64	0.42	0.56	0.79	0.55	0.81

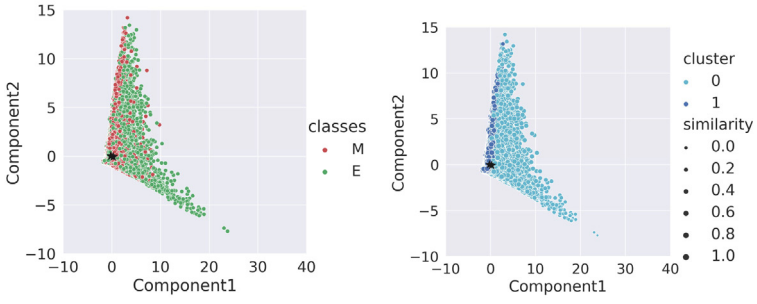
Precision

	Labeling **25%**		Labeling **50%**		Labeling **75%**		Labeling **100%**	
Chunk	*H*	*D*	*H*	*D*	*H*	*D*	*H*	*D*
#5	0.41	0.52	0.48	0.65	0.50	0.62	0.51	0.61
#8	0.09	0.9	0.09	0.92	0.21	0.95	0.22	0.95

returned by DISSFCM (second column). Cluster prototypes are represented by stars, and the size of the dots represents the membership of data to the cluster it has been assigned. A subset of chunks has been selected to show how the algorithm is able to dynamically adapt the model according to the data distribution, even when abrupt changes occur. We can observe that chunk#4 contains only samples belonging to the class mania. Conversely, chunk#5 contains both mania and euthymia samples (red and green). The algorithm creates a new prototype to describe the frames belonging to this class. Then, in chunk#6 a new change occurred, indicating that the patient moved from a mania state to the healthy one (euthymia). All the frames belong to this state. Again, DISSFCM dynamically adapts to the new data by removing the prototype that is no more used. In this example the number of classes and clusters is equal, however, if necessary, more than one cluster could be used to describe a single class. Then, a new change occurs in chunk#8 and the hypomania class appears. Differently from chunk#5 where the two classes were not completely overlapped, frames belonging to hypomania are higher in number, than those belonging to euthymia, and the two classes are quite completely overlapped. Euthymic data are localized on the left part of this Figure. Since these data are under-represented, there are no sufficient labels to construct the model. In this case, the unsupervised part of the optimization function prevailed and the splitting mechanism was activated in order to create

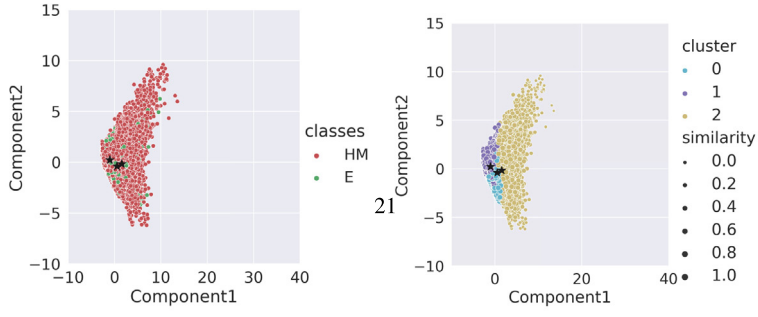


(a) chunk#4



(b) chunk#5

(c) chunk#6



(d) chunk#8

Fig. 4. PCA visualization of training data for chunks #4, #5, #6 and #8, with 100% labeling. In the first column the original distribution of classes in data. In the second column the models derived by DISSFCM. Prototypes are denoted by a ☆, whilst the size of dots indicates their membership to the cluster they have been assigned.

more clusters to describe the data structure. Indeed, clusters 1 and 2 were used to describe hypomania data, whilst cluster 0 for euthymia.

From Fig. 4 we can observe that the distribution of data in different chunks changes. It is the case of mania data frames (red dots) that in the first chunk are spanned in the top-most part of the Figure, whilst in the second chunk they are

concentrated in the top-left part. Similarly, euthymic data completely changes their distribution from chunk#2 to chunk#3. Nevertheless, DISSFCM was able to capture these changes and to evolve its model according to them. Moreover, Fig. 5 shows heatmap visualizations of the prototypes plotted in Fig. 4. Prototypes summarize data by capturing patterns that can be used to identify the corresponding patient state. Indeed, a prototype of chunk#4 describes mania data, the first prototype for chunk#5 describes mania data and the second euthymia, chunk#6 contains only euthymia data, whilst for chunk#8 clusters 1 and 2 are used to describe hypomania data, and cluster 0 for euthymia. Colors are used to highlight patterns in data. It is also observed that feature spectral centroid assumes relatively high values.

Complementary graphs are shown in Fig. 6. Swarmplots represent data distribution for each feature for a given cluster, chunk and labeling percentage. Fig. 6 describes data belonging to clusters 0 and 1 of chunk#5, with 100% of labels. Colors are used to represent the original classes in data. The two clusters are not able to completely separate the two classes, as was suggested by the quantitative evaluation. However, we can observe that the feature *spectralharmonicity* has high variance, and some outliers as the feature *energy*.

