Journal of Biomedical Informatics 134 (2022) 104172

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

Learning prognostic models using a mixture of biclustering and triclustering: Predicting the need for non-invasive ventilation in Amyotrophic Lateral Sclerosis

Diogo F. Soares [a,*], Rui Henriques [b], Marta Gromicho [c], Mamede de Carvalho [c], Sara C. Madeira [a,*]

[a] LASIGE, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
[b] INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[c] Instituto de Medicina Molecular, Instituto de Fisiologia, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

ARTICLE INFO

Keywords:
Biclustering
Triclustering
Three-way data
Prognostic
Disease progression patterns
Amyotrophic Lateral Sclerosis

ABSTRACT

Longitudinal cohort studies to study disease progression generally combine temporal features produced under periodic assessments (clinical follow-up) with static features associated with single-time assessments, genetic, psychophysiological, and demographic profiles. Subspace clustering, including biclustering and triclustering stances, enables the discovery of local and discriminative patterns from such multidimensional cohort data. These patterns, highly interpretable, are relevant to identifying groups of patients with similar traits or progression patterns. Despite their potential, their use for improving predictive tasks in clinical domains remains unexplored.

In this work, we propose to learn predictive models from static and temporal data using discriminative patterns, obtained via biclustering and triclustering, as features within a state-of-the-art classifier, thus enhancing model interpretation. `triCluster` is extended to find time-contiguous triclusters in temporal data (temporal patterns) and a biclustering algorithm to discover coherent patterns in static data. The transformed data space, composed of bicluster and tricluster features, capture local and cross-variable associations with discriminative power, yielding unique statistical properties of interest.

As a case study, we applied our methodology to follow-up data from Portuguese patients with Amyotrophic Lateral Sclerosis (ALS) to predict the need for non-invasive ventilation (NIV) since the last appointment. The results showed that, in general, our methodology outperformed baseline results using the original features. Furthermore, the bicluster/tricluster-based patterns used by the classifier can be used by clinicians to understand the models by highlighting relevant prognostic patterns.

1. Introduction

Subspace clustering aims to find clusters within different subspaces (a selection of one or more features) in multivariate data spaces. Its main advantage is the possibility of finding coherent clusters defined in only a subset of all features. Given these clusters, discriminate patterns can be extracted along their specific data subspaces [1]. In this context, biclustering performs simultaneous grouping of objects and features when analysing two-way data [2]. Triclustering extends biclustering to find groups of objects, features and contexts by enabling the simultaneous clustering of three-way data, where contexts generally refer to the temporal dimension [3].

When learning from high dimensional data spaces, feature selection, sparse priors, and data space transformations [4] are generally used to improve performance and avoid the curse of dimensionality. However, feature selection can prevent relevant features from being used if their relevance is only observed in a subgroup of observations or when analysed with other features (a subgroup of features). In addition, data transformations for dimensionality reduction generally hamper the interpretability of the learned models [5]. In this context, subspace clustering can be applied to find subsets of features corresponding to a subspace where a subgroup of observations show coherent values, overcoming the limitations of feature selection while improving predictive performance and promoting model interpretability pushed by the actionable self-explanatory nature of the found subspaces. In particular, in clinical domains, biclustering and triclustering are promising approaches that have been largely used separately to identify groups of patients with correlated clinical features along time [3,6], that could be used as discriminative patterns [6].

* Corresponding authors at: LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal.
E-mail addresses: dfsoares@ciencias.ulisboa.pt (D.F. Soares), sacmadeira@ciencias.ulisboa.pt (S.C. Madeira).

https://doi.org/10.1016/j.jbi.2022.104172
Received 22 October 2021; Received in revised form 31 March 2022; Accepted 15 August 2022
Available online 30 August 2022
1532-0464/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

In this scenario, we propose BicTric, a classification approach to learn predictive models using a mixture of biclustering and triclustering, that takes advantage of two types of data usually available in clinical contexts: static data, which does not change with time, and dynamic (temporal) data, collected over time. In longitudinal cohort studies, static data generally correspond to single-time exams, germline mutations, psychophysiological traits and demographic profiles. In contrast, temporal data generally correspond to periodic or irregular assessments associated with the clinical follow-up of individuals partaking in the cohort.

