Fuzzy Sets and Systems 500 (2025) 109205



Contents lists available at ScienceDirect

Fuzzy Sets and Systems

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



Incremental learning and granular computing from evolving data streams: An application to speech-based bipolar disorder diagnosis

Daniel Leite [a,*], Gabriella Casalino [b], Katarzyna Kaczmarek-Majer [c,d], Giovanna Castellano [b]

[a] Department of Computer Science, Paderborn University, Paderborn, Germany
[b] Department of Computer Science, University of Bari Aldo Moro, Bari, Italy
[c] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
[d] Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, Ostrava, Czech Republic

ARTICLE INFO

Keywords:
Signal processing
Granular computing
Online machine learning
Explainable artificial intelligence
AI in medicine

ABSTRACT

We apply an evolving granular-computing modeling approach, called evolving Optimal Granular System (eOGS), to bipolar mood disorder (BD) diagnosis based on speech data streams. The eOGS online learning algorithm reveals information granules in the flow and design the structure and parameters of a granular rule-based model with a certain degree of interpretability based on acoustic attributes obtained from phone calls made over 7 months to the Psychiatry department of a hospital. A multi-objective programming problem that trades-off information specificity, model compactness, and numerical and granular error indices is presented. Spectral and prosodic attributes are ranked and selected based on a hybrid Pearson-Spearman correlation coefficient. Low attribute-class correlation, ranging from 0.03 to 0.07, is observed, as well as high class overlap, which is typical in the psychiatric field. eOGS models for BD recognition overcome alternative computational-intelligence models, namely, Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) and Fuzzy-set-Based evolving Modeling (FBeM-Gauss), by a small margin in both best and average cases; followed by eXtended Takagi-Sugeno (xTS) and evolving Takagi Sugeno (eTS) types of models. The proposed eOGS model using only 8 of the original acoustic attributes, and about 15 'If-Then' inference rules, has exhibited the best root mean square error, 0.1361, and 91.8% accuracy in sharp BD class estimates. Granules associated to linguistic labels and a granular input-output map offer human understandability with relation to the inherent process of generating class estimates. Linguistically readable eOGS rules may assist physicians in explaining symptoms and making a diagnosis.

1. Introduction

Bipolar Disorder (BD) is a mental illness characterized by unusual changes of mood, energy, concentration, activity level, and ability of an individual to carry out a daily routine. Moods range from 'highs' (maniac or hypomaniac episodes) to 'lows' (depressive episodes). During an episode, symptoms may last several days or weeks. Although continuous monitoring of patients over time is

* Corresponding author.
E-mail addresses: daniel.leite@uni-paderborn.de (D. Leite), gabriella.casalino@uniba.it (G. Casalino), k.kaczmarek@ibspan.waw.pl (K. Kaczmarek-Majer), giovanna.castellano@uniba.it (G. Castellano).

https://doi.org/10.1016/j.fss.2024.109205
Received 21 May 2023; Received in revised form 18 September 2024; Accepted 15 November 2024

Available online 19 November 2024
0165-0114/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

highly recommended, medical visits for control are uncommon. Monitoring, objectively diagnosing, and providing evidence-based intervention for individuals with mental illnesses is a major concern [1]. Access to traditional psychiatric services due to financial, geographical, or practical issues remains a key barrier. Remote technologies have become highly relevant.

Computational models and dedicated mobile applications able to recognize and monitor BD mood changes during phone calls have been developed [2] [3]. Acoustic attributes extracted from speech have proved effective to some degree in describing patients status [1]. Speech processing by means of machine learning algorithms has been investigated for monitoring not only bipolar disorder [4] [5], but also other medical conditions such as Alzheimer's dementia [6], and Parkinson's disease [7]. Machine learning models trained on speech recordings could improve the outcomes, save clinicians time, reduce health care costs, and make treatment planning more efficient and customized. Systematic remote medical assessment tools may facilitate a better personalization of care and improve clinical services across the medical practice.

Toward the development of computational models based on voice and speech data for BD monitoring, the two most fundamental aspects are prediction accuracy and model interpretability or explainability [8] [9]. Practically speaking, interpretability and explainability are tied or very close concepts, often used interchangeably [10]. In-model interpretability – also called transparency – comes from models that are intrinsically interpretable, e.g., some concise fuzzy rule-based models supported by information granules of balanced size [11] [12]; some modular neural networks whose synaptic weights refer to the relevance of the problem attributes and network modules, or whose parameters have some straightforward real-world implication or correspondence [9] [13]. Interpretable models imply that the inherent mathematical operations and basic data-processing entities can be understood by humans.

Explainability is a common term in the deep learning literature [8] [14] to refer to post hoc methods, i.e., methods applied after model training to explain some characteristics and behaviors of *black-box* deep neural models. Nonetheless, post hoc explainable methods can be applied to intrinsically interpretable models since they are usually decoupled from the main model design. In this case, post hoc methods can possibly add further interpretability. Lipton [14] argues that post hoc interpretability answers the question *what else can the model tell us*, whereas intrinsic interpretability answers the question *how the model works*. Herein, the terms interpretability and explainability will be used free from a mathematical definition [11] [15] [16], in the broad general sense of understandability in human terms.

Granular Computing (GrC) frameworks have been proposed to uncover meaningful knowledge from large data streams in form of information granules that denote data items connected by similarity, proximity or functionality in a domain [13] [17] [18]. GrC methods examine the information flow in dynamic environments to develop and update a granular model that can be linguistically understood once expressed by 'If-Then' rules [11] [19]. Granular rule-based models have an inherent level of interpretability. In-model interpretability is achieved by means of: (i) generalized constraints [20], which delimit the basic information-processing units [21]; (ii) hyperparameters that come from domain knowledge and expert interactions with the environment – default values can be assumed; and (iii) linguistic values attributed to granules depending on their spatial location and organization. A prediction is given together with a linguistic statement that clarifies why the prediction has been offered. Several reasons are outlined in [22] to emphasize why explainable black boxes should be avoided in high stake decisions, such as medical decisions. Explain black box models to alleviate immediate problems is argued to perpetuate bad practice. Conversely, GrC provides models with the potential to be interpretable within a domain.

While typical machine learning methods train models on a subset of data and test performance on the rest of the data not used for training to give insight into how well they may generalize, online evolving GrC methods provide estimates based on the input-output granular map they know at the moment. When the true output value is available, a GrC method performs a single and quick training pass over active granules and associated parameters using the input-output pair. The incremental granulation and parametric updating process aims at retaining the essence of the streaming data as granular objects. Direct application of machine learning methods to data streams is usually infeasible since it is difficult to maintain all the data in memory. Computing with adaptive granules and If-Then rules in online environment tends to provide more *human-centered* models.

The field of evolving intelligence in general is rapidly developing with numerous applications in online data-driven modeling [23] [24]. With the ability to adapt model parameters and structure in real-time, the burden and costs related to redesigning models periodically are avoided. Many incremental learning mechanisms, both deterministic and heuristic, have been proposed to aid in the development of evolving neural networks and rule-based systems from numerical and uncertain data streams [13] [25] [26] [27] [28]. While significant progress has been made towards achieving practical solutions for immediate goals, there is still a need to ensure that certain conditions, stability, and performance will be optimally accomplished in dynamic environments. This is critical in medical diagnosis, and in areas such as unmanned vehicles, privacy and security, logistics, and closed-loop systems, where reliability and safety are paramount. Despite these challenges, the potential benefits of evolving artificial intelligence are vast, and ongoing research efforts have delivered sophisticated and effective solutions.

This paper proposes and elaborates on an evolving GrC approach as a potentially interpretable method to address the BD diagnosis problem using speech data. The eOGS (evolving Optimal Granular System) – first introduced in [21] as a general-purpose method to autonomously create and update predictors, controllers, and classifiers – herein produces a fuzzy-interval predictive model which uses spectral and prosodic features to remotely detect the presence of BD in an individual. The method uses piecewise affine and inclusion functions associated with Gaussian and hyper-rectangular granules to provide estimates of a time-varying BD function. eOGS self-adapts its structure whenever a new concept emerges or a drift occurs in the speech data stream while minimizing a multi-objective function based on α-level sets and Pareto fronts. The method trades-off accurate estimates and the amount and specificity of granules needed to represent BD patterns.

The contributions of this paper are:

- a nonlinear and time-varying granular rule-based model that maps speech-extracted data into classes of bipolar disorder. Linguistic labels are assigned to granules for potentially interpretable supporting statements;
- a customized online learning method, called eOGS, that keeps track of spatio-temporal patterns in never-ending data streams. Different from other online algorithms, eOGS is driven by a multi-criteria optimization function where accuracy, information specificity, and model compactness are important;
- a suggestion of combined Pearson-Spearman correlation coefficient to rank and select acoustic attributes generated by individuals with bipolar disorder. The effect of attribute selection is studied;
- comparative results considering 5 evolving fuzzy and neuro-fuzzy methods, namely, eOGS [21], evolving Takagi Sugeno (eTS) [29], eXtended Takagi-Sugeno (xTS) [30], Fuzzy-set-Based evolving Modeling (FBeM-Gauss) [31], and Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) [32].

2. Literature review on speech-based intelligent models for bipolar disorder classification

This section provides an overview of some research papers that report the application of machine learning methods to speech-based BD diagnosis. We do not intend to present an exhaustive literature review, but to summarize a few recent studies closely related to the topics of the present paper.

2.1. General machine learning methods

In [33], machine learning methods are used to distinguish clinical conditions (BD, depressive disorder, schizophrenia, and generalized anxiety) in psychiatric emergencies. Time and frequency acoustic attributes play the role of a digital marker to spotlight changes in vocal patterns. A Random Forest method composed by 300 trees outperformed methods such as Support Vector Machine (SVM), Naive Bayes, and Multi-Layer Perceptron and Radial-Basis Function neural networks in classification performance. Authors emphasized the speech model-based approach as being inexpensive, remote, and non-invasive. Different from [33], the eOGS approach described in the present paper – instead of being based on hundreds of trees – is driven by a multi-objective function where compactness (few parameters and rules) and the specificity of information granules are continually taken into account. The evolution of the classifier is focused not only on accuracy but also on providing a degree of interpretability to support decision making.

Similar to [33], Random Forest achieved the most accurate estimates in [34] compared to estimates provided by Decision Trees, SVM, and Logistic Regression setups for independent BD patients. Data from a smart watch, daily self-reports, and medical observations at regular appointments were evaluated. The classes refer to different states of patients who had already been diagnosed with depressive BD. Interestingly, none of the obtained models and parameters were the best for all independent patients – which suggests that unique customized solutions for disorder monitoring, in opposition to a general and unanimous machine-learning solution that works for every person, may be attractive as a research path. Besides, class recognition using a subset of selected attributes produced better results. The most relevant subset of variables was different for different independent subjects. This once more corroborates with customized healthcare solutions. Such outcomes are consistent with the results we present in this paper. Different evolving models generate better BD predictions observing different amounts of speech attributes. The continuous eOGS learning process constitutes a framework for the development of customized models.