Even if these graphical representations can help to understand the results, they are still relatively difficult for clinicians. For this reason, linguistic summarization is subsequently performed taking advantage of the information about obtained clusters and their prototypes.

6.5. Linguistic Summarization

Starting from prototypes derived by the DISSFCM, we create the linguistic summaries according to the approach described in Section 5. The first step is to construct membership functions for the linguistic terms sets describing the acoustic parameters. Fig. 7 illustrates the fuzzy numbers representing low-level acoustic data. It needs to be noted that most of the observed acoustic data come from distributions that have long tails. It can be also observed that the fuzzy numbers representing *medium* term resulting from the quartiles of prototypes are relatively narrow. It needs to be noted that for the considered data, for most chunks, only one cluster was assigned for each class.

Additionally, quantifiers *most, minority, about half, some* were defined as fuzzy numbers. Next, we linguistically summarize the results to see whether there are differences between data identified as healthy and the considered disease states (mania and hypomania). Considering that we have four high-level features and three linguistic terms, 12 linguistic

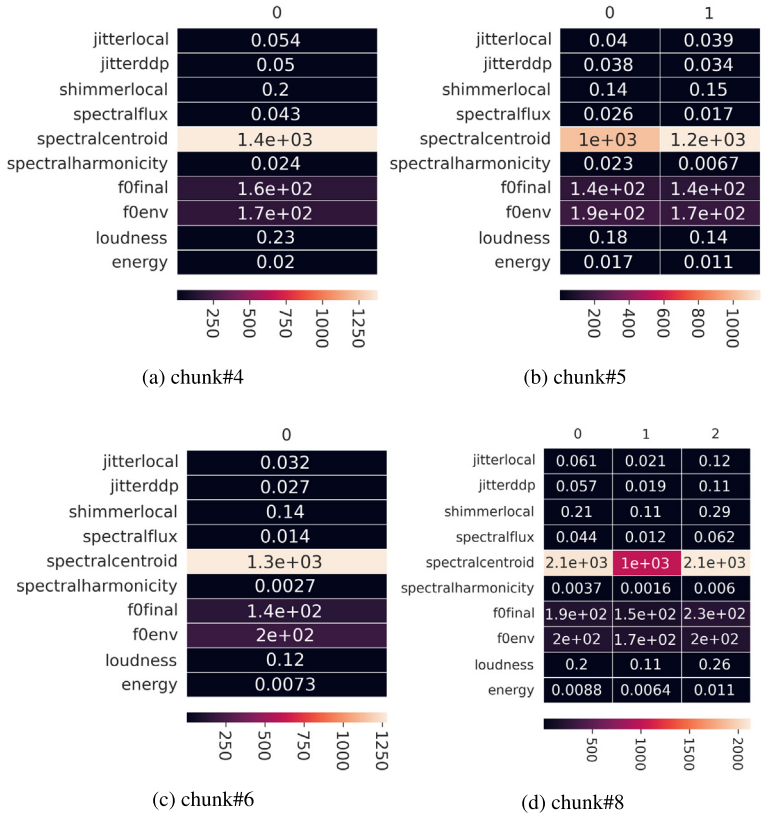


Fig. 5. Heatmap visualization of the prototypes of chunk#4 (0 = M), chunk#5 (0 = E 1 = M), chunk#6 (0 = E) and chunk#8 (0 = E 1 = HM 2 = HM), with 100% labeling.

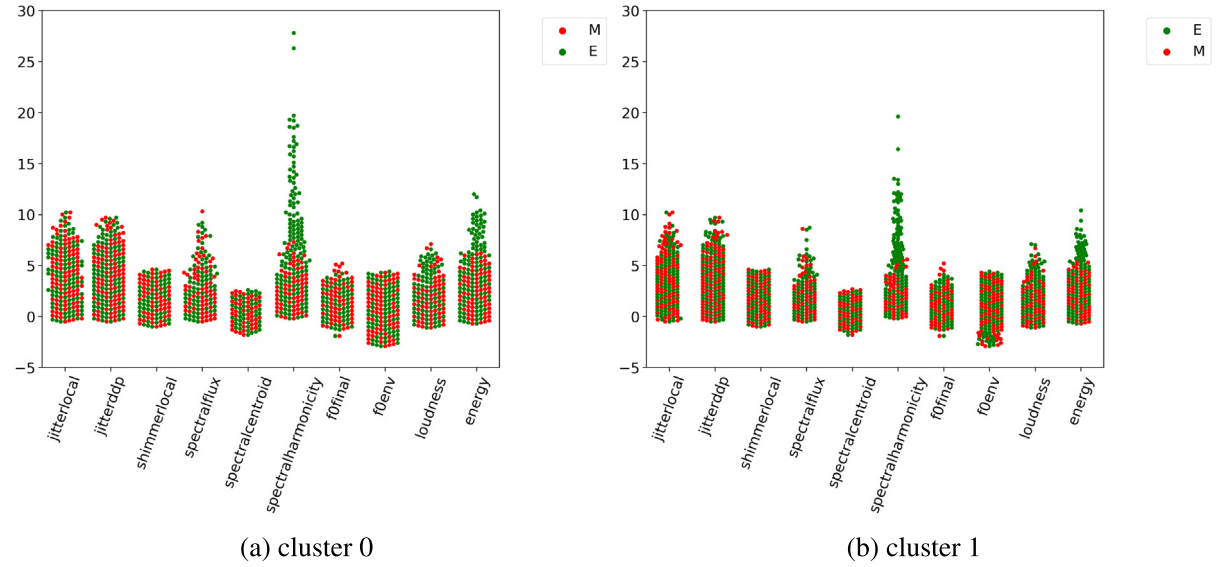


Fig. 6. Swarmplot visualization of original classes distributions in training set data belonging to the two clusters in chunk#5, with 100% labeling.