To this end, BicTric first searches for discriminative patterns resulting from biclustering static data and triclustering temporal data and then uses them as features within (interpretative) state-of-the-art classifiers.

Given the irregularity of follow-up assessments in longitudinal, our methodology further extends the principles proposed by Carreiro et al. [7] to compute learning examples from patient assessments of varying number and misalignment degrees.

Although the extent of contributions on feature extraction from complex clinical data domains [3,8,9], our study is the first bridging biclustering and triclustering stances for predictive tasks. BicTric can be positioned as a transformation from structured data spaces described by static and temporal variables into new pattern-centric data spaces able to model local cross-variable associations with unique discriminative properties of interest.

In this work, we performed a case study in patients suffering from Amyotrophic Lateral Sclerosis (ALS), a neurodegenerative disorder causing progressive loss of muscle control, with the aim of learning prognostic models to predict the need for non-invasive ventilation in a clinically relevant time window (90 days). Effective predictive models would allow anticipating respiratory insufficiency, a common symptom developed by ALS patients, in whom the rapid intervention/administration of non-invasive ventilation has shown to be effective in prolonging life and improving quality of life [10,11]. Data were gathered from a large cohort of Portuguese patients, where biclusters together with triclusters learnt from patients' follow-up data can be interpreted as disease progression patterns.

The major contributions are thus the following:

- a new classification approach, BicTric, to learn predictive models using a mixture of Biclustering and Triclustering;
- a new approach to create dynamic learning examples from varying number of misaligned temporal observations using windowing and sampling principles;
- a case study in ALS patients, where we learn prognostic models using BicTric able to predict the need for non-invasive ventilation using a set of clinical appointments (patients' follow-up), together with their static data;
- comprehensive assessment of BicTric predictive properties, revealing state-of-the-art performance and interpretive power, overcoming competitive baselines and unravelling disease progression patterns.

The remainder of the paper is organized as follows: Section 2 discusses related work, Section 3 formalize the subspace clustering problem stating the variables needed to understand the methodology, Section 4 describes the proposed BicTric methods, Section 5 concretize the study in ALS, Section 6 presents the BicTric results in ALS prognostic prediction, Section 7 discuss the results, and finally Section 8 concludes the paper.

2. Related work

The Revised ALS Functional Rating Scale (ALSFRS-R) is broadly used in clinical practice to evaluate disease progression in ALS patients [10]. However, disease heterogeneity challenges prognosis, pinpointing the need for advanced machine learning approaches to learning explainable disease progression models that clinicians can effectively use to aid prognostic prediction, increase survival, and promote patient care. These models should be able to learn from heterogeneous temporal data and benefit from patient stratification models whenever possible.

In this context, recent years have witnessed an increasing awareness of the potentialities of machine learning amongst ALS researchers, leading to several applications over ALS cohort data studies [7,12–15]. The great potential of learning stratification models has also shown opportunities for future clinical trials, besides promoting more accurate and trustable predictions by learning group-specific prognostic models [12,16,17].

Carreiro et al. [7] was a pioneer in proposing prognostic models to predict the need for NIV in ALS based on clinically defined time windows. More recently, Pires et al. [16] stratified patients according to their state of disease progression achieving three groups of progressors (slow, neutral and fast) and proposed specialized learning models according to these groups. They further used patient and clinical profiles with promising results [17]. Recently, Martins et al. [18] proposed to combine itemset mining with sequential pattern mining to unravel disease presentation and disease progression patterns and use these patterns to predict the need for NIV in ALS patients. Despite their relevant results, the success depends on proper data discretization. Also, it does not consider the contiguity constraint imposed by the temporality nature of the patient's follow-up data.

Matos et al. [19] proposed a biclustering-based classifier. Biclustering was used to find groups of patients with coherent values in subsets of clinical features (biclusters), then used as features together with static data. Besides promising, none of these approaches took into account the temporal dependence between the features. Soares et al. [20] proposed a triclustering-based classification approach to analyse ALS temporal data with promising results. Recently, Henriques and Madeira [6] assessed the impact of discriminative patterns with varying coherence and quality on associative classification.