Responses in interviews with 15 healthy controls and 30 individuals with BD and unipolar depression, after they watched 6 emotion-eliciting videos, are evaluated in [35]. Hierarchical Spectral clustering over an emotion dataset is used to map the instances to a mood dataset for alleviating bias inherent to the emotion space. A Convolutional neural network with attention mechanism generates an emotion profile for the elicited speech. Then, a Long Short-Term Memory characterizes the temporal evolution of emotion profiles. The proposed approach reached a 75.6% classification accuracy – outperforming SVM, 62.2%, and a standalone Convolutional net, 66.7%. Different from [35], the acoustic attributes used in the experiments of the present paper were obtained during natural phone calls to the Psychiatry department of a hospital – visits for interviews or consultations needless. The eOGS information granulation method works on the space of known selected acoustic attributes – instead of working on a more abstract space with nonlinearly-transformed data representations (produced after the application of a chain of convolutional filters and activation functions), where the information has no correspondence with reality.

An ensemble of Convolutional neural nets as weak learners followed by a Multi-Layer Perceptron meta-model for BD classification (stacking scheme) is addressed in [5]. The ensemble is fed with speech data produced in interviews with structured questions. Spectral and cepstral attributes such as Mel-frequency cepstral coefficients and Geneva minimalistic acoustic parameters are extracted from the audio signals. Mania, hypomania, and remission are the clinical states annotated by psychiatrists. The stacked ensemble outperformed other methods on the BD Turkish corpus [5]. In [36], a multimodal (acoustic-linguistic-visual) decision system for 3-level mania classification is also applied to the Turkish BD corpus. The offline-designed multimodal system achieved a 64.8% recall score – the state-of-the-art performance on the underlying dataset. Different from [5] and [36], this paper uses some spectral (centroid, harmonics) and prosodic (jitter, shimmer, loudness, and log of energy) unimodal (speech) attributes. The latter refers to long-time phoneme-level variations in intonation, stress, and speech rhythm. Moreover, while convolutional and fully-connected perceptron-based neural nets are known to be offline designed black boxes; eOGS is a transparent online method, since it learns from a data stream and offers a human-centered granular solution. eOGS processes original attributes (with greater connection with the real world) rather than hundreds of features from transformed spaces with low coherence with the human way of thinking.

Although deep learning is the prevalent technique across the speech processing field, the corpora available for health applications are small concerning both the amount of data and evaluated patients. In [37], a learning framework for constructing more reliable

deep nets for classifying 3 levels of mania in BD patients using small datasets is addressed. Speech clips are segmented into chunks, which are input to Feed-Forward, Recurrent, and Convolutional deep neural nets. Each individual chunk is weakly labeled – since they are annotated with the label of the corresponding speech clip, but may not be indicative of that label. An ensemble of deep models performed better than, or matched, the state-of-the-art methods on the BD corpus of the 2018 Audio-Visual Emotion Challenge (3-class mania classification) [38]. Speech segments with weak labels may mislead learning methods. Imprecise labels occasionally assigned by clinicians may affect the trustfulness and benefits of a long-term intelligent monitoring system, especially if predictions are not supported by explanations.

2.2. Remarks on model interpretability and method explainability

The priority of the artificial intelligence has long been on the predictive power of models, while the ability to explain decision processes has taken a back seat. The demand for the ability to understand and trust machine learning and artificial intelligence systems has increased recently in domains such as healthcare, criminal justice, and computer vision, influenced by deep neural nets and ensembles of black box models [8]. More interpretable models and explanation methods have been developed over the past few years [9]. The next section describes some of these developments with emphasis on BD diagnosis [39] [40] [41]. Understandability in the psychiatric field is of utmost importance to ensure continued confidence in the deployment of intelligent software.

In BD monitoring, getting a prediction is not enough. A correct prediction partially solves the original problem. The model must describe how it came to the prediction. Model interpretability is a subjective, hard to formalize, and domain-specific concept [8] [15]. Depending on the situation, patient needs, and medical expertise, different types of descriptions might be useful. For example, a physician might want to know a couple of reasons why a patient was diagnosed as 'potential state of disorder', and another one as 'healthy' based on differences in vocal loudness and speech rate. In another use case, the physician wants to know many more details about the speech features of all his/her patients to get in depth insights. The greatest satisfaction about an adequate model granularity depends on the occasion, purpose, and audience.

In this paper we use the term interpretability according to its non-mathematical definition, such as in [15] and [16], namely, *"interpretability is the degree to which a human can understand the cause of a decision"*. The interpretability of an evolving model is intuitively higher if it is easier for a person to reason and trace back why a prediction was made. In-model interpretability concerns models that have inherent interpretability potential in them through the use of meaningful basic entities and constraints [9]. Different from in-model intrinsic interpretability potential, post-model interpretability refers to improving the interpretability degree after building the model (post hoc). A granular rule-based model, such as the proposed eOGS model, belongs to the former category.

2.3. Explainable granular computing methods

In general, psychiatrists using a speech-based decision support software do not have deep knowledge of low-level acoustic descriptors, artificial intelligence, and nonlinear manifolds. They are more inclined to rely on general and abstract information granules so that links to their prior medical studies and clinical practice are established. For example, a psychiatrist may state that during the BD depressive state, the speech activity is reduced, and pause-related voice attributes are intensified. Attribute extraction is accompanied by uncertainties, e.g., patients use devices with different microphone quality. Background noise during phone calls and involuntary interruptions are also challenging issues. As discussed in Section 1, GrC [42] [43] is a general-purpose framework to learn from data, handle uncertainties, and allow expression of expert knowledge to facilitate human-machine communication.

An explainable neuro-fuzzy classification approach for BD, called PLENARY, is presented in [39]. Neural network learning based on two levels of labels assigned to the data is exploited. Subsequently, a fuzzy linguistic summary is built through the Shapley Additive Explanation tool [44]. The PLENARY model is validated on pre-processed speech signals collected from smartphones from individuals with BD. The model provides improved explainability because it aggregates low-level attributes into high-level granules, and incorporates vague domain knowledge into a multi-task post-hoc-explained neural net. Different from [39], the eOGS method given in this paper is a plug-and-play approach in which the rule-based model jointly learns to predict and explain. Linguistic rules ensure a degree of clarity of the model at any time along the never-ending learning process. Additionally, human experts can express preferences by means of optimization constraints and hyperparameters. PLENARY and eOGS are also different in terms of the multi-criteria problem that guides the online structural optimization of the latter.

An online semi-supervised fuzzy clustering method, called LS-FC [40] [41], has been applied to acoustic data recorded from BD patients using smartphones in uncontrolled environment. The goal is to distinguish the states of mania/hypomania from that of euthymia. Clusters are projected onto axes to shape membership functions linked to linguistic values such as 'low loudness' and 'high jitter'. Linguistic summaries are further developed for a human-centered prediction of the subject mental states. Psychiatrists incorporate information about cluster prototypes in accordance with clinical observations. The LS-FC method translates detailed data into information granules. Similar to LS-FC, the eOGS method deals with dynamic data streams. Its multicriteria-driven algorithm, however, balances granule specificity and rule base compactness with the typical numerical error index that guides LS-FC and the vast majority of evolving and incremental methods [23]. In other words, the attainment of a concise optimal classifier is an explicit concern within the eOGS framework. Additionally, in this paper we propose the possibility of a third class, namely 'Potential state of Bipolar Disorder'. This class refers to data occurrences to which psychiatrists are unable to designate labels in a safe and timely manner.

3. Preliminary concepts for optimal granular models from data streams

Sections 3 and 4 provide a brief overview of the basic concepts of optimal granular systems from evolving data streams [21]. The construction of an eOGS model consists in keeping an updated summary of the streaming data by using an incremental learning algorithm. A set of granules $\gamma = \{\gamma^1, ..., \gamma^c\}$ is used to represent spatial and temporal information from real-valued vectors $x^{[h]} = (x_1, ..., x_j, ..., x_n)^{[h]}$, $h = 1, ...$ Each granule γ^i can accommodate an instance $x^{[h]}$. If $x^{[h]}$ is not sufficiently related to any existing γ^i, $i = 1, ..., c$, a new granule γ^{c+1} is created, expanding the collection γ. Thus, γ is a structural adaptive granular model that describes the stream $x^{[h]}$, $h = 1, ...$

In particular, an interval granule γ^i depends basically on the lower l_j^i and upper L_j^i endpoints, $j = 1, ..., n$, of n-dimensional hyper-rectangles aligned to the axes. These endpoints are determined by spanning the α-level sets of the uni-dimensional Gaussians Π_j^i. Assume a Gaussian membership function $\Pi_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$ as the local representation of the j-th attribute of the i-th granule of a collection. We are concerned with the range of values that the instances typically fall. Parameter α trims Π_j^i at both sides. The truncation of Π_j^i generates the interval $[l_j^i, L_j^i]$, whose projection over the Cartesian product space forms a granule [21]. Interval granules pave the data space in a non-uniform manner, while fuzzy Gaussians represent the essence of the information within the granule.

Depending on the constriction of Π_j^i and the value of $\alpha \in [0, 1]$, granules and granular map can achieve any level of specificity. With the use of an online algorithm to update μ_j^i and σ_j^i, the Gaussian Π_j^i can expand, contract or drift as needed to follow nonstationary behaviors. At the same time, by adapting the parameter $\alpha \in [0, 1]$, the algorithm controls the width of $[l_j^i, L_j^i]$ (a slice of Π_j^i) to manage the specificity of interval granules. Overall, eOGS models are highly flexible and can be tailored to specific needs by managing the parameters involved. From a perspective, eOGS models are aligned with the principle of justifiable granularity [45] [46], which concerns developing meaningful granules from experimental data.

The principle consists of conflicting requirements for guiding algorithms in the search for an optimal granularity [45] [47]. In essence, the numerical evidence supporting individual granules should be as great as possible. The greater the amount of data covered by a granule, the better. Conversely, granules should be as specific [48] as possible to describe the nuances of the data and retain semantics in the resulting model. Several formulations and algorithms have been considered to address this optimization problem [47]. Besides data coverage and information specificity, other criteria to bear a rationale for the development of granules have been discussed [45]. In eOGS, *granular prediction error* and *model structural compactness* are additional criteria in a multi-objective function that guides the granulation of stream data and the adaptation of local functions.

3.1. Information specificity

The concept of specificity revolves around the idea of how much information a granule carries [48]. Measures of specificity have their maximum values when an object reduces to a single element. The values of specificity are expressed within the interval $[0, 1]$, where a value of 0 indicates that all points within a given domain are equally possible, while a value of 1 suggests singularity.

Consider a subset Π of a continuous domain X [48]. A broad class of specificity is defined by

$$\text{Sp}(\Pi) = \int_0^{\alpha_{\max}} F(\lambda(\Pi_\alpha))\text{d}\alpha, \tag{1}$$

in which $\Pi_\alpha = \{x : \Pi(x) \geq \alpha\}$ is the α-level of Π. The height of Π is represented by $\alpha_{\max}$, λ is a monotonic measure, and $F : [0, 1] \to [0, 1]$ is a function that satisfies $F(0) = 1$, $F(1) = 0$, and $0 \leq F(x_1) \leq F(x_2)$, for $x_1 > x_2$.

Consider a Gaussian function Π in X with modal value μ and dispersion σ,

$$\Pi = e^{-(x-\mu)^2/2\sigma^2}. \tag{2}$$

Π is positive everywhere, and $\int_{-\infty}^{\infty} \Pi \text{d}x$ does not necessarily equal 1 since Π is normal. We define the α-level interval of Π as Π_α. It specifies the range of values of x for which $\Pi(x)$ is greater or equal to α. Rearranging (2), $\alpha > 0$ cuts Π at $\mu \pm \sqrt{-2\sigma^2 \ln\ \alpha}$. The α-level interval of Π is $[\mu - \sqrt{-2\sigma^2 \ln \alpha},\ \mu + \sqrt{-2\sigma^2 \ln \alpha}]$.