summaries based on the short protoform were generated for each quantifier and each of the two affective states, namely mania, and hypomania. Similarly, 108 (12 x 3 x 3) linguistic summaries were generated based on the extended protoform for each quantifier and each of the affective states. Results for the quantifier *most* defined as the trapezoidal number [0.45, 0.6, 1, 1] are presented in the remaining of this Section since this quantifier was most preferred by the clinicians due to its informativeness.

Table 6 gathers in the upper rows relative summaries in mania and hypomania based on short protoforms for which the degree of truth equals 1. Prototypes for the labeling percentage of 100% were applied. Next, linguistic summaries based on extended protoform in mania and hypomania are also gathered in Table 6. As observed in Table 6, there are two types of summaries that are true in both considered disease states, informing about low spectrum and low quality compared to the state of euthymia. Interestingly, for the state of hypomania, most calls have low loudness. Then we analyze the relative extended linguistic summaries to understand further relations between high-level parameters.

We can further learn that e.g., *most calls with high loudness in hypomania have a high spectrum compared to the state of euthymia.* This information is complementary to the linguistic summaries based on the short protoforms. Indeed, the extended protoforms enable retrieval of less frequent patterns. It is also observed that *most calls with a low pitch in hypomania have a low spectrum compared to the state of euthymia.* Interestingly, most of the summaries for the manic state are consistent with the hypomanic state, but there are some exceptions, e.g., it is observed that *most calls with high loudness in mania have low spectrum compared to the state of euthymia.*

6.6. Discussion

The experimental results indicate that summaries derived by the proposed summarization approach provide the clinicians with very clear and usable information about the patient's affective state. Since certain speech properties seem to change with the phase change, the psychiatrist could be informed that, for example, the patient is no longer in euthymia state. This is a basic and extremely important conclusion that can be derived from the voice analysis along with the linguistic summaries. For example, based on the results of this study, it can be assumed that during most calls in hypomania/mania state, both the quality of patient's voice (reflected mainly in jitter and shimmer parameters) and the dynamics of change in the spectrum signal (reflected in spectral flux) are low compared to euthymia. In other words, a patient in hypomania/mania speaks less clearly, speech is more slurred and the intensity of the voice fluctuates less. This is consistent with clinical observation, as these patients also tend to speak faster. This is also in line with the available, limited literature as it was found that voice quality differed significantly between depressive and manic states in BD [46]. Surprisingly, the obtained results have also indicated that during most calls in hypomania, the patient's speech is quieter than in euthymia. To date, loudness-related parameters have been rarely studied, we have identified only two studies evaluating these features [58,46]. They brought inconclusive results. Thus, it is unclear whether they are capable of distinguishing affective states in BD. Nevertheless, clinical observations indicate that patients in hypomania tend to speak simply faster, while the loudness increases as manic symptoms increase up to full-blown mania.

Also, the results of linguistic summaries based on extended protoforms turn out to be informative for the clinician. It is possible to infer from their conclusions about the possible association of certain speech parameters and hence presume a