In contrast with the large extent of contributions towards feature extraction from temporal data domains [8,21], the use of biclustering and triclustering stances to aid predictive tasks is recent and still scarce [3], mainly being represented by the studies mentioned above. In addition, and to our knowledge, no other study integrates biclustering and triclustering tasks for predictive ends.

3. Problem formulation

Definition 1 (*Three-way Dataset*)**.** A three-way dataset, D, is defined by N objects $X = \{x_1, \dots, x_N\}$, M features $Y = \{y_1, \dots, y_M\}$, and p contexts $Z = \{z_1, \dots, z_O\}$, where the elements a_{ijk} relate object x_i, feature y_j, and context z_k.

In the context of our work, we consider heterogeneous data is characterized by the presence of N individuals described by a set of static variables Y_{static}, associated with a two-way or multivariate dataset (where elements a_{ij} relate individual x_i and feature $y_j \in Y_{static}$); and temporal variables $Y_{temporal}$, associated with a three-way dataset (where elements a_{ijk} relate individual x_i, variable $y_j \in Y_{temporal}$, and time reference z_k).

Definition 2 (*Bicluster*)**.** Given a two-way data space, a bicluster $B = (I, J)$ is a subspace given by a subset of objects, $I \subseteq X$, and a subset of features, $J \subseteq Y$.

Definition 3 (*Tricluster*)**.** Given a three-way data space, a tricluster $\mathcal{T} = (I, J, Z)$, is a subspace defined by $I \subseteq X$ objects, $J \subseteq Y$ features and $K \subseteq Z$ contexts, where a_{ijk} denotes the elements of $\mathcal{T}$. A tricluster $\mathcal{T}$ can be alternatively represented as a set of $|K|$ biclusters, $\mathcal{T} = \{B_1, B_2, \dots, B_{|K|}\}$, where $B_i = (I, J, z_i)$,

$$B_1 = \begin{bmatrix} a_{i_1 j_1 z_1} & a_{i_1 j_2 z_1} & \cdots & a_{i_1 j_{|J|} z_1} \\ a_{i_2 j_1 z_1} & a_{i_2 j_2 z_1} & \cdots & a_{i_2 j_{|J|} z_1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_{|I|} j_1 z_1} & a_{i_{|I|} j_2 z_1} & \cdots & a_{i_{|I|} j_{|J|} z_1} \end{bmatrix}$$

$$B_2 = \begin{bmatrix} a_{i_1 j_1 z_2} & a_{i_1 j_2 z_2} & \cdots & a_{i_1 j_{|J|} z_2} \\ a_{i_2 j_1 z_2} & a_{i_2 j_2 z_2} & \cdots & a_{i_2 j_{|J|} z_2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_{|I|} j_1 z_2} & a_{i_{|I|} j_2 z_2} & \cdots & a_{i_{|I|} j_{|J|} z_2} \end{bmatrix}$$

$$\vdots$$

$$B_{|K|} = \begin{bmatrix} a_{i_1 j_1 z_{|K|}} & a_{i_1 j_2 z_{|K|}} & \cdots & a_{i_1 j_{|J|} z_{|K|}} \\ a_{i_2 j_1 z_{|K|}} & a_{i_1 j_2 z_{|K|}} & \cdots & a_{i_1 j_{|J|} z_{|K|}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_{|I|} j_1 z_{|K|}} & a_{i_{|I|} j_2 z_{|K|}} & \cdots & a_{i_{|I|} j_{|J|} z_{|K|}} \end{bmatrix}.$$

Definition 4 (*Subspace Clustering*)**.** Given a dataset, subspace clustering aims at identifying all relevant subspaces, whether given by biclusters in multivariate data or triclusters in three-way data, satisfying specific criteria of homogeneity and statistical significance.

Homogeneity criteria are commonly guaranteed through the use of a merit function, such as the variance of the values in the subspace or correlation across objects [2,3]. Merit functions are typically applied to guide the formation of biclusters and triclusters in greedy and exhaustive searches. In stochastic approaches, a set of parameters that describe the biclustering solution are learned by optimizing a merit (likelihood) function. The pursued homogeneity determines the coherence, quality and structure of a subspace clustering solution [22], where the *coherence* is determined by the observed form of correlation among its elements; the *quality* is defined by the type and amount of accommodated noise; and the *structure* described by the shape and positioning of subspaces. A flexible structure is characterized by an arbitrary number of (possibly overlapping) subspaces. *Statistical significance* criteria, in addition to homogeneity criteria, guarantees that the probability of a given subspace to occur (against a null model) deviates from expectations [3,23].