Assuming $X = [c, d]$ as a bounded domain, we can use the definition of specificity (1) and the α-level interval Π_α to write the specificity of Π in terms of σ and α, that is,

$$\text{Sp}(\Pi) \ = \ 1 - \frac{2\sqrt{2}\sigma\sqrt{-\ln(\alpha)}}{d - c}. \tag{3}$$

The concept of specificity can be extended to higher dimensional spaces. Let $\Pi^i = \{\Pi_1^i, ..., \Pi_j^i, ..., \Pi_n^i\}$ be a set of Gaussians defined on the domain $X^n = X_1 \times ... \times X_j \times ... \times X_n$. By taking α-cuts, we can construct an interval granule γ^i in X^n. The specificity of an n-dimensional granule γ^i is defined as the average of the specificity measure of each of its n components [21],

$$\text{Spa}(\gamma^i) \ = \ 1 - \frac{2\sqrt{2}}{n} \sum_{j=1}^{n} \frac{\sigma_j^i \sqrt{-\ln(\alpha)}}{d_j - c_j}, \tag{4}$$

where c_j and d_j are the lower and upper bounds of the domain X_j.

3.2. Multi-objective decision making

Let $\mathbb{P}$ denote the set of all eOGS parameters. For example, the α level to cut Gaussians at each tail to produce intervals is a parameter. Other eOGS parameters will be discussed in the next section. There are several possible criteria that can guide the adaptation of $\mathbb{P}$ including the numerical (E_n) and granular (E_g) estimation errors, specificity of the granular map (Spa(γ)), and number of rules (c). The goal is to minimize numerical and granular errors, increase the specificity and meaningfulness of the granules, and keep the rule base concise. However, clear trade-offs should be considered. For instance, a small number of rules can result in less specific maps and may generate better or worse predictions depending on the data stream.

A multi-objective function to be minimized is

$$\begin{aligned} F(\mathbb{P}) \;&=\; \min\; [f_1(\mathbb{P}),\; f_2(\mathbb{P}),\; f_3(\mathbb{P}),\; f_4(\mathbb{P})] \\ &\text{s.t. } \mathbb{P} \in \Omega \end{aligned} \tag{5}$$

in which Ω represents the space of parameters and

$$\begin{aligned} E_n \;&\triangleq\; f_1(\mathbb{P}), \\ E_g \;&\triangleq\; f_2(\mathbb{P}), \\ -\mathrm{Spa}(\gamma) = -\sum_{i=1}^{c} \mathrm{Spa}(\gamma^i) \;&\triangleq\; f_3(\mathbb{P}), \\ c \;&\triangleq\; f_4(\mathbb{P}). \end{aligned} \tag{6}$$

The objectives of (5) are in conflict. The solution is not unique. Improving one objective can come at the expense of the others; however, Pareto optimality is preserved [21]. The next section discusses the parameters $\mathbb{P}$.

When using the ϵ-constraint method [49], we choose a priority objective f_s for optimization, and set upper bounds on the remaining objectives f_t to convert them into constraints. The resulting form of problem (5) is

$$\begin{aligned} F(\mathbb{P}) = \min\; & f_s(\mathbb{P}) \\ \text{s.t. } & f_t(\mathbb{P}) \le \epsilon_t,\; \forall t,\; t \neq s \\ & \mathbb{P} \in \Omega \end{aligned} \tag{7}$$

By establishing a priority objective and imposing upper bounds on the remaining objectives, the eOGS algorithm generates local non-inferior solutions in terms of numerical and granular errors, rule-base compactness, and granule specificity. These solutions are weakly Pareto optimal [21] [49].

4. eOGS: evolving optimal granular system

Evolving granular models provide numerical and granular predictions of nonlinear and nonstationary functions [31] [50]. An incremental algorithm is used to build the model structure and adapt parameters when new patterns are found in the data. eOGS can detect changes and deal with uncertainty, while also allowing human experts to incorporate their intuition and experience. As shown in Fig. 1, the framework supports both learning from data streams and learning from experience. The expert in Fig. 1 has a mental model of the environment, which may change when an explanation is found for an expected event. In the same way, explanations and, thus, inherent interpretability potential, are essential to support the predictions. Experts can adjust eOGS hyperparameters $\mathbb{P}$ and impose constraints (5)-(6) to guide eOGS learning to their specific needs and interests.

The fundamental working principle is to group similar data into hyper-rectangles. This simplifies computations and makes the model readable in linguistic terms. The model itself becomes the source of knowledge instead of the detailed data. Granules make the knowledge explicit in a convenient manner for humans. A granular mapping from X to Y consists of If-Then rules that assign granules in X to granules in Y. An eOGS model encodes these mappings, with the granular output given either by an inclusion function or by the bounds of output granules. The numerical output is determined by a local algebraic function p associated with each rule. In the following sections, we explore the structure of eOGS rules and the associated incremental algorithm.

4.1. Inference rules

A granule γ^i in eOGS is defined by spanning α-level sets of membership functions $\Pi_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$, where $j = 1, ..., n$, in the Cartesian product space $x_1 \times ... \times x_n$ [31]. The granule γ^i is characterized by a rule of the form:

$$\begin{aligned} R^i\text{: IF } \;& (l_1^i \le x_1 \le L_1^i) \text{ AND } \;\ldots\; \text{AND } (l_n^i \le x_n \le L_n^i) \\ & \text{THEN } \;(u_1^i \le y_1 \le U_1^i) \;\text{ AND }\; \bar{y}_1^i = p_1^i(x_1, ..., x_n) \;\text{ AND} \end{aligned}$$

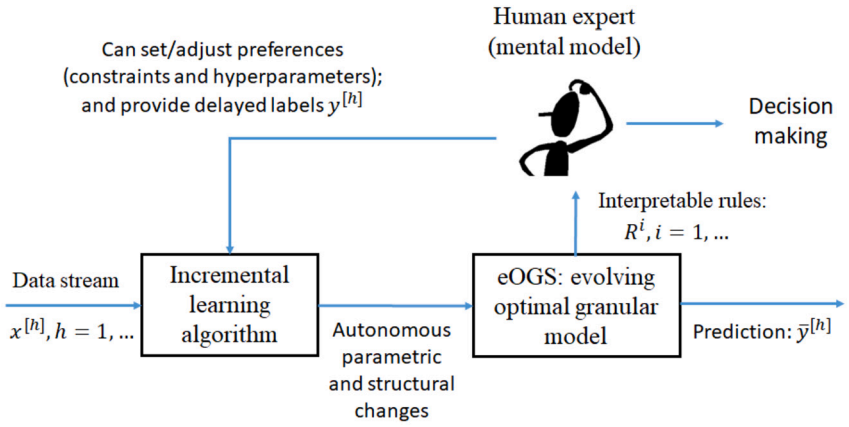


Fig. 1. eOGS general-purpose framework for designing optimal and potentially-interpretable granular systems from evolving data stream.

$$\ldots$$
$$(u_m^i \leq y_m \leq U_m^i) \text{ AND } \bar{y}_m^i = p_m^i(x_1, ..., x_n)$$

In rule R^i, l_j^i and L_j^i are the j-th lower and upper bounds of the attribute x_j according to the i-th rule, while u_k^i and U_k^i are the k-th lower and upper bounds of the i-th partition of the target output y_k. The conjunctive operator 'AND' can be realized by any triangular norm T. The minimum (Gödel) T-norm is used in this paper. Additionally, p_k^i for $k = 1, ..., m$ are numerical affine functions [51],

$$\bar{y}_k^i = p_k^i(x_1, ..., x_n) = a_{0k}^i + \sum_{j=1}^{n} a_{jk}^i x_j. \tag{8}$$

Moreover, y_k, the target output variable, becomes available in a subsequent time step for a supervised incremental learning step. The domain of the output variable y_k is granulated into $[u_k^i, U_k^i]$, $i = 1, ..., c$, similar to the domain of the input features. The model maps granules from different domains and carries information within them, such as Gaussian membership functions, local affine functions, and counters, which are used to compose rules and support the model outcomes toward a higher level of interpretability for humans.

Functions p_k^i can be of various types in general. The hyper-rectangle γ^i stores Gaussian membership functions $\Pi_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$, $j = 1, ..., n$, as its internal representation. This additional information summarizes past data belonging to γ^i. Modal values μ_j^i and dispersion σ_j^i are captured from the data stream in a recursive way. Moreover, when the i-th rule is active for an instance $x = (x_1, ..., x_n)$, the Recursive Least Squares (RLS) algorithm [21] [50] determines the coefficients a_{jk}, $j = 0, ..., n$, $k = 1, ..., m$, of the functions p_k^i, which generate numerical estimates.

4.2. The Stigler initialization approach

The process of creating granules and rules in eOGS involves checking whether $x^{[h]} \notin [l_j^i, L_j^i]$ $\forall i$ and at least one input dimension j per γ^i, or $y^{[h]} \notin [u_k^i, U_k^i]$ $\forall i$ and at least one output dimension k per γ^i. The endpoints of the intervals used to define the hyper-rectangles depend on the value of α, which cuts Π_j^i and Π_k^i. As α approaches zero, the hyper-rectangles cover the entire input and output spaces, forming unit n and m-dimensional hyper-boxes for data scaled within $[0, 1]$. Conversely, as α approaches one, the granules degenerate into points.

In order to initialize the parameters of a new granule, one possible approach is to use Stigler's standard Gaussian function [21] [52]. Specifically, the modal value of the new granule γ^{c+1} is

$$\mu_{j,k}^{c+1} = (x_j, y_k)^{[h]} \ \forall j, k, \tag{9}$$

and the dispersion is

$$(\sigma_{j,k}^{c+1})^2 = 1/2\pi \ \forall j, k. \tag{10}$$

To obtain a hyperbox centered at $(x, y)^{[h]}$, the Cartesian product of uni-dimensional α-level sets is taken. The bounds of the hyperbox are as follows:

$$[l_j^{c+1}, L_j^{c+1}] = x_j^{[h]} \pm \sqrt{-2(\sigma_j^{c+1})^2 \ln(\alpha)} \tag{11}$$

and

$$[u_k^{c+1}, U_k^{c+1}] = y_k^{[h]} \pm \sqrt{-2(\sigma_k^{c+1})^2 \ln(\alpha)}, \tag{12}$$

$0 < \alpha < 1$. Coefficients of consequent functions are initialized as follows:

$$a_{0k}^{c+1} = y_k^{[h]},\ \forall k;\ \text{and}\ a_{jk}^{c+1} = 0,\ \forall k,\ j = 1, ..., n. \tag{13}$$

The initial granule's specificity is calculated from (4).

The Stigler initialization approach (*starting big*) may require some data instances to belong and activate the respective granule in the future, in order to effectively shrink and size the new granule appropriately. New instances that fall within the bounds of the new granule are used to update modal values, dispersion, and consequent parameters.

4.3. Online parameter adaptation

Data streams change over time. Adapting eOGS rules involves steps such as enlarging or contracting the bounds of granules, updating modal values and dispersion, and simultaneously changing the coefficients of consequent functions to better fit local behaviors. These steps help ensure that eOGS remains up-to-date and effective in modeling the behavior of a system.