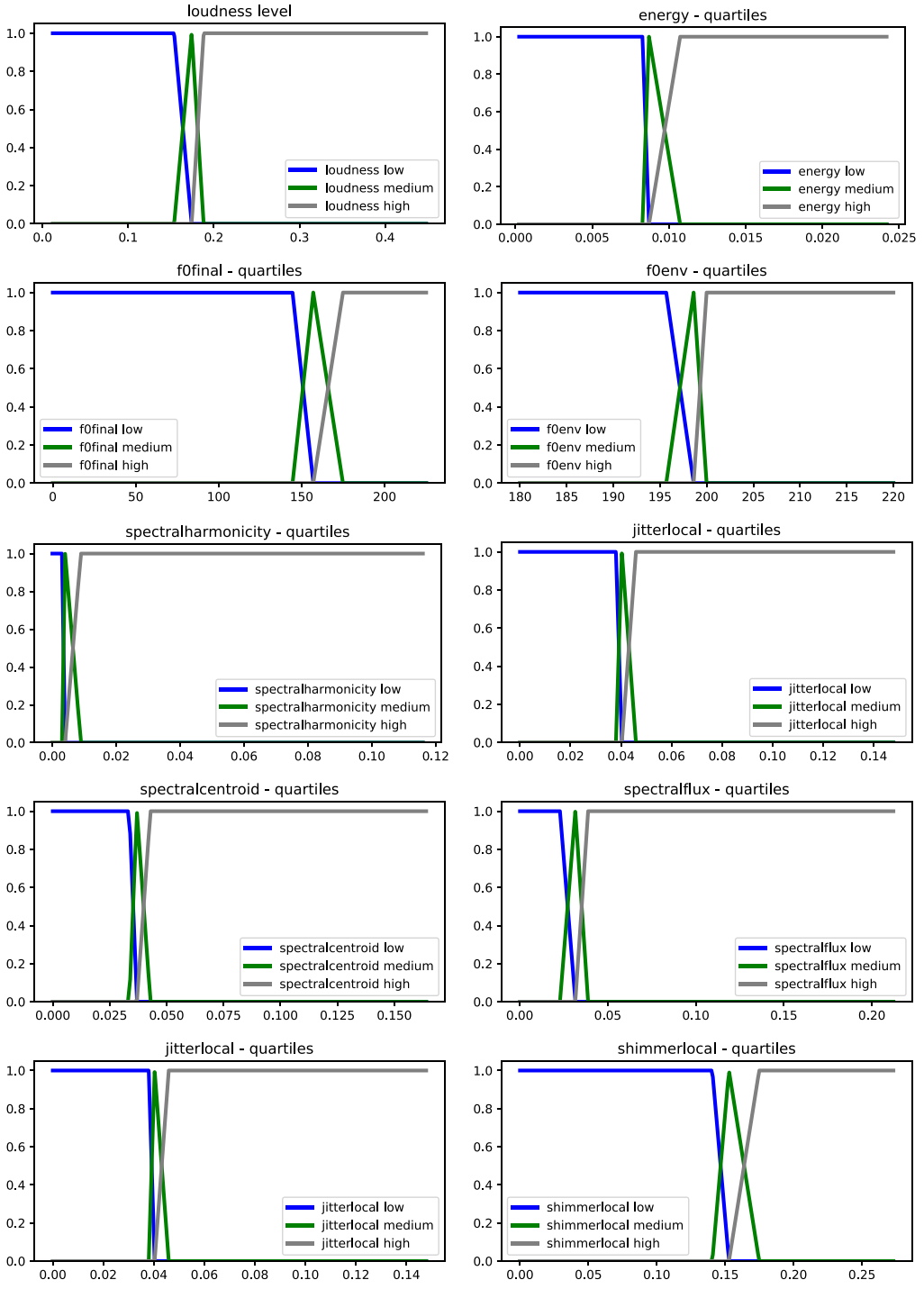


Fig. 7. Linguistic variables and their fuzzy sets derived from prototypes. Prototypes derived assuming 100% labeling percentage.

given mental state. It seems very intuitive to observe that if a patient speaks loudly in most calls in mania, he/she would probably also tend to speak more clearly and the intensity of the voice will fluctuate less. The numerical results of this study have demonstrated such associations. This can make it very easy for mental health professionals to identify an episode, especially one that occurred some time ago, between visits.

Table 6
Relative linguistic summaries based on short protoforms for mania and hypomania episodes (LS with $T = 1.0$) and extended protoforms for mania and hypomania episodes (LS with $T > 0.5$).

Relative LS based on short protoform	*T*
Most calls in the state of mania have low spectrum compared to the state of euthymia.	1.0
Most calls in the state of mania have low quality compared to the state of euthymia.	1.0
Most calls in the state of hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls in the state of hypomania have low loudness compared to the state of euthymia.	1.0
Most calls in the state of hypomania have low qualty compared to the state of euthymia.	1.0
Relative LS based on extended protoform - HYPOMANIA	*T*
Most calls with low loudness in hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls with low loudness in hypomania have low quality compared to the state of euthymia.	1.0
Most calls with high loudness in hypomania have high spectrum compared to the state of euthymia.	1.0
Most calls with high loudness in hypomania have high quality compared to the state of euthymia.	1.0
Most calls with low pitch in hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls with low pitch in hypomania have low loudness compared to the state of euthymia.	1.0
Most calls with low pitch in hypomania have low quality compared to the state of euthymia.	1.0
Most calls with low spectrum in hypomania have low loudness compared to the state of euthymia.	1.0
Most calls with low spectrum in hypomania have low quality compared to the state of euthymia.	1.0
Most calls with high spectrum in hypomania have high loudness compared to the state of euthymia.	1.0
Most calls with high spectrum in hypomania have high quality compared to the state of euthymia.	1.0
Most calls with low quality in hypomania have low loudness compared to the state of euthymia.	1.0
Most calls with low quality in hypomania have low spectrum compared to the state of euthymia.	1.0
Most calls with high quality in hypomania have high loudness compared to the state of euthymia.	1.0
Most calls with high quality in hypomania have high spectrum compared to the state of euthymia.	1.0
Relative LS based on extended protoform - MANIA	*T*
Most calls with low loudness in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with low loudness in mania have low pitch compared to the state of euthymia.	0.6
Most calls with low loudness in mania have low quality compared to the state of euthymia.	1.0
Most calls with high loudness in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with low pitch in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with low pitch in mania have low loudness compared to the state of euthymia.	1.0
Most calls with low pitch in mania have low quality compared to the state of euthymia.	1.0
Most calls with low spectrum in mania have low pitch compared to the state of euthymia.	0.7
Most calls with low spectrum in mania have low loudness compared to the state of euthymia.	1.0
Most calls with low spectrum in mania have low quality compared to the state of euthymia.	1.0
Most calls with medium spectrum in mania have high loudness compared to the state of euthymia.	0.8
Most calls with medium spectrum in mania have low quality compared to the state of euthymia.	0.6
Most calls with high spectrum in mania have high loudness compared to the state of euthymia.	1.0
Most calls with low quality in mania have low loudness compared to the state of euthymia.	1.0
Most calls with low quality in mania have low spectrum compared to the state of euthymia.	1.0
Most calls with high quality in mania have high loudness compared to the state of euthymia.	1.0
Most calls with high quality in mania have high spectrum compared to the state of euthymia.	1.0