Given multivariate and three-way data, this work aims at assessing the quantitative impact produced by pattern-centric feature spaces, using biclustering and triclustering stances (Definition 4), to guide predictive tasks.

4. BicTric methodology

This section introduces BicTric, the proposed approach to learning predictive models from feature spaces produced under a mixture of biclustering and triclustering stances.

In this work, we propose a generic methodology able to work with any biclustering and triclustering algorithm. Fig. 1 depicts the BicTric workflow. To this end, consider the presence of heterogeneous data with static variables (two-way dataset) and temporal variables (three-way dataset). In the initial step, biclustering and triclustering are performed over the two-way and three-way data, respectively, to obtain biclusters and triclusters, seen as discriminative features.

In particular, when considering temporal data, we can either use each tricluster as a single feature or use its respective biclusters individually as multiple features. The latter case might be helpful depending on the heterogeneity of biclusters composing a tricluster.

After this initial step, additional principles are necessary to produce the target feature space, i.e. to map the original data space $Y_{static}, T_{temporal} \in \mathcal{Y}$ into the pattern-centric space $P_{bic}, P_{tric} \in \mathcal{P}$,

$$BicTric : \mathcal{Y} \rightarrow \mathcal{P}. \tag{1}$$

Sections 4.1 to 4.4 detail the necessary steps to yield this transformation for subsequent predictive modelling tasks.

4.1. Processing input data

Temporal observations from heterogeneous data may not be readily available within a three-way data structure. The number of time points could not be equal for all the instances in the original input dataset (e.g. patient follow-up screening). Some instances can have more records than others. To increase the performance prediction, we can transform the original data into a new data space where instances have the same number of time points before the instant target. However, this approach will lead us to discard a (possible) high number of individuals with fewer time points. In this context, we propose to group consecutive time points in which sets of these points will have a maximum length $\min(L, P)$, where L is a predefined set length and P is the available number of instance's time points.

This preprocessing step is independent of the nature of the data. It can be used with any type of three-way labelled data. In Fig. 2 we present a possible example of preprocessing in clinical data. In this case, we have patients with a different number of appointments, and the target is to predict if a patient will evolve (E) to a state that needs non-invasive ventilation support (E = Y) or not (E = N). To transform the original data into a dataset where each patient has a predefined size of 3 appointments ($L = 3$), we create sets of 3 consecutive appointments. The label of each set is now the corresponding label of its last appointment. Note that in this example, the patient P4 has only 2 ($P = 2$) available appointments, so will be created a set of size defined by $\min(3, 2) = 2$ and not to be removed.

The transformed dataset containing these sets of time-points for each individual will be the input for our proposed BicTric classifier.

4.2. Pattern discovery

Once the temporal observations are preprocessed, the resulting data is described by N instances and M variables, consisting of single-time measurements along M_{static} static variables, $Y_{static} \subset Y$, and a fixed number of periodic measurements along $M_{temporal}$ temporal variables, $Y_{static} \subset Y$, such that $Y_{static} \cap Y_{temporal} = \emptyset \wedge Y_{static} \cup Y_{temporal} = Y \wedge M = M_{static} + M_{temporal}$, forming a two-way (static) dataset and three-way (temporal) dataset, respectively. Biclustering and triclustering algorithms are then eligible to be applied over the produced two-way and three-way datasets, respectively, for the retrieval of statistically significant local associations.

4.2.1. Biclustering and triclustering searches

Given the end predictive goal, the target subspaces by both biclustering and triclustering tasks should show a correlation between instances, i.e. pattern coherence should be preserved across objects.