Let a time window be denoted as $(x, y)^{[h-\nu]}$, $(x, y)^{[h-\nu+1]}$, ..., $(x, y)^{[h]}$, where ν represents the window's length, and h represents the current time index. In the event that $x^{[h]}$ is in $[l_j^i, L_j^i]\ \forall j$, and $y^{[h]}$ is in $[u_k^i, U_k^i]\ \forall k$, for a specific i, the parameters of the Gaussians Π^i are updated recursively from

$$\mu_{j,k}^i(\text{new}) = \frac{\nu^i \mu_{j,k}^i(\text{old}) + (x_j, y_k)^{[h]}}{\nu^i + 1}, \forall j, k \tag{14}$$

and

$$\left(\sigma_{j,k}^i(\text{new})\right)^2 = \beta^i \left(\sigma_{j,k}^i(\text{old})\right)^2, \forall j, k, \tag{15}$$

in which

$$\beta^i \triangleq \frac{\nu^i(\mu_{j,k}^i - (l_j, u_k)) + \Psi(|\mu_{j,k}^i - (x_j, y_k)^{[h]}|)}{(\nu^i + 1)(\mu_{j,k}^i - (l_j, u_k))}. \tag{16}$$

Variable ν^i is a counter of instances that belong to γ^i out of the previous ν instances. The hyperparameter Ψ is set to a default value of 2. Superior or inferior values cause greater expansion or contraction of the Gaussians over time. It is worth noting from (14)-(16) that storing data either inside or outside the window is needless as only cumulative values are computed.

The α-level set of the updated granule, γ^i, is obtained from

$$[(l_j^i, u_k^i), (L_j^i, U_k^i)] = \mu_{j,k}^i(\text{new}) \pm \sqrt{-2\ln(\alpha)\left(\sigma_{j,k}^i(\text{new})\right)^2} \tag{17}$$

$\forall j, k$. The specificity of the underlying granule, $\text{Spa}(\gamma^i)$, is calculated using (4).

The ϵ-constraint method (7) is used to determine the appropriate values of α and Ψ to satisfy the constraints f_t and minimize the priority objective f_s. The values are then are used in (15) and (17) to update the dispersion of Π_j^i and Π_k^i, and to determine the bounds $[l_j^i, L_j^i]$ and $[u_k^i, U_k^i]$. The preferences of users or experts are expressed by selecting the priority objective and specifying acceptable values for the constraints ϵ_t for all t.

The procedures outlined in this section are recursive, meaning that all values are updated incrementally. After being processed, a pair $(x, y)^{[h]}$ is discarded. The RLS algorithm [50] adjusts the coefficients a_{jk}^i for instances that activate γ^i. Additionally, the length of the time window ν determines the lifespan of information within the short-term memory of an eOGS model. The model considers only the past ν instances to track the evolution of the system.

4.4. Merging granules

Merging partially overlapping granules carrying similar information, say γ^{i_1} and γ^{i_2}, can be beneficial for a compact model. To merge granules, we calculate the minimum 2-norm of the difference between their modal values, i.e.,

$$\arg \min_{i_1, i_2 = 1, ..., c;\ i_1 \neq i_2} \frac{||\mu^{i_1} - \mu^{i_2}||_2}{n}, \tag{18}$$

where n is the number of attributes, or dimensions of μ^i. If the closest granules, γ^{i_1} and γ^{i_2}, satisfy the condition

$$\frac{||\mu^{i_1} - \mu^{i_2}||_2}{n} \leq \omega, \tag{19}$$

where ω is a time-varying threshold, then the respective granules are merged. An adaptation heuristic for the hyperparameter ω is discussed in Section 5.3.

The input-output granule γ^{c+1} that results from merging γ^{i_1} and γ^{i_2} is formed by $n + m$ Gaussian membership functions, whose modal values are determined by the number of instances that each merged granule previously represented. Specifically,

$$\mu_{j,k}^{c+1} = \frac{\nu^{i_1}\mu_{j,k}^{i_1} + \nu^{i_2}\mu_{j,k}^{i_2}}{\nu^{i_1} + \nu^{i_2}},\ \forall j,k. \tag{20}$$

in which ν^{i_1} and ν^{i_2} are counters of instances represented by γ^{i_1} and γ^{i_2}.

The Stigler dispersion approach [52], which seeks to maximize the spread of the local data, takes

$$\sigma_{j,k}^{c+1} = \max(\sigma_{j,k}^{i_1}, \sigma_{j,k}^{i_2}),\ \forall j,k, \tag{21}$$

as the dispersion of the new γ^{c+1}. This approach aims to preserve information from a variety of instances in the merging region. The bounds and specificity of γ^{c+1} are obtained from (17) and (4), respectively. The consequent coefficients are initially set as

$$a_{jk}^{c+1} = \frac{a_{jk}^{i_1} + a_{jk}^{i_2}}{2},\ j = 0,...,n; k = 1,...,m, \tag{22}$$

where $a_{jk}^{i_1}$ and $a_{jk}^{i_2}$ are the consequent coefficients of γ^{i_1} and γ^{i_2}. After merging, γ^{i_1} and γ^{i_2} are deleted.

4.5. Deleting inactive rules

Rules that remain inactive for a certain number of time steps can be removed. Inactivity may indicate that the data source has changed, making it practical to eliminate unnecessary granules and keep the rule base compact. After all, having too many rules can make models more difficult to be understood. Consider ν^i, which represents the number of instances that activated granule γ^i out of the latest ν instances, $x^{[h-\nu]},...,x^{[h]}$. For inactive granules, $\nu^i = 0$, and the corresponding granule and rule can be removed.

4.6. Interval consequent function

The image of γ^i through a real function p_k^i is

$$p_k^i([l_1^i, L_1^i],...,[l_n^i, L_n^i]) = \{p_k^i(x_1,...,x_n) : x_j \in [l_j^i, L_j^i],\ j = 1,...,n\}. \tag{23}$$

The image of γ^i through p_k^i is typically not a hyper-rectangle, does not have a convex shape, and can be difficult to calculate in a closed form. To overcome this, an inclusion function P_k^i, which is a hyper-rectangle within the range of p_k^i, i.e., $[u_k^i, U_k^i]$, can be approximated.

An interval function P_k^i is referred to as an inclusion function if $p_k^i \subseteq P_k^i$ for all i. Inclusion functions are not unique, and their construction depends on the selection of P [50]. An inclusion function P_k^i is optimal if it is the convex hull of p_k^i [51]. Put simply, the optimal inclusion function for p_k^i is the smallest hyper-rectangle $P_k^{i^*}$ that encloses p_k^i. This optimal inclusion function $P_k^{i^*}$ is unique and presents the highest level of specificity that ensures inclusion.

Suppose p_k^i is monotonically increasing in $[l_j^i, L_j^i]$, $j = 1,...,n$. In that case, we can calculate p_k^i from

$$p_k^i(x) = [p_k^i(l_1^i,...,l_n^i),\ p_k^i(L_1^i,...,L_n^i)]. \tag{24}$$

Thus, for any $x \in \gamma^i$, it holds that $p_k^i(x) \subseteq [p_k(l^i), p_k(L^i)]$. When p_k^i is a monotonic decreasing function, we have

$$p_k^i(x) = [p_k^i(L_1^i,...,L_n^i),\ p_k^i(l_1^i,...,l_n^i)]. \tag{25}$$

Consider

$$p_k^i(x \in \gamma^i) = a_{0k}^i + \sum_{j=1}^{n} a_{jk}^i [l_j^i, L_j^i], \tag{26}$$

in which $(a_{0k}^i,...,a_{nk}^i)$ are degenerated intervals, i.e., real-valued parameters. If the slope $a_{jk}^i < 0$, then, the bounds in (26) should be inverted, which becomes $[L_j^i, l_j^i]$.

The interval vector generated by (26) is restricted by intervals $[u_k^i, U_k^i]$, which are derived after granulating output data $y^{[h]}$ and taking α-level sets. Namely, intervals $[u_k^i, U_k^i]$ place limits on the values produced by (26). This is realized by the AND connector in the rule consequent which enforces interval intersection – as explained in Section 4.1. The eOGS not only provides a granular estimate (an enclosure of output data) but also a numerical estimate through the use of $x^{[h]}$ in (26), as shown in (8).

Both numerical and granular estimates are meaningful in many contexts, such as time series prediction, decision-making under uncertainty, and dynamic systems control. In the specific context of BD diagnosis addressed in this paper, only the numerical estimate is needed. Therefore, while both estimates have their uses, the numerical estimate is the most relevant for the problem formulated in this paper.

4.7. eOGS: incremental learning algorithm

The eOGS learning algorithm consists of the steps provided below. We emphasize that the algorithm is deterministic, as it produces the same model for the same data flow.

BEGIN
Set hyperparameters $\mathbb{P} = \{\alpha, \Psi, \nu, \omega\}$ (default values can be chosen: Section 5.3);
Specify a priority objective f_s;
Specify admissible values for the remaining objectives $\epsilon_t \forall t$;
for $h = 1, \ldots$ // Data stream
  Read input data $x^{[h]}$;
  if $h = 1$
    Create the first granule γ^1 and rule R^1 (Eqs. (9)-(13) and (4));
    Provide numerical $\bar{y}^{[h]}$ (8), and granular $[u^1, U^1]^{[h]}$ (26) estimates;
  else
    Provide numerical $\bar{y}^{[h]}$ (8), and granular $[u^i, U^i]^{[h]}$ (26) estimates;
    // The actual output $y^{[h]}$ becomes available;
    if $x_j^{[h]} \notin [l_j^i, L_j^i]$ $\forall i$ and some j // Structural learning
      Create granule γ^{c+1}, local functions $\Pi_{j,k}^i$ and p_k^i, and rule R^{c+1} (Eqs. (9)-(13) and (4));
    else // Parametric learning
      Update active granules γ^i (Eqs. (14)-(17) and (4));
      Update coefficients $a_{jk}^i, \forall j, k$, using any variation of RLS algorithm, e.g. [50];
    end
  end
  // Structural and hyperparametric optimization
  Merge granules and rules if needed (Eqs. (18)-(22), (4), (17));
  Delete inactive granules and rules if needed (Section 4.5);
  Update the hyperparameters $\mathbb{P}$ to minimize f_s while respecting $\epsilon_t \forall t$ (Section 5.3);
end
END

5. Methodology

We describe the BD classification task, exploratory data analysis, and how we rank and select speech attributes. The autonomous operation and initial setup of the eOGS algorithm for the application is given. Alternative evolving intelligent methods and their hyperparameters for performance comparison are outlined.

5.1. Dataset characterization and Pearson-Spearman attribute selection

The BD dataset consists of spectral and prosodic acoustic attributes from the speech of an individual during multiple phone calls made over 7 months to the Psychiatry department of a hospital. The attributes derived from speech recordings, which intend to reveal mood swings that range from depression to mania, are:

x_1: Local Jitter → Local Simple Moving Average (SMA) jitter;
x_2: Local Jitter Derivative → First order derivative of local SMA jitter;
x_3: Local Shimmer → Local SMA shimmer;
x_4: Spectral Flux → Pulse Code Modulation (PCM), average magnitude after Fast Fourier Transform (FFT);
x_5: Spectral Centroid → PCM, average magnitude after FFT;
x_6: Spectral Harmonic → PCM, average magnitude after FFT;
x_7: Final Fundamental Frequency → Final SMA;
x_8: Fundamental Frequency 10% Envelope → Env SMA (useful to identify trends);
x_9: Loudness → Auditory spectrum components, SMA 3;
x_{10}: Log Energy → PCM, average magnitude after FFT.