One of the main limitations of the proposed approach is the need for prior knowledge about the acoustic parameters to group them adequately. This is a challenging task also due to the fact that acoustic data coming from sensors are subject to various transformations. Interdisciplinary work is needed to transform them into human-centred characteristics to be used by clinicians in their practice.

It should be noted, that the obtained results illustrate the performance of the proposed approach and its efficiency in delivering explainable information granules about large and evolving data streams depending on the disease state. However, caution is necessary when interpreting results due to the fact that the degree of truth of the linguistic summaries discovered in the current work was calculated for relatively small sample size, and thus, future work will be devoted to running further experiments on data collected from other BD patients.

7. Conclusions and Future Work

In this work, acoustic data streams, representing frames of phone calls from a patient affected by Bipolar Disorder have been analyzed. We have proposed an approach to dynamically classify data by adaptive fuzzy clustering and elaborate the resulting information about clusters for explainability purposes through fuzzy logic and linguistic summarization. To cope with the absence of labeled data characterizing the considered healthcare context (where only few data acquired during the control visits can be labeled) we have proposed the use of a semi-supervised clustering algorithm that is able to learn cluster prototypes from subsequent chunks of partially labeled data. The clustering algorithm was able to capture abrupt changes occurring in data when the patient' states changed. Moreover, a comparison with a benchmark algorithm on different stream datasets has shown the effectiveness of the DISSFCM to accurately classify data.

Linguistic summaries have been derived from prototypes representing patterns discovered in data. Furthermore, in order to make summaries easier to understand, acoustic features have been granulated in four semantic classes namely, loudness,

pitch, spectrum, and quality of voice, according to the physicians' knowledge. Relative linguistic summaries based on the short and extended protoforms were retrieved about the acoustic data in various affective states. Numerical experiments confirmed that relative linguistic summaries can be considered as promising information granules and demonstrate the potential of using them in the context of smartphone-based monitoring of bipolar disorder. One of the main strengths of the proposed approach is its ability to deliver relatively easy to interpret and informative linguistic descriptions bridging the gap between hardly interpretable acoustic data streams and sparse labels coming from the psychiatric assessments about patient's mental state.

In this work, we focused on phone calls of a single patient. Future work will be addressed to compare linguistic summaries coming from different patients. Furthermore, the idea of this paper is to collect all cluster prototypes to establish membership functions for linguistic terms. A more advanced approach would be to summarize and reason about the centers of clusters themselves and observe how they change in sequential chunks. This would lead to the definition of evolving linguistic terms that could be well represented by more sophisticated forms of fuzzy sets, such as shadowed fuzzy sets. Another aspect that could be investigated in future work for this particular application context and the broader field of Speech Analysis, would be alleviating the sparseness of labeled data and trying to make efficient use of the big amounts of available speech in online repositories. Finally, another direction for future work is the extension of the considered acoustic parameters to other high-level concepts including other affective states and emotions.

Code availability

Generation of fuzzy linguistic descriptions have been implemented in Python with the support of Scikit-fuzzy module. The program code is open and available for download from the Github repository: https://github.com/kasiakaczmarek/LS-FC.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

BIPOLAR data streams were collected in the CHAD project - entitled "Smartphone-based diagnostics of phase changes in the course of bipolar disorder" (RPMA.01.02.00-14-5706/16-00) that was financed from EU funds (Regional Operational Program for Mazovia) in 2017-2018. The authors thank the researchers Karol Opara and Weronika Radziszewska from Systems Research Institute, Polish Academy of Sciences for their support in data preparation and analysis. M. Dominiak acknowledges funding from the Department of Pharmacology and Physiology of the Nervous System, Institute of Psychiatry and Neurology, Warsaw, Poland. The authors thank researchers Choiru Za'in and Mahardhika Pratama for their advice and support in performing the numerical evaluation of their WeScatterNet algorithm [35]. Choiru Za'in acknowledges the use of Pawsey computing resources to run the WeScatternet for the experiment. G. Casalino acknowledges funding from the Italian Ministry of Education, University, and Research through the European PON project AIM (Attraction and International Mobility), nr. 1852414, activity 2, line 1. This work was partially supported by INdAM GNCS within the research project "Computational Intelligence methods for Digital Health". G. Casalino and G. Castellano are with the CITEL - Centro Interdipartimentale di Telemedicina, University of Bari Aldo Moro.