For the target biclustering task, this can be accomplished by seeking subspaces, $B = (I, J)$, whose elements, $a_{ij} \in (I, J)$, are described by

$$a_{ij} = c_j + \gamma_i + \eta_{ij},$$

where c_j is the expected value of variable $y_j \in Y_{static}$, γ_i is the adjustment for instance x_i, and η_{ij} is the noise factor of a_{ij}. The bicluster *pattern* φ_B is the set of expected values in the absence of adjustments and noise $\{c_j \mid y_j \in J\}$. This stance can be considered for pursuing biclusters with *constant coherence* (assuming $\gamma_i = 0$), and *additive coherence* ($\gamma_i \neq 0$).

Similar assumptions are extended towards the triclustering task. In this context, given a three-way dataset, the elements, a_{ijk}, of a subspace, $\mathcal{T} = (I, J, K)$, are described by

$$a_{ijk} = c_j + \gamma_i + \beta_k + \eta_{ijk},$$

where c_j is the expected value of variable $y_j \in Y_{temporal}$, γ_i is the adjustment for instance x_i, β_k is the temporal variation on time point z_k, and η_{ijk} is the noise factor of a_{ijk}.

The tricluster patterns $\varphi_{\mathcal{T}}$ are the set of expected values in the absence of adjustments and noise $\{c_j \mid y_j \in J\}_k$. A consensus across