The dataset contains 1963 ten-attribute instances, i.e., $x^{[h]} = (x_1 \ \ldots \ x_{10})^{[h]}$, $h = 1, \ldots, 1963$. Attributes are linearly and independently re-scaled into the $[0, 1]$ range. The target or output variable, $y^{[h]}$, is a class $C = \{C_1, C_2, C_3\}$, in which C_1 stands for 'Bipolar Disorder'; C_2 is an intermediate state in which psychiatrists are not able to confirm the occurrence of the disorder or not, i.e., C_2 means 'Unknown' or 'Potential state of Disorder'; and C_3 attests the 'Healthy' state of the individual. The true class $y^{[h]}$ for instances $x^{[h]}$ were annotated by expert psychiatrists. We assume human experts as being right, and evolving computational models as approximators. The dataset contains 357, 1461, and 145 instances of each class, C_1, C_2, and C_3, respectively.

Estimated outputs $\bar{y}^{[h]}$ provided by the evolving models are real values in the $[0, 1]$ range. In case a sharp decision is required, such real values are rounded to the closest class ($C_1 \rightarrow y = 0$; $C_2 \rightarrow y = 0.5$; and $C_3 \rightarrow y = 1$). For example, an estimate $\bar{y}^{[h]} = 0.15$ is closer to C_1 (disorder); an estimate $\bar{y}^{[h]} = 0.35$ is closer to C_2 (potential state of disorder). The original data are summarized in Fig. 2 by using box-and-whisker plots to illustrate the mean, median, percentiles, symmetry and variance, per attribute and per class. Notice

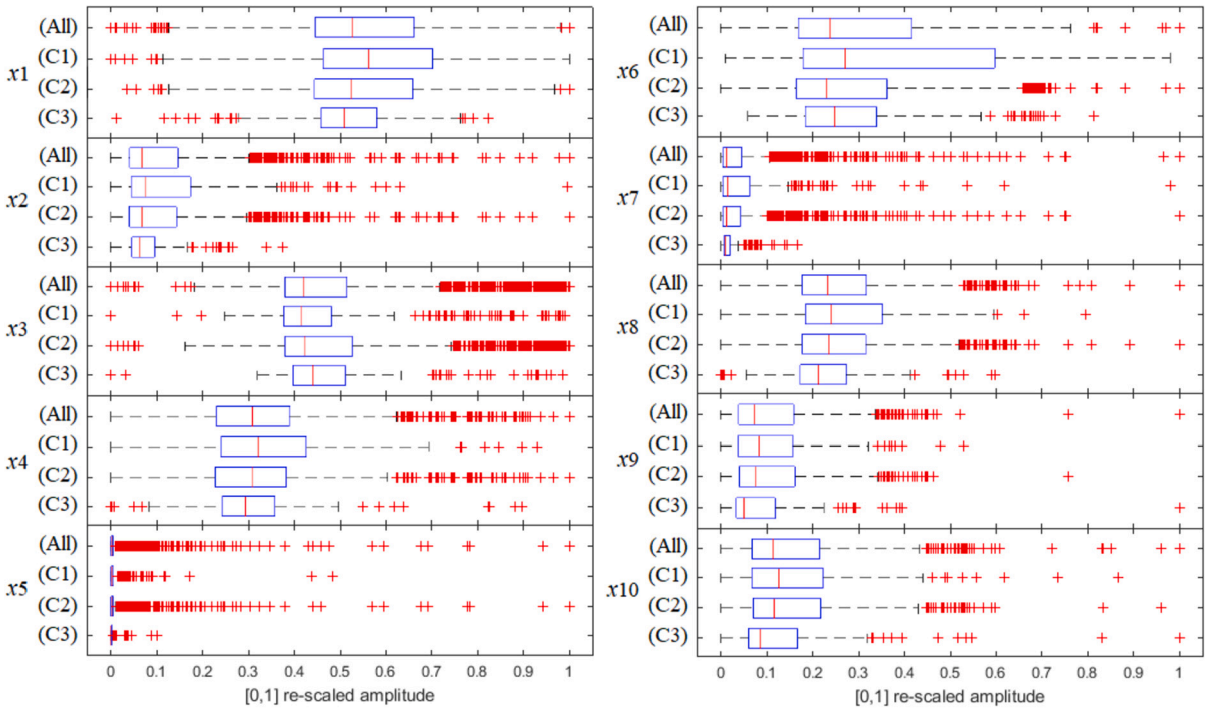


Fig. 2. Bipolar disorder data: box plot with summarized information about symmetry, skew and variance per attribute and per class. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

highly-superposed statistical components, and difficulty in identifying class-distinctive attributes at first glance – which is common in the psychiatric field. Online machine learning methods should identify spatio-temporal patterns on streaming data for effectiveness.

In Fig. 2, assume, for the sake of argument, that a BP class manifests in a single region of the domain of a feature x_j. Then, its representative data could be modeled by a single box (25th-75th percentiles), with no red plus marks outside the whiskers. The whiskers extend to the most extreme data that are not outliers. Outliers are plotted individually using red plus markers to illustrate that the acoustic features require multiple granulations per class and nonlinear class decision boundaries. In other words, the BD classification problem is complex. The features have low correlation with the classes, and incremental learning applied to data-driven nonlinear models is useful for continuously refining the model to achieve better predictions.

To preserve a level of interpretability of evolving models in the medical domain, feature extraction and space projection methods, such as variations of Principal Component Analysis or Linear Discriminant Analysis, are avoided. Instead, linear and monotonic correlation analysis [53] by averaging Pearson and Spearman coefficients are carried out. After attribute ranking, we perform a simple leave-one-attribute-out-at-a-time method to possibly attain more compact evolving models (with less input attributes and parameters) which may nonetheless achieve equal or greater class recognition accuracy. The proposed hybrid Pearson-Spearman coefficient assigned to each attribute plays a similar role to that of the Shapley Additive Explanation tool from the cooperative game theory [44], which aims to provide an explanation to a particular prediction by means of additive feature attributions. Given a model and an input vector, the Shapley method assigns each attribute an importance weight, measuring its marginal relevance to the model output. Fig. 3 shows the result of the combined Pearson-Spearman correlation.

The combined Pearson-Spearman correlation coefficient between a specific input x_j and the output y, say $\mathfrak{S}_j$, is given by the blue bars in Fig. 3. The higher $\mathfrak{S}_j$, the most discriminant the attribute x_j. Correlation analysis suggests an order for attribute removal, viz.:

1st to remove: x_6 ($\mathfrak{S}_6 = 0.0283$),
2nd: x_2 ($\mathfrak{S}_2 = 0.0313$),
3rd: x_1 ($\mathfrak{S}_1 = 0.0380$),
4th: x_7 ($\mathfrak{S}_7 = 0.0438$),
5th: x_3 ($\mathfrak{S}_3 = 0.0499$),
6th: x_{10} ($\mathfrak{S}_{10} = 0.0562$),
7th: x_9 ($\mathfrak{S}_9 = 0.0566$),
8th: x_4 ($\mathfrak{S}_4 = 0.0602$),
9th: x_5 ($\mathfrak{S}_5 = 0.0685$)

– being x_8 ($\mathfrak{S}_8 = 0.0747$) the variable that carries the most significant information to assist evolving classifiers.

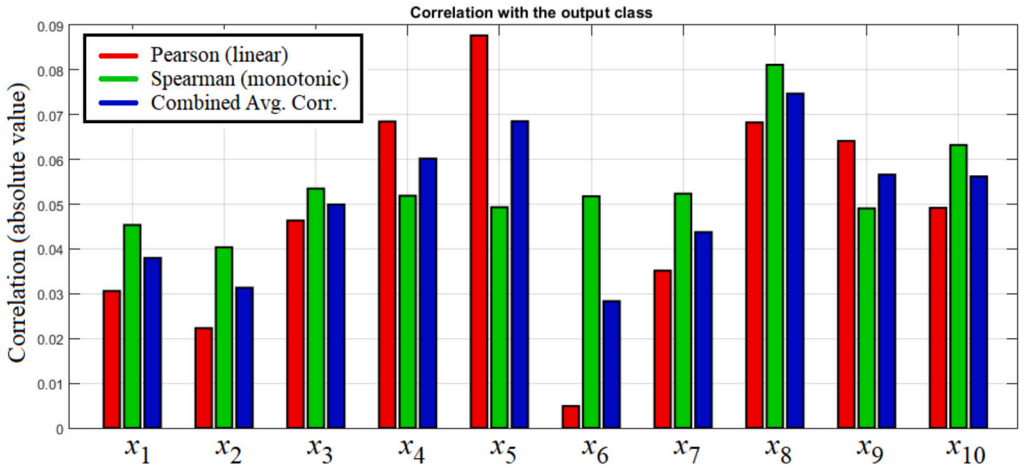


Fig. 3. Linear (Pearson), monotonic (Spearman), and combined correlation between the 10 acoustic attributes and the output class.

5.2. Error indices

The root mean square error (RMSE) is a commonly used metric to evaluate the accuracy of numerical predictions. Being H the current amount of processed instances, the RMSE is calculated as

$$\text{RMSE} = \frac{1}{H}\sum_{h=1}^{H}\sqrt{\frac{1}{m}\sum_{k=1}^{m}(y_k^{[h]} - \bar{y}_k^{[h]})^2}, \tag{27}$$

in which m is the number of outputs ($m = 1$ in the present BD classification task). The RMSE measures the average deviation between the predicted and the actual values.

An error metric for granular predictions that takes into consideration both the inclusion of the actual output values and the narrowness of the predicted enclosure is the mean granular error (MGE) [21],

$$\text{MGE} = \frac{1}{kH}\sum_{h=1}^{H}\sum_{k=1}^{m} 1 - \mathfrak{z}_k^{[h]}(1 - (U_k^{[h]} - u_k^{[h]})), \tag{28}$$

in which

$$\mathfrak{z}_k^{[h]} = \begin{cases} 1 & \text{if } y_k^{[h]} \in [u_k, U_k]^{[h]} \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

Although the actual value $y_k^{[h]}$ can always be included for wider interval predictions $[u_k, U_k]^{[h]}$, the factor $1 - (U_k^{[h]} - u_k^{[h]})$ penalizes such wider intervals that include the actual value. The more constricted the bounds $[u_k, U_k]^{[h]}$ that enclose $y_k^{[h]}$, the smaller the MGE, but the riskier it is that $y_k^{[h]}$ lies beyond the limits.

5.3. eOGS: autonomous operation and hyperparameters

A common formulation of the optimization problem (7) selects the numerical error as the primary objective. The goal is to minimize the prediction error observing the following constraints:

$$\begin{aligned} F(\mathbb{P}) &= \min \quad \text{RMSE} \\ \text{s.t.} \quad & \text{MGE} \leq \epsilon_1 \\ & c \leq \epsilon_2 \\ & \text{Spa}(\gamma) \geq \epsilon_3 \end{aligned} \tag{30}$$

The parameters $\mathbb{P} = \{\alpha, \Psi, \nu, \omega\}$ are automatically adjusted by the eOGS algorithm to handle violations of constraints in the optimization problem (30). If a constraint is violated, the algorithm should react as quickly as possible by adjusting these parameters. We propose default initial values for $\mathbb{P}$, which are $\alpha = 0.1$, $\Psi = 2$, $\nu = 200$, and $\omega = 0.005$. If a small set of instances is available *a priori*, a data stream can be simulated within seconds, and it may be possible to find a more appropriate initial set of values for $\mathbb{P}$.