References

[1] S. Allen, Artificial intelligence and the future of psychiatry, IEEE Pulse 11 (3) (2020) 2–6, https://doi.org/10.1109/MPULS.2020.2993657.
[2] M. Faurholt-Jepsen, J. Busk, M. Frost, J.E. Bardram, M. Vinberg, L.V. Kessing, Objective smartphone data as a potential diagnostic marker of bipolar disorder, Australian & New Zealand Journal of Psychiatry 53 (2) (2019) 119–128, pMID: 30387368, doi: 10.1177/0004867418808900..
[3] E. Vieta, M. Berk, T.G. Schulze, A.F. Carvalho, T. Suppes, J.R. Calabrese, K. Gao, K.W. Miskowiak, I. Grande, Bipolar disorders, Nature Reviews Disease Primers 4 (1) (2018) 1–16, https://doi.org/10.1038/nrdp.2018.8.
[4] A. Adadi, M. Berrada, Explainable AI for healthcare: From black box to interpretable models, in: in: Embedded Systems and Artificial Intelligence, Springer, 2020, pp. 327–337.
[5] M.T. Ribeiro, S. Singh, C. Guestrin, "Why should i trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, Association for Computing Machinery, 2016, pp. 1135–1144.
[6] L.A. Zadeh, Fuzzy sets, Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh, World Scientific, 1996, pp. 394–432, doi:10.1142/2895.
[7] J.M. Alonso, C. Castiello, L. Magdalena, C. Mencar, Explainable fuzzy systems: Paving the way from interpretable fuzzy systems to explainable AI systems, Studies in Computational Intelligence, Springer Nature, Cham, Switzerland, 2021, https://doi.org/10.1007/978-3-030-71098-9.
[8] R. Seising, From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis, Artificial Intelligence in Medicine 38 (3) (2006) 237–256, https://doi.org/10.1016/j.artmed.2006.06.004.
[9] K. Kaczmarek-Majer, O. Hryniewicz, K.R. Opara, W. Radziszewska, A. Olwert, J.W. Owsinski, S. Zadrozny, Control charts designed using model averaging approach for phase change detection in bipolar disorder, in: S. Destercke (Ed.), Uncertainty Modelling in Data Science, of Advances in Intell. Systems and Computing, 832, Springer International, 2019, pp. 115–123.
[10] O. Kamińska, K. Kaczmarek-Majer, K. Opara, W. Jakuczun, M. Dominiak, A. Antosik-Wójcińska, Ł. Świścicki, O. Hryniewicz, Self-organizing maps using acoustic features for prediction of state change in bipolar disorder, Artificial Intelligence in Medicine, Knowledge Representation and Transparent and Explainable Systems, doi:10.1007/978-3-030-37446-4_12.