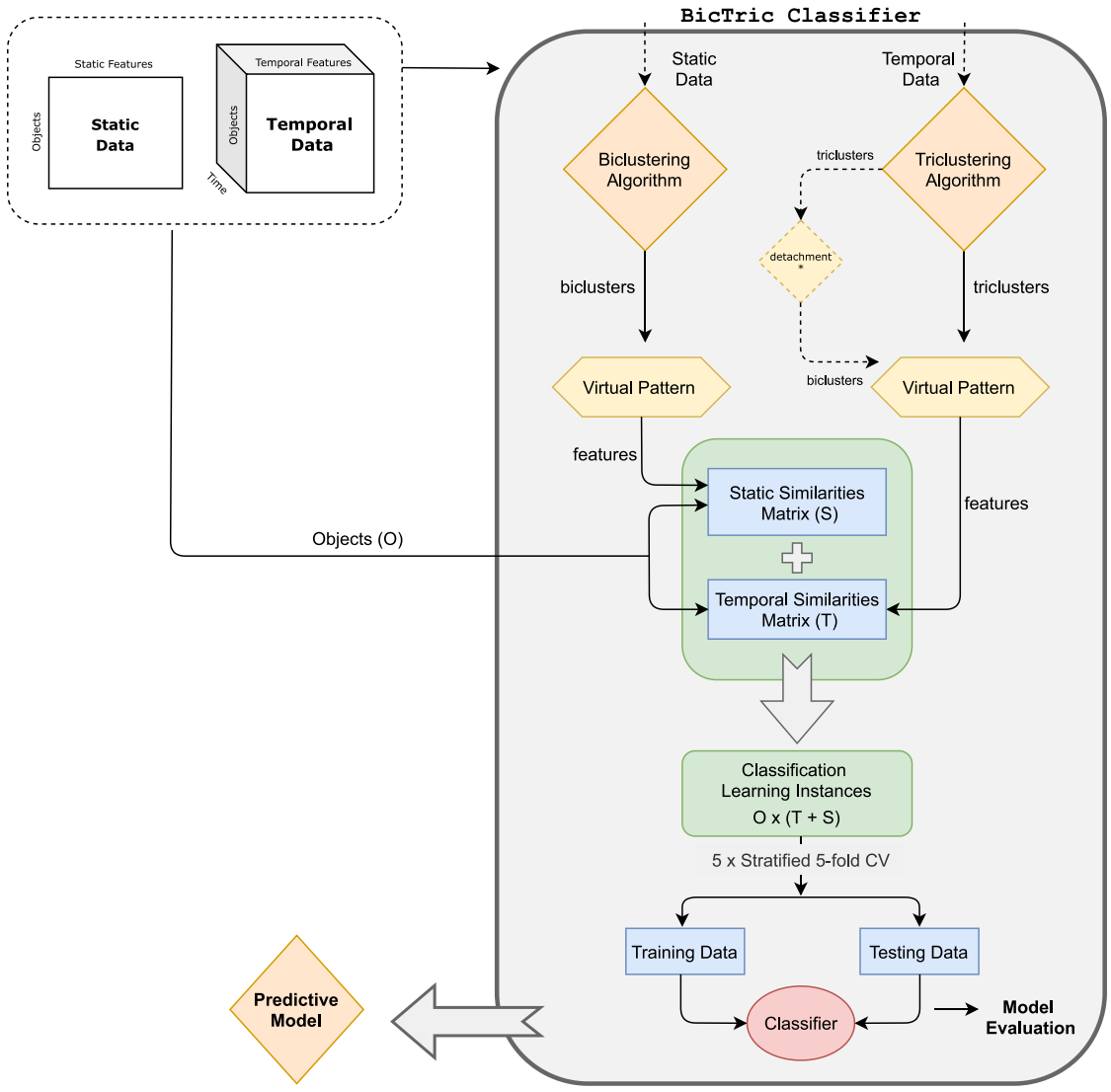

Fig. 1. BicTric Workflow: Learning predictive models using transformations onto pattern-centric feature spaces.

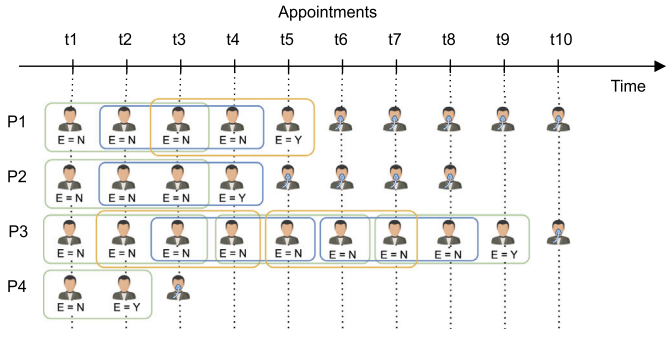

Fig. 2. Example of computing sets of snapshots with maximum length $\min(L, P)$, in this case, L = 3 and P is represented by the number of snapshots (where the patient was not using NIV) available for each patient. P4 has only 2 ($P = 2$) snapshots before NIV, and only one set with this 2 snapshots was considered.

time points can be considered in order to produce a single pattern from a tricluster, $\{\hat{c}_j \mid y_j \in J\}$.

Fig. 3 provides an illustrative view of additive bicluster and tricluster subspaces, along with their patterns.

In particular, BicTric is by default equipped with biclustering and triclustering searches able to find additive subspaces given their well-recognized relevance in clinical domains [3,6]. For this purpose, we consider the TRICLUSTER algorithm [24] to find both biclusters over static data and triclusters over temporal data. The biclustering search undertaken in TRICLUSTER is exhaustive and capable of accommodating additive factors on numerical variables. To this end, the Pearson coefficient is used to indicate the correlation strength between objects and a subset of overall variables.

In the context of the triclustering task, we instantiate the triclustering step considering TRICLUSTER [24], a pioneer, highly cited, and state-of-the-art triclustering approach proposed by Zhao and Zaki in 2005. It is a quasi-exhaustive approach, able to mine arbitrarily positioned and overlapping triclusters with constant, scaling, and shifting patterns from three-way data. Given that TRICLUSTER was proposed to mine coherent triclusters in three-way gene expression data (gene-sample-time), at this point, it is important to understand that clinical data can be preprocessed in order to yield a similar structure, in which gene-sample-time data becomes patient-feature-time data, for instance.

TRICLUSTER has 3 main steps: (1) construct a multigraph with similar value ranges between all pairs of samples; (2) mine maximal biclusters from the multigraph formed for each time point (slices of the 3D dataset); and (3) extract triclusters by merging similar biclusters from