During the online operation of the algorithm, if a constraint is violated, the parameters $\mathbb{P}$ are modified as follows:

- If MGE exceeds the threshold ϵ_1 (30), then: (i) the α cut is stepped up; and (ii) the initial variance σ^2 for new granules (see (10)) is stepped down. However, the value of σ^2 is constrained to be within a minimal and the Stigler value, namely, $0.01 \leq \sigma^2 \leq 1/2\pi$.
- If c exceeds the threshold ϵ_2 (30), then: (i) the α cut is stepped down; and (ii) the merging parameter ω is stepped up. After a few time steps, when c returns to a feasible value, ω is reset to its default value.
- If Spa(γ) falls below the threshold ϵ_3 (30), then: (i) the α cut is stepped up; and (ii) the Gaussians' expansion-contraction parameter Ψ is stepped down. After a few time steps, when Spa(γ) returns to a feasible value, Ψ is reset to its default value.

If the threshold values ϵ_t are set to unrealistically low or high values, the parameters $\mathbb{P}$ reach their limits and prevent further adjustment. In this case, eOGS will present the best possible solution that it can achieve given the constraints imposed. Selecting suitable values for ϵ_t is an important aspect of using eOGS in practice.

For the BD diagnosis problem, the right-hand sides of the constraints were set to specific values: $\epsilon_1 = 0.5$, $\epsilon_2 = 15$, and $\epsilon_3 = 0.6$. These values have the following implications for the eOGS model: (i) The MGE (28)-(29) has a relatively high acceptable value of 0.5. This is because in classification problems, the primary concern is with numerical estimates and the closest class; (ii) the eOGS model should retain no more than 15 granules to offer 15 decision rules at most to psychiatrists. Having more rules than ϵ_2 is assumed to hinder interpretability from a human perspective; and (iii) the average specificity for Gaussians (4) should be greater than ϵ_3, which is set to 0.6. This is to ensure that the granules are not too wide, as this could prevent eOGS from effectively modeling important details and patterns in the data with sufficient nonlinearity.

5.4. Alternative learning methods for comparisons

In order to evaluate the performance of eOGS in the BD diagnosis problem, other machine learning methods are used for comparison. The methods include: evolving Takagi-Sugeno (eTS) [29], eXtended Takagi-Sugeno (xTS) [30], Fuzzy-set-Based evolving Modeling (FBeM-Gauss) [31], and Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) [32].

The eTS modeling method from data streams [29] includes an online clustering algorithm and a weighted version of Recursive Least-Squares algorithm. eTS uses the concept of potential of new data instances given by a Cauchy-type function – similar to the well-known Subtractive Clustering algorithm. After grid search aiming the best accuracy in BD diagnosis using less than or approximately 15 clusters, eTS uses the hyperparameters: radius $R = 0.3$, and $\omega = 50$. The xTS [30] method combines zero and first-order Takagi-Sugeno fuzzy models. In addition to the concept of data potential, xTS introduces ideas such as adaptive zone of influence to learn the data distribution for each cluster, exclusion of contradictory rules, and a measure of cluster quality based on their age and support. After grid search, with focus on the best accuracy using a maximum of 15 rules, xTS makes use of the hyperparameters: $\omega = 50$ and 'local optimization', for the BD recognition task.

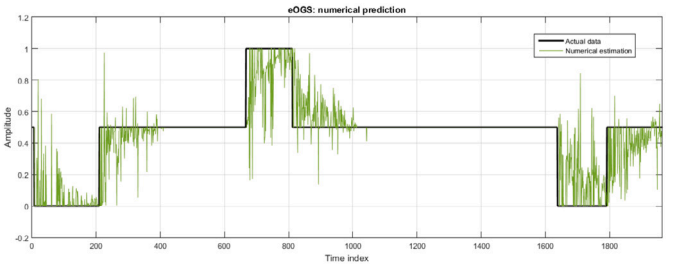
FBeM-Gauss [31] [54] is an incremental method that utilizes Gaussian fuzzy granules to represent uncertainty in the data. A scattering-type clustering procedure and structural updating equations are addressed in [54]. The modeling approach allows experts to express their domain knowledge. FBeM-Gauss supports both online learning from data streams and learning from experience. We used FBeM-Gauss for BD diagnosis and performed grid search optimization to set the best hyperparameters that offer high accuracy using less than 15 rules. The hyperparameters are $\rho^{[0]} = 0.03$, $\alpha = 0.05$, $\Delta = 0.1$, and $h_r = 200$. DENFIS [32] is a neural net that can learn spatio-temporal sequences through a hybrid supervised/unsupervised learning method. We consider the dynamic creation of a first-order Takagi-Sugeno rule base for DENFIS, which is accomplished by means of the evolving Clustering Method, eCM, and a version of Recursive Least Squares [32]. Similar to FBeM-Gauss, experts can insert rules into DENFIS prior to and during the learning process. We considered learning from scratch for all methods. Grid search is performed to find the best hyperparameters for DENFIS, which are $d_{thr} = 0.18$, and an optimization interval $I = 10$ for BD classification.

6. Results and discussions

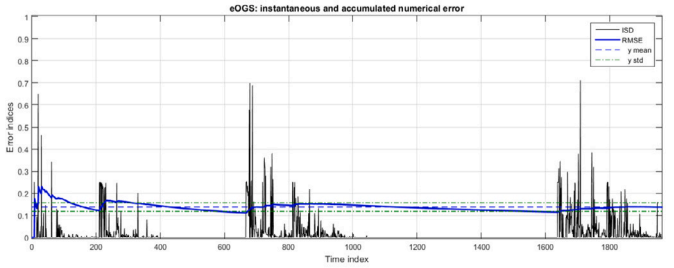
Experiments based on speech data produced by an individual with BD, expressing emotional mood swings, are carried out to verify the effectiveness of the incremental learning methods. The methods are: the proposed eOGS, and eTS, xTS, FBeM-Gauss, and DENFIS. The goal is to provide compact and accurate models. Automated data processing in the diagnosis of BD helps providing quantitative medical indicators and allows monitoring patients for longer periods. We highlight eOGS linguistic rules to support medical decision making.

6.1. eOGS for bipolar disorder diagnosis

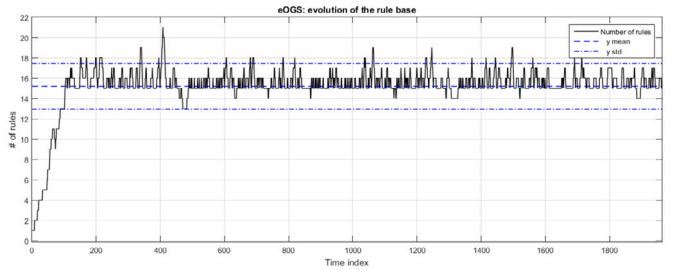
Speech-extracted time-dependent data $x^{[h]}$, $h = 1, ..., 1963$, are presented to eOGS. In the first experiment, full 10-attribute vectors are processed. The respective eOGS model is built from scratch – with no rules nor pre-training. Notice that the application of cross-validation and separation of train and test data is not in question since the data is not maintained in memory. An instance of the data stream is read, an estimation is provided by the current model, and finally, an incremental learning step is applied to update the parameters and model structure (instance-per-instance test-before-training approach). We use the default initial hyperparameters (Section 5.3). Experiments run on an i7-8550U CPU dual-core 1.88-1.99 GHz processor with 8 GB of RAM, and on a Windows 11 operating system. Fig. 4(a) shows the numerical estimates of the classes ($C_1 \rightarrow y = 0$, the 'Bipolar Disorder' condition; $C_2 \rightarrow y = 0.5$, the 'Potential state of Disorder'; and $C_3 \rightarrow y = 1$, the 'Healthy' state). The estimates are rounded to the closest class for sharp decision making. Moreover, Figs. 4(b)-(e) display, respectively, the evolution of the RMSE and instantaneous error; the number of eOGS rules



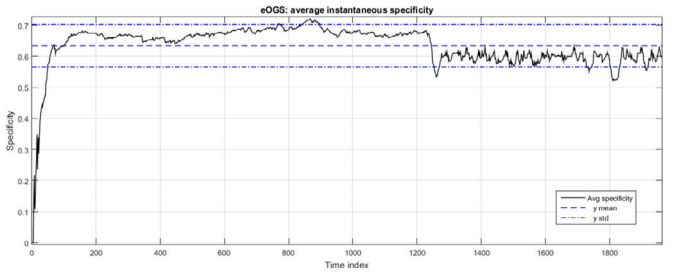
(a)



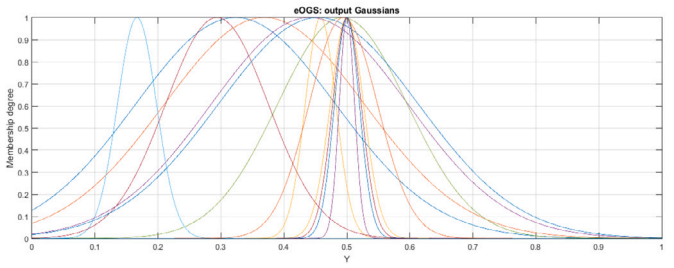
(b)



(c)



(d)



(e)

Fig. 4. eOGS results: (a) Numerical prediction of the bipolar disorder function; Amplitude 0: Bipolar Disorder; 0.5: Undefined State or Potential Disorder; 1: Healthy; (b) Evolution of the RMSE and instantaneous error; (c) Evolution of the granular rule base; (d) Average Gaussian specificity; (e) Snapshot of the dynamic output Gaussians at $h = 1963$.

over time; the average specificity of the granular map; and a snapshot of the position of the output Gaussians at $h = 1963$ as an example.

Notice in Fig. 4(a) that the actual class $y^{[h]}$ associated to the current input $x^{[h]}$ is represented by the black line according to psychiatrists. eOGS does not access this information at first. After providing an estimate $\bar{y}^{[h]}$, the algorithm updates active granules γ^{i*} that form antecedent terms of rules. This is done in an unsupervised way in the sense of paving the data space with either new granules or more appropriate locations to existing ones. The objective is to provide better estimates the next time a similar input is presented. In the near future, the true or annotated class $y^{[h]}$ becomes available, and a step of the Recursive Least Squares algorithm updates the coefficients a_j^{i*}, $j = 0, ..., n$, of the consequent function using the pair $(x, y)^{[h]}$. We notice in Fig. 4(a) that the prediction in green tracks the true class (the drifts in the speech) by means of eOGS nonlinear capacity and adaptive parameters.

Fig. 4(b) corroborates to the incremental learning potential of eOGS. The model incorporates and tracks the novelties reflected by the abrupt changes of the data. We remark that the class transients may not necessarily be associated to modeling errors, but partially to dubious or hesitant intermediate occurrences between classes, and the need of a prompt and sharp class decision by psychiatrists. Transients are interesting information from the medical point of view since the onset of mania and depression episodes can be prevented during the transients. The RMSE reduced to 0.1370 over time, namely, the model is effectively learning.