[11] G. Casalino, G. Castellano, F. Galetta, K. Kaczmarek-Majer, Dynamic incremental semi-supervised fuzzy clustering for bipolar disorder episode prediction, in: A. Appice, et al. (Eds.), Discovery Science. DS 2020, 2020, doi: 10.1007/978-3-030-61527-7_6.
[12] G. Casalino, G. Castellano, C. Mencar, Data stream classification by dynamic incremental semi-supervised fuzzy clustering, International Journal on Artificial Intelligence Tools 28 (08) (2019), https://doi.org/10.1142/S0218213019600091.
[13] G. Casalino, G. Castellano, K. Kaczmarek-Majer, O. Hryniewicz, Intelligent analysis of data streams about phonecalls for bipolar disorder monitoring, in: Proc. of 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2021), 2021, doi: 10.1109/FUZZ45933.2021.9494512.
[14] K. Kaczmarek-Majer, O. Hryniewicz, M. Dominiak, Personalized linguistic summaries in smartphone-based monitoring of bipolar disorder patients, in: 11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), IEEE, 2019, https://doi.org/10.1109/FUZZ45933.2021.9494512.
[15] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, T.F. Quatieri, A review of depression and suicide risk assessment using speech analysis, Speech Communication 71 (2015) 10–49, https://doi.org/10.1016/j.specom.2015.03.004.
[16] J.C. Mundt, A.P. Vogel, D.E. Feltner, W.R. Lenderking, Vocal acoustic biomarkers of depression severity and treatment response, Biological psychiatry 72 (7) (2012) 580–587, https://doi.org/10.1016/j.biopsych.2012.03.015.
[17] A. Guidi, J. Schoentgen, G. Bertschy, C. Gentili, E.P. Scilingo, N. Vanello, Features of vocal frequency contour and speech rhythm in bipolar disorder, Biomedical Signal Processing and Control 37 (2017) 23–31, https://doi.org/10.1016/j.bspc.2017.01.017.
[18] A.Z. Antosik-Wójcińska, M. Dominiak, M. Chojnacka, K. Kaczmarek-Majer, K.R. Opara, W. Radziszewska, A. Olwert, Ł. Swiecicki, Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling, Int J Med Inform 138 (2020) 104131, https://doi.org/10.1016/j.ijmedinf.2020.104131.
[19] S. Graham, C. Depp, E.E. Lee, C. Nebeker, X. Tu, H.-C. Kim, D.V. Jeste, Artificial intelligence for mental health and mental illnesses: an overview, Current psychiatry reports 21 (11) (2019) 1–18, https://doi.org/10.1007/s11920-019-1094-0.
[20] J. Moreno-Garcia, J. Abián-Vicén, L. Jimenez-Linares, L. Rodriguez-Benitez, Description of multivariate time series by means of trends characterization in the fuzzy domain, Fuzzy Sets and Systems 285 (2016) 118–139, https://doi.org/10.1016/j.fss.2015.05.011.
[21] R.R. Yager, A new approach to the summarization of data, Information Sciences 28 (1) (1982) 69–86, https://doi.org/10.1016/0020-0255(82)90033-0.
[22] J. Kacprzyk, R.R. Yager, J.M. Merigo, Towards human-centric aggregation via ordered weighted aggregation operators and linguistic data summaries: A new perspective on zadeh's inspirations, IEEE Computational Intelligence Magazine 14 (1) (2019) 16–30, https://doi.org/10.1109/MCI.2018.2881641.
[23] A. Ramos-Soto, P. Martin-Rodilla, Enriching linguistic descriptions of data: A framework for composite protoforms, Fuzzy Sets and Systems 407 (2019) 1–26, https://doi.org/10.1016/j.fss.2019.11.013.
[24] R. Castillo-Ortega, N. Mann, D. Sánchez, Linguistic local change comparison of time series, in: 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), IEEE, 2011, pp. 2909–2915.
[25] M.J. Lesot, G. Moyse, B. Bouchon-Meunier, Interpretability of fuzzy linguistic summaries, Fuzzy Sets and Systems 292 (2016) 307–317, https://doi.org/10.1016/j.fss.2014.10.019.
[26] Q. Pang, H. Wang, Z. Xu, Probabilistic linguistic term sets in multi-attribute group decision making, Information Sciences 369 (2016) 128–143, https://doi.org/10.1016/j.ins.2016.06.021.
[27] S. Ryan, R. Corizzo, I. Kiringa, N. Japkowicz, Deep learning versus conventional learning in data streams with concept drifts, in: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), IEEE, 2019, pp. 1306–1313.
[28] M. Das, M. Pratama, T. Tjahjowidodo, A self-evolving mutually-operative recurrent network-based model for online tool condition monitoring in delay scenario, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 2775–2783, https://doi.org/10.1145/3394486.3403328.
[29] J. Read, F. Perez-Cruz, A. Bifet, Deep learning in partially-labeled data streams, in: Proceedings of the 30th Annual ACM Symposium on Applied Computing, Association for Computing Machinery, 2015, pp. 954–959, https://doi.org/10.1145/2695664.2695871.
[30] N. Tajbakhsh, Y. Hu, J. Cao, X. Yan, Y. Xiao, Y. Lu, J. Liang, D. Terzopoulos, X. Ding, Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, pp. 1251–1255, https://doi.org/10.1109/ISBI.2019.8759553.
[31] Y.-R. Van Eycke, A. Foucart, C. Decaestecker, Strategies to reduce the expert supervision required for deep learning-based segmentation of histopathological images, Frontiers in Medicine 6 (2019) 222, https://doi.org/10.3389/fmed.2019.00222.
[32] E. Lughofer, Improving the robustness of recursive consequent parameters learning in evolving neuro-fuzzy systems, Information Sciences 545 (2021) 555–574, https://doi.org/10.1016/j.ins.2020.09.026.
[33] E. Lughofer, M. Pratama, I. Škrjanc, Online bagging of evolving fuzzy systems, Information Sciences 570 (2021) 16–33, https://doi.org/10.1016/j.ins.2021.04.041.
[34] X. Gu, P. Angelov, Z. Zhao, Self-organizing fuzzy inference ensemble system for big streaming data classification, Knowledge-Based Systems 218 (2021), https://doi.org/10.1016/j.knosys.2021.106870 106870.
[35] M. Pratama, C. Za'in, E. Lughofer, E. Pardede, D.A. Rahayu, Scalable teacher forcing network for semi-supervised large scale data streams, Information Sciences 576 (2021) 407–431, https://doi.org/10.1016/j.ins.2021.06.075.
[36] A. Abdullatif, F. Masulli, S. Rovetta, Clustering of nonstationary data streams: A survey of fuzzy partitional methods, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (4) (2018), https://doi.org/10.1002/widm.1258.
[37] L.A.Q. Cordovil, P.H.S. Coutinho, I.V. de Bessa, M.F.S.V. D'Angelo, R.M. Palhares, Uncertain data modeling based on evolving ellipsoidal fuzzy information granules, IEEE Transactions on Fuzzy Systems 28 (10) (2019) 2427–2436, https://doi.org/10.1109/TFUZZ.2019.2937052.
[38] D. Upadhyay, S. Jain, A. Jain, A Fuzzy Clustering Algorithm for High Dimensional Streaming Data, Journal of Information Engineering and Applications 3 (10) (2013) 1–10.
[39] M.J. Patwary, X.-Z. Wang, Sensitivity analysis on initial classifier accuracy in fuzziness based semi-supervised learning, Information Sciences 490 (2019) 93–112, https://doi.org/10.1016/j.ins.2019.03.036.
[40] M.J. Patwary, X.-Z. Wang, D. Yan, Impact of fuzziness measures on the performance of semi-supervised learning, International Journal of Fuzzy Systems 21 (5) (2019) 1430–1442, https://doi.org/10.1007/s40815-019-00666-2.
[41] D. Leite, P. Costa, F. Gomide, Evolving granular neural network for semi-supervised data stream classification, in: The 2010 International joint Conference on Neural Networks (IJCNN), IEEE, 2010, pp. 1–8.
[42] F. Eyben, F. Weninger, F. Gross, B. Schuller, Recent developments in opensmile, the munich open-source multimedia feature extractor, in: Proc. of the 21st ACM Int. Conf, on Multimedia, 2013, pp. 835–838, https://doi.org/10.1145/2502081.2502224.
[43] F. Eyben, K.R. Scherer, B.W. Schuller, J. Sundberg, E. André, C. Busso, L.Y. Devillers, J. Epps, P. Laukka, S.S. Narayanan, K.P. Truong, et al, The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing, IEEE Transactions on Affective Computing 7 (2) (2016) 190–202, https://doi.org/10.1109/TAFFC.2015.2457417.
[44] L. Dm, B. Kh, G.S. Kessing, Automated assessment of psychiatric disorders using speech: A systematic review, Laryngoscope Investig Otolaryngol 31;5 (1) (2020) 96–116, https://doi.org/10.1002/lio2.354.
[45] J. Zhang, Z. Pan, C. Gui, T. Xue, Y. Lin, J. Zhu, D. Cui, Analysis on speech signal features of manic patients, Journal of psychiatric research 98 (2018) 59–63, https://doi.org/10.1016/j.jpsychires.2017.12.012.
[46] Z.N. Karam, E.M. Provost, S. Singh, J. Montgomery, C. Archer, G. Harrington, M.G. Mcinnis, Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech, in: 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2014, pp. 4858–4862.
[47] M. Faurholt-Jepsen, J. Busk, M. Frost, M. Vinberg, E.M. Christensen, O. Winther, J.E. Bardram, L.V. Kessing, Voice analysis as an objective state marker in bipolar disorder, Translational psychiatry 6 (7) (2016) e856, https://doi.org/10.1038/tp.2016.123.