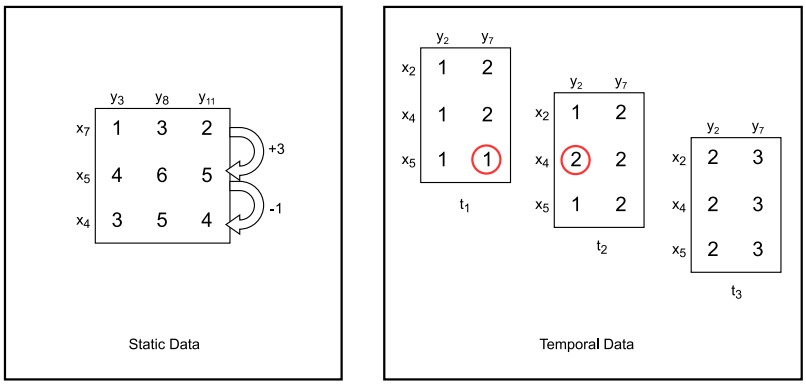

Fig. 3. Illustrative additive bicluster B and tricluster $\mathcal{T}$ from integer static and temporal data. Bicluster B has pattern $\varphi_B = \{c_3 = 1, c_8 = 3, c_{11} = 2\}$ and adjustment factors $\gamma = \{\gamma_4 = 2, \gamma_5 = 3, \gamma_7 = 0\}$. The tricluster pattern is defined by the contributions $c = \{c_2 = 1, c_7 = 2\}$ along the variable dimension, $\gamma = \{\gamma_2 = 0, \gamma_4 = 0, \gamma_5 = 0\}$ along the observation (patient) dimension, and $\beta = \{\beta_2 = 0, \beta_3 = 0, \beta_4 = 1\}$ along the time dimension. We highlighted noisy elements in red circles.

different time points. Optionally, it can delete or merge triclusters, according to the overlapping criteria used.

BicTric is applied with an extended version of the TRICLUSTER since the original version is unable to consider contiguity constraints along the time dimension. As our goal is to mine temporal three-way data, meaning the Z dimension (contexts) is time, we introduce temporal contiguity constraints similar to the stance taken in CCC-Biclustering [25], a state-of-the-art and highly efficient temporal biclustering algorithm. As a result, BicTric mines Time-Contiguous Triclusters, triclusters with consecutive time-points. The new TCTRICLUSTER algorithm considers this time constraint on its third phase. The use of contiguity constraints is associated with significant efficiency gains, leverages the interpretability of triclustering solutions, and potentially promotes predictive accuracy [3].

4.3. Virtual patterns

Once the patterns are found, we then assess how well each bicluster and tricluster describe a given object in order to produce the $a'_{ij} \in \mathbb{R}$ entries of the new pattern-centric space, $\mathcal{Y} \rightarrow \mathcal{P}$. To this end, we use the bicluster/tricluster most representative pattern and the object pattern. The object's (local) pattern corresponds to its values in the features in the bicluster or the features and contexts in the tricluster. To obtain the most representative pattern of a bicluster, we use the Virtual Pattern 2D introduced in [26]. To compute the tricluster's most representative pattern, we extended the Virtual Pattern 2D as follows:

Definition 5 (*Virtual Pattern 3D*)**.** Given a tricluster $\mathcal{T}$, its virtual pattern P is defined as a set of elements $P = \{\rho_1, \rho_2, \ldots, \rho_I\}$, where $\rho_i, 1 \leq i \leq I$, is defined as the mean (or the mode, in case of categorical features) of values in the ith row for each context:

$$\rho_i = \frac{1}{|J||Z|} \sum_{z=1}^{Z} \sum_{j=1}^{J} a_{ijz}. \tag{2}$$

4.4. Similarity matrices

Let S be the number of static virtual patterns given by the found biclusters, and T be the number of temporal patterns corresponding to the number of unique patterns found in the output triclusters (or corresponding biclusters).

In order to assess how similar each of the N objects (patients) is to the produced bicluster and tricluster features, we propose the calculus of the similarities between the virtual pattern P and the correspondent object pattern (same features and contexts). To this end, we consider two major approaches: (1) compute the Euclidean distance; or (2) compute Pearson correlation. The similarities are computed for temporal (resulted in a matrix with size $N \times T$) and static patterns (resulted in a matrix with size $N \times S$) and then merged into a single similarity matrix with size $N \times (S+T)$. This matrix will then be used to compute the learning instances by adding the class label to each feature vector.

4.5. BicTric in predictive tasks

Given a set of training examples, BicTric's is applied to learn and produce the target pattern-centric feature space autonomously. Any state-of-the-art classifier or regressor can be subsequently applied over this new data space to learn a predictive model.