The autonomous evolution of the rule base from scratch is shown in Fig. 4(c). The model produced approximately 15 ± 2.3 rules to represent the speech patterns. This number of rules respects the users' choice, which is expressed by the parameter ϵ_2 in the constrained optimization problem (30) that guides the learning process (Section 5.3). The intuition behind choosing an ϵ_2 is to ensure a reasonable number of rules to support medical decisions at any time. Note that setting a maximum is not required. In this case, ϵ_2 can be any large number.

Aiming to understand the reasoning behind a prediction, the idea is to zoom in on a single instance. This is done by verifying the most active granules (a small region of interest) and their rules. This surrogate local model, while not optimal, is a local approximation with higher interpretability degree. Locally, the prediction might depend linearly on the available data. An example of a highly active eOGS rule – considering an 8-attribute model (dimensionality reduction discussed in the next section) – at the time step $h = 1963$ is:

R^i: IF (x_1 'LOCAL JITTER' is 'LOW' $[-0.1424, 0.3626]$) AND
(x_3 'LOCAL SHIMMER' is 'LOW' $[-0.0426, 0.4699]$) AND
(x_4 'SPECTRAL FLUX' is 'VERY LOW' $[-0.2005, 0.2253]$) AND
(x_5 'SPECTRAL CENTROID' is 'LOW' $[0.0105, 0.4710]$) AND
(x_7 'FINAL FUNDAMENTAL FREQUENCY' is 'MEDIUM' $[0.0443, 0.5597]$) AND
(x_8 'FUNDAMENTAL FREQUENCY 10% ENVELOPE' is 'MEDIUM' $[0.1574, 0.6725]$) AND
(x_9 'LOUDNESS' is 'VERY LOW' $[-0.1832, 0.3134]$) AND
(x_{10} 'LOG ENERGY' is 'MEDIUM' $[0.2507, 0.7659]$),
THEN (y_1 'PREDICTED CLASS' is 'BP.DISORDER' $[0.1828, 0.4659]$) AND
($\bar{y}^i = 0.9140 + 0.6680x_1 - 0.7578x_3 - 11.2237x_4 + 0.3251x_5 + 0.1961x_7 - 0.9475x_8 + 4.4705x_9 - 0.6062x_{10}$)

Notice: (i) the pointwise estimate $\bar{y}^i$ is local – partial influence of other active rules will further compose the global estimate; (ii) the granular estimate $[u^i, U^i]$ limits $\bar{y}^i$; and (iii) typically from 3 to 5 labels, namely, 'very low', 'low', 'medium', 'high', and 'very high', can be assigned per attribute, which depends on the order of the approximately 15 granules along the attribute axis and a subjective expert decision about the linguistic understandability of the model. Moreover, the consequent coefficients $a^i = (0.9140, \ ..., \ -0.6062)$ say about the local sensitivity of the output with respect to specific attributes x_j, which adds explainability to the model for decision support.

Finally, Figs. 4(d) and 4(e) show the evolution of the specificity $\text{Spa}(\gamma)$ of the Gaussians involved in the granular map, and the shape of the output Gaussians at $h = 1963$. As ϵ_3 was set to 0.6 (Section 5.3), we notice a surplus value for the respective constraint until $h = 1250$. This means that eOGS worked on the optimization problem with no concern about the width of the granules. The constraint binds at $h = 1250$ since some Gaussians – specifically those representing the C_2 class whose data exemplars were persistently incoming – started to spread too much on such area of the data space. The algorithm reacted at $h = 1250$ to avoid a spread larger than that acceptable (given by ϵ_3) by stepping up α to reduce the level set (see effect on the size of the granules in Eqs. (11), (12), and (17)), and stepping down Ψ (see outcome in Eq. (15)) to compel granule contraction. The specificity was maintained at the acceptable ϵ_3 level – with no perceptible effect on the overall performance.

The snapshot of the output Gaussians covers especially the most recent classes, i.e., 'Bipolar Disorder' ($y = 0$), and 'Potential State of Disorder' ($y = 0.5$). Gaussian granules move according to the available data. Snapshots of granules and rules representing any of the classes can be taken at any time for further decision making support. In general, as a decision support system, the result may recommend, for example, new medications or hospital visits. The total CPU time to process 1963 instances was 2.31 seconds, which indicates that the algorithm can operate in real-time considering high-frequency data streams produced by a number of acoustic indicators and multimodalities.

6.2. Comparison with other incremental methods

Alternative models, namely, eTS, xTS, FBeM-Gauss, and DENFIS, were developed using both complete input vectors $(x_1, ..., x_{10})^{[h]}$, and reduced vectors after attribute selection. The order of attribute removal based on the Pearson-Spearman coefficient $\mathfrak{S}_j$ (Section 5.1) is x_6, x_2, x_1, x_7, x_3, x_{10}, x_9, x_4, and x_5. We leave one attribute out at a time to possibly attain more compact models. The

Table 1
Bipolar disorder diagnosis: Evaluating different incremental learning methods and models.

Model	# Attributes	# Rules	$RMSE$
eTS [29]	10	15.0	0.2420
	9	15.0	0.2413
	8	17.0	0.2430
	7	15.0	0.2423
	6	**14.0**	**0.2357**
	5	13.0	0.2426
	4	11.0	0.2432
	3	10.0	0.2426
	2	8.0	0.2391
	Average (5th) →		**0.2413**
xTS [30]	10	11.0	0.2176
	9	11.0	0.2144
	8	12.0	0.2108
	7	11.0	0.2184
	6	8.0	0.2193
	5	9.0	0.2237
	4	8.0	0.2322
	3	7.0	0.2337
	2	**6.0**	**0.2095**
	Average (4th) →		**0.2200**
FBeM-Gauss [31]	**10**	**14.8**	**0.1529**
	9	15.1	0.1749
	8	14.6	0.1726
	7	13.9	0.1731
	6	11.7	0.1764
	5	9.9	0.1626
	4	8.2	0.1862
	3	6.6	0.1720
	2	5.1	0.1925
	Average (3th) →		**0.1737**
DENFIS [32]	10	14.0	0.1740
	9	15.0	0.1730
	8	16.0	0.1700
	7	15.0	0.1660
	6	13.0	0.1650
	5	12.0	0.1580
	4	9.0	0.1520
	3	10.0	0.1530
	2	**10.0**	**0.1440**
	Average (2nd) →		**0.1617**
eOGS (This paper)	10	15.2	0.1370
	9	15.2	0.1378
	8	**15.3**	**0.1361**
	7	15.2	0.1390
	6	15.1	0.1434
	5	14.9	0.1490
	4	14.6	0.1525
	3	14.0	0.1638
	2	7.4	0.1728
	Average (1st) →		**0.1479**

best hyperparameters of each method were obtained through grid search (Sections 5.3 and 5.4). Table 1 compares eOGS and the other incremental methods.

Table 1 shows that eOGS overcomes DENFIS and FBeM-Gauss by a small margin in both best and average cases. The best model – eOGS using 8 attributes (PCM of the spectral harmonic, and first order derivative of the local jitter deleted) – achieved a mean estimation error of 0.1361. Interestingly, while eOGS, eTS and FBeM-Gauss reached the best accuracy using a larger number of attributes, 8, 6 and 10, respectively; xTS and DENFIS reached their best predictions using only 2 attributes (fundamental frequency, and PCM of the spectral centroid). Such result may suggest that information about the bipolar disorder conveyed by the original speech attributes is highly superposed and/or white noise composes a great part of the data. xTS and DENFIS models reveal such class overlapping and a level of attribute redundancy. Contrariwise, FBeM-Gauss and eOGS managed to learn small nuances in other dimensions x_j, therefore showing a significant RMSE enhancement, about 0.04, when inspecting more attributes.

By rounding real-valued estimates to the closest class for a sharp classification outcome, the best 8-attribute eOGS model has exhibited an accuracy of 91.85%. The Confusion matrix of such eOGS model, as shown in Fig. 5, offers new insights. We notice

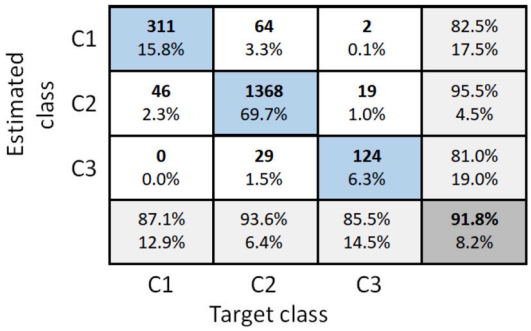


Fig. 5. Confusion Matrix of the 8-attribute eOGS model: the best overall case.

that the majority of wrong classifications involve the undefined intermediate class, C_2 (target class). The C_2 estimates inevitably call further psychiatric attention and deeper analysis as a potential indication of the disorder onset. Confusions between the extreme cases, $C_1 \rightarrow C_3$ and $C_3 \rightarrow C_1$, are certainly more inconvenient since they may mislead physicians in explaining the symptoms and making a diagnosis. In this context, only 2 instances out of 437 were wrongly classified.

To summarize, the eOGS incremental algorithm and model have shown promising outcomes in diagnosing bipolar disorder by utilizing attributes extracted from the human voice and speech. The model evolves continuously over time and can identify and track new patterns in the data stream autonomously. The modeling approach balances multiple objectives to give not only accurate predictions but also decision rules with a degree of interpretability to aid medical decision making. The domain of speech attributes is granulated. Linguistic labels are assigned to granules to enhance understandability and comprehension.

7. Conclusion

Bipolar mood disorder diagnosis based on speech data has been addressed by using Evolving Granular Computing. A general-purpose incremental learning approach, eOGS, performs optimal design of the structure and parameters of a rule-based model from acoustic attributes obtained from phone calls. We present a multi-objective formulation that trades-off information specificity, model compactness, and numerical and granular error indices. Granules associated to linguistic labels and a granular input-output map give a high level of human understandability with relation to the inherent process of producing class predictions.

A stream with 10 spectral and prosodic acoustic attributes is taken into consideration. Attributes were ranked and selected based on a hybrid Pearson-Spearman correlation coefficient. Low input-output correlations – varying from 0.0283 to 0.0747 – were noticed, as well as high class overlap, which is typical in the psychiatric field. The best eOGS hyperparameters as well as those of alternative incremental methods, namely, eTS, xTS, FBeM-Gauss, and DENFIS, were determined from grid search optimization. The search focused not only on estimation accuracy, but also on the development of concise classifiers using a maximum of 15 rules, which was considered a sufficient amount of rules to support a level of model interpretability and medical decision making.

eOGS models overcame DENFIS and FBeM-Gauss models in a 3-class classification setup ('bipolar disorder', 'potential state of disorder', and 'healthy' classes) by a small margin considering both the best (RMSE = 0.1361 against 0.1440 and 0.1529, respectively) and average (RMSE = 0.1479 against 0.1617 and 0.1737) cases; followed by xTS and eTS models. eOGS using only 8 attributes (PCM of the spectral harmonic, and first order derivative of the local jitter eliminated) reached the best overall result. It exhibited an accuracy of 91.85% in sharp class estimates. We notice from its confusion matrix that the majority of wrong classifications involve the intermediate class, which would inevitably call further clinical attention as a potential indication of the disorder onset. Nonetheless, a minimal number of confusions between the extreme healthy-disorder classes happened (2 out of 437 cases). Such confusions are inconvenient since they may mislead physicians in explaining symptoms and making a diagnosis.