[48] G. Kiss, K. Vicsi, Mono-and multi-lingual depression prediction based on speech processing, International Journal of Speech Technology 20 (4) (2017) 919–935, https://doi.org/10.1007/s10772-017-9455-8.
[49] C.R. Marmar, A.D. Brown, M. Qian, E. Laska, C. Siegel, M. Li, D. Abu-Amara, A. Tsiartas, C. Richey, J. Smith, et al, Speech-based markers for posttraumatic stress disorder in us veterans, Depression and Anxiety 36 (7) (2019) 607–616, https://doi.org/10.1002/da.22890.
[50] O. Kaminska, K. Kaczmarek-Majer, O. Hryniewicz, Acoustic feature selection with fuzzy clustering, self organizing maps and psychiatric assessments, Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2020, doi:10.1007/978-3-030-50146-4_26..
[51] W. Pedrycz, J. Waletzky, Fuzzy clustering with partial supervision, IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics 27 (5) (1997) 787–795, https://doi.org/10.1109/3477.623232.
[52] W. Pedrycz, K. Hirota, Fuzzy vector quantization with the particle swarm optimization: A study in fuzzy granulation-degranulation information processing, Signal Processing 87 (9) (2007) 2061–2074, https://doi.org/10.1016/j.sigpro.2007.02.001.
[53] W. Pedrycz, Conditional Fuzzy C-Means, Pattern Recognition Letters 17 (6) (1996) 625–631, https://doi.org/10.1016/0167-8655(96)00027-X.
[54] K. Kaczmarek-Majer, O. Hryniewicz, Application of linguistic summarization methods in time series forecasting, Information Sciences 478 (2019) 580–594, https://doi.org/10.1016/j.ins.2018.11.036.
[55] F.E. Boran, D. Akay, R.R. Yager, An overview of methods for linguistic summarization with fuzzy sets, Expert Systems with Applications 61 (2016) 356–377, https://doi.org/10.1016/j.eswa.2016.05.044.
[56] A. Grünerbl, A. Muaremi, V. Osmani, Smartphone-based recognition of states and state changes in bipolar disorder patients, IEEE Journal of Biomedical and Health Informatics 19 (1) (2015), https://doi.org/10.1109/JBHI.2014.2343154.
[57] M. Pratama, S.G. Anavatti, P.P. Angelov, E. Lughofer, Panfis: A novel incremental learning machine, IEEE Transactions on Neural Networks and Learning Systems 25 (1) (2014) 55–68, https://doi.org/10.1109/TNNLS.2013.2271933.
[58] A. Muaremi, F. Gravenhorst, A. Grünerbl, B. Arnrich, G. Tröster, Assessing bipolar episodes using speech cues derived from phone calls, in: International Symposium on pervasive computing paradigms for mental health, Springer, 2014, pp. 103–114, https://doi.org/10.1007/978-3-319-11564-1_11.