During testing time, the learned mapping, $\mathcal{Y} \rightarrow \mathcal{P}$, is further considered to transform testing examples to the pattern-centric feature space. To this end, we first compute the similarities between the new object and the virtual patterns obtained in the learning phase. This array can then be passed to the trained classification or regression model, returning a class or quantity estimate for the new object as output.

Using the introduced principles, BicTric can be used to support the learning of predictive models from heterogeneous static and temporal data. Fig. 4 depicts how our proposed methodology can be used to classify a new object.

4.6. BicTric model evaluation

We proposed four well-established metrics to evaluate the performance of BicTric models: Area under the Receiver Operator Characteristic (ROC) curve (AUC), Accuracy, Sensitivity and Specificity [27].

The Receiver Operator Characteristic (ROC) curve assesses the true positive rate against the false-positive rate at various threshold values in binary classification settings. The Area Under the ROC Curve (AUC) offers a summary statistic of the ROC curve, ranging from 0 and 1 (correct decisions with high confidence) and measuring of the predictive ability of the classifier under different decision thresholds [27].

Accuracy is the ratio of the number of correct predictions to the total number of input samples [27]. Sensitivity identifies the portion of positive instances that were correctly classified, while specificity is the portion of the negative instances that were correctly classified. Illustrating, sensitivity (specificity) informs on the percentage of individuals with (without) a disease that were correctly diagnosed [27].

5. Case study

A large longitudinal cohort was led by the authors at Lisbon ALS Clinic (Centro Hospitalar Lisboa Norte) to study Amyotrophic Lateral Sclerosis (ALS). ALS is a neurodegenerative disorder associated with progressive muscular impairments without known cause or cure to

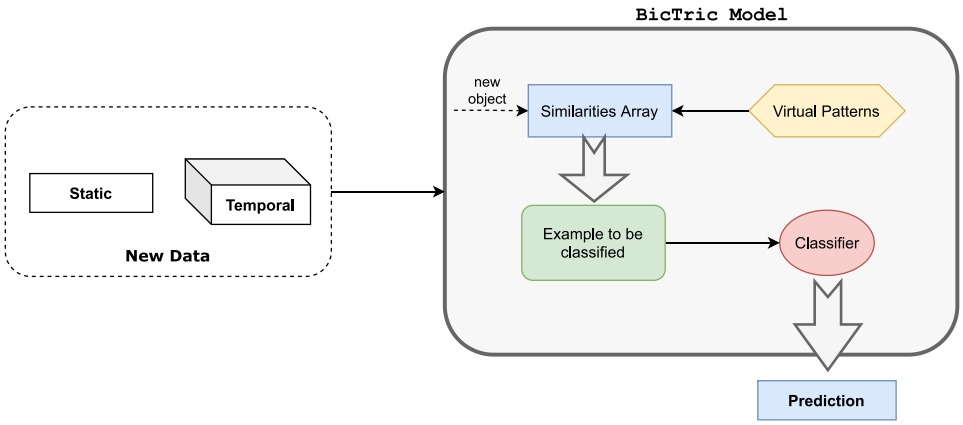

Fig. 4. BicTric Workflow at testing time: Using the BicTric model to classify new instances.

Table 1
Static and temporal features considered in the ALS case study. The following temporal features are concerning ALFRS-R scale [29]: ALSFRS (original score); ALSFRS-R (revised score); ALSFRSb (bulbar subscore); R (respiratory subscore); ALSFRSsUL (upper limb function); ALSFRSsLL (lower limb function); ALSFRSr (question 10 of the original ALSFRS score evaluating respiratory symptoms).

Static	Gender Body Mass Index (BMI) Age at onset Family History of Motor Neuron Disease Predominant Upper vs. Lower motor Neuron clinical phenotype Onset Form Diagnostic Delay El Escorial Revised Criteria Positive vs. Negative C9orf72 mutation
Temporal	ALSFRS ALSFRS-R ALSFRSb R ALSFRSsUL ALSFRSsLL ALSFRSr Forced Vital Capacity (FVC) Maximum Inspiratory Pressure (MIP) Maximum Expiratory Pressure (MEP)

date [28]. In particular, we aim at predicting the need for non-invasive ventilation (NIV) within 90 days since the last appointment. The 90 days frame is clinically defined based on its actionability towards care monitoring protocols.

This section describes the target data cohort used