The advantage of the incremental granular method over other machine learning methods lies in its granular structure as additional information to assist decision making, and in the possibility to trade-off multiple objectives. In the future, attributes from wavelet transforms will be investigated. Additionally, the 'bipolar disorder' class will be decomposed in 3 diagnoses: bipolar I, bipolar II, and cyclothymia.

CRediT authorship contribution statement

Daniel Leite: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Gabriella Casalino:** Writing – review & editing, Investigation, Formal analysis. **Katarzyna Kaczmarek-Majer:** Writing – review & editing, Data curation, Investigation. **Giovanna Castellano:** Writing – review & editing, Investigation, Formal analysis.

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.

Acknowledgement

The first author acknowledges Project SAIL funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia under the grant no. NW21-059D. The second author acknowledges funding from the European Union PON project Ricerca e Innovazione 2014-2020, DM 1062/2021. The third author is supported by the project "Research of Excellence on Digital Technologies and Well-being CZ.02.01.01/00/22-008/0004583", co-financed by the EU. The fourth author acknowledges the support of the PNRR project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI (CUP H97G22000210007) under the NRRP MUR program funded by the NextGenerationEU.

Data availability

Data will be made available on request.

References

[1] D. Low, K. Bentley, S. Ghosh, Automated assessment of psychiatric disorders using speech: a systematic review, Laryngoscope Investig. Otolaryngol. 5 (1) (2020) 96–116.
[2] D. Highland, G. Zhou, A review of detection techniques for depression and bipolar disorder, Smart Health 24 (2022) 100282.
[3] M. Faurholt-Jepsen, J. Busk, M. Frost, M. Vinberg, E. Christensen, O. Winther, J. Bardram, L. Kessing, Voice analysis as an objective state marker in bipolar disorder, Transl. Psychiatry 6 (2016) e856.
[4] A. Maxhuni, A. Munoz-Melendez, V. Osmani, H. Perez, O. Mayora, E. Morales, Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients, Pervasive Mob. Comput. 31 (2016) 50–66.
[5] N. Abaeikoupaei, H.A. Osman, A multi-modal stacked ensemble model for bipolar disorder classification, IEEE Trans. Affect. Comput. 14 (1) (2023) 236–244.
[6] P. Mahajan, V. Baths, Acoustic and language based deep learning approaches for Alzheimer's dementia detection from spontaneous speech, Front. Aging Neurosci. 13 (2021) 623607.
[7] C. Sakar, G. Serbes, A. Gunduz, H. Tunc, H. Nizam, B. Sakar, M. Tutuncu, T. Aydin, M. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, Appl. Soft Comput. 74 (2019) 255–263.
[8] D.V. Carvalho, E.M. Pereira, J.S. Cardoso, Machine learning interpretability: a survey on methods and metrics, Electronics 8 (832) (2019) 1–34.
[9] D. Leite, A. Sharma, C. Demir, A.-C. Ngomo, Interpretability index based on balanced volumes for transparent models and agnostic explainers, in: World Congress on Computational Intelligence, WCCI, FUZZ-IEEE, Yokohama, 2024, pp. 1–10.
[10] W. Silva, K. Fernandes, M.J. Cardoso, J.S. Cardoso, Towards complementary explanations using deep neural networks, in: Lecture Notes in Computer Science: Understanding and Interpreting Machine Learning in Medical Image Computing Applications, Springer, 2018, pp. 133–140.
[11] D. Leite, I. Skrjanc, S. Blazic, A. Zdesar, F. Gomide, Interval incremental learning of interval data streams and application to vehicle tracking, Inf. Sci. 630 (2023) 1–22.
[12] D. Wang, H. Liu, W. Pedrycz, W. Song, H. Li, Design gaussian information granule based on the principle of justifiable granularity: a multi-dimensional perspective, Expert Syst. Appl. 197 (2022) 116763.
[13] D. Leite, P. Costa, F. Gomide, Evolving granular neural networks from fuzzy data streams, Neural Netw. 38 (2013) 1–16.
[14] Z.C. Lipton, The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery, ACM Queue: New York 16 (3) (2018) 31–57.
[15] T. Miller, Explanation in artificial intelligence: insights from the social sciences, Artif. Intell. 267 (2018) 1–38.
[16] B. Kim, R. Khanna, O. Koyejo, Examples are not enough, learn to criticize! Criticism for interpretability, in: Advances in Neural Information Processing Systems, 2016, pp. 2280–2288.
[17] L. Maciel, R. Ballini, F. Gomide, Evolving granular analytics for interval time series forecasting, Granul. Comput. 1 (2016) 213–224.
[18] Y. Yao, A triarchic theory of granular computing, Granul. Comput. 1 (2016) 145–157.
[19] A. Bargiela, W. Pedrycz, Granular mappings, IEEE Trans. Syst. Man Cybern., Part A 35 (2) (2005) 292–297.
[20] L. Zadeh, Generalized theory of uncertainty (GTU): principal concepts and ideas, Comput. Stat. Data Anal. 51 (1) (2006) 15–46.
[21] D. Leite, G. Andonovski, I. Skrjanc, F. Gomide, Optimal rule-based granular systems from data streams, IEEE Trans. Fuzzy Syst. 28 (3) (2020) 583–596.
[22] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (2019) 206–215.
[23] I. Skrjanc, J.A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, F. Gomide, Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: a survey, Inf. Sci. 490 (2019) 344–368.
[24] E. Lughofer, Evolving Fuzzy Systems: Methodologies, Advanced Concepts and Applications, Springer, 2011.
[25] N. Kasabov, Evolving and spiking connectionist systems for brain-inspired artificial intelligence, in: R. Kozma, C. Alippi, Y. Choe, F. Morabito (Eds.), Artificial Intelligence in the Age of Neural Networks and Brain Computing, Academic Press, 2019, pp. 111–138, Ch. 6.
[26] X. Gu, J. Han, Q. Shen, P. Angelov, Autonomous learning for fuzzy systems: a review, Artif. Intell. Rev. (2022) 1–47.
[27] P.V. Souza, E. Lughofer, EFNC-Exp: an evolving fuzzy neural classifier integrating expert rules and uncertainty, Fuzzy Sets Syst. (2022) 1–33.
[28] E. Lughofer, M. Pratama, Evolving multi-user fuzzy classifier system with advanced explainability and interpretability aspects, Inf. Fusion 91 (2023) 458–476.
[29] P. Angelov, D. Filev, An approach to online identification of Takagi-Sugeno fuzzy models, IEEE Trans. Syst. Man Cybern., Part B 34 (1) (2004) 484–498.
[30] P. Angelov, X. Zhou, Evolving fuzzy systems from data streams in real-time, in: International Symposium on Evolving Fuzzy Systems, 2006, pp. 29–35.
[31] D. Leite, R. Ballini, P. Costa, F. Gomide, Evolving fuzzy granular modeling from nonstationary fuzzy data streams, Evolv. Syst. 3 (2012) 65–79.
[32] N. Kasabov, Q. Song, DENFIS: dynamic evolving neural-fuzzy inference system and its application for time series prediction, IEEE Trans. Fuzzy Syst. 10 (2) (2002) 144–154.
[33] C.W. Espinola, J.C. Gomes, J.M.S. Pereira, W.P. dos Santos, Detection of major depressive disorder, bipolar disorder, schizophrenia and generalized anxiety disorder using vocal acoustic analysis and machine learning: an exploratory study, Res. Biomed. Eng. 38 (2022) 813–829.
[34] P. Llamocca, V. Lopez, M. Santos, M. Cukic, Personalized characterization of emotional states in patients with bipolar disorder, Mathematics 9 (11) (2021) 1–18.
[35] K.-Y. Huang, C.-H. Wu, M.-H. Su, Attention-based convolutional neural network and long short-term memory for short-term detection of mood disorders based on elicited speech responses, Pattern Recognit. 88 (2019) 668–678.
[36] P. Baki, H. Kaya, E. Ciftci, H. Gulec, A.A. Salahv, A multimodal approach for mania level prediction in bipolar disorder, IEEE Trans. Affect. Comput. 13 (4) (2022) 2119–2131.
[37] Z. Ren, J. Han, N. Cummins, Q. Kong, M. Plumbley, B. Schuller, Multi-instance learning for bipolar disorder diagnosis using weakly labelled speech data, in: International Conference on Digital Public Health, 2020, pp. 79–83.

[38] F. Ringeval, B. Schuller, M. Valstar, R. Cowie, H. Kaya, M. Schmitt, S. Amiriparian, N. Cummins, D. Lalanne, A. Michaud, E. Ciftci, H. Gulec, A.A. Salah, M. Pantic, Multi-instance learning for bipolar disorder diagnosis using weakly labelled speech data, in: AVEC 2018 Workshop and Challenge: Bipolar Disorder and Cross-Cultural Affect Recognition, 2018, pp. 3–13.
[39] K. Kaczmarek-Majer, G. Casalino, G. Castellano, M. Dominiak, O. Hryniewicz, O. Kaminska, G. Vessio, N. D.-Rodriguez, Plenary: explaining black-box models in natural language through fuzzy linguistic summaries, Inf. Sci. 614 (2022) 374–399.
[40] K. Kaczmarek-Majer, G. Casalino, G. Castellano, O. Hryniewicz, M. Dominiak, Explaining smartphone-based acoustic data in bipolar disorder: semi-supervised fuzzy clustering and relative linguistic summaries, Inf. Sci. 588 (2022) 174–195.
[41] K. Kaczmarek-Majer, G. Casalino, G. Castellano, D. Leite, O. Hryniewicz, Fuzzy linguistic summaries for explaining online semi-supervised learning, in: IEEE International Conference on Intelligent Systems, 2022, pp. 1–8.
[42] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets Syst. 90 (2) (1997) 111–127.
[43] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989.
[44] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: International Conference on Neural Information Processing Systems, 2017, pp. 4768–4777.
[45] J. Qin, L. Martínez, W. Pedrycz, X. Ma, Y. Liang, An overview of granular computing in decision-making: extensions, applications, and challenges, Inf. Fusion 98 (2023) 101833.
[46] W. Pedrycz, W. Homenda, Building the fundamentals of granular computing: a principle of justifiable granularity, Appl. Soft Comput. 13 (10) (2013) 4209–4218.
[47] D. Wang, W. Pedrycz, Z. Li, Design of granular interval-valued information granules with the use of the principle of justifiable granularity and their applications to system modeling of higher type, Soft Comput. 20 (6) (2016) 2119–2134.
[48] R. Yager, Measures of specificity over continuous spaces under similarity relations, Fuzzy Sets Syst. 159 (17) (2008) 2193–2210.
[49] V. Chankong, Y.Y. Haimes, Multiobjective Decision Making: Theory and Methodology, Courier Dover Publications, 2008.
[50] D. Leite, I. Skrjanc, Ensemble of evolving optimal granular experts, OWA aggregation, and time series prediction, Inf. Sci. 504 (2019) 95–112.
[51] R. Moore, R.B. Kearfott, M.J. Cloud, Introduction to Interval Analysis, Society for Industrial and Applied Mathematics, 2009.
[52] S.M. Stigler, A modest proposal: a new standard for the normal, Am. Stat. 36 (2) (1982) 137–138.
[53] Y. Dodge, The Concise Encyclopedia of Statistics, Springer-Verlag, New York, 2010.
[54] D. Leite, F. Gomide, R. Ballini, P. Costa, Fuzzy granular evolving modeling for time series prediction, in: IEEE International Conference on Fuzzy Systems, 2011, pp. 1–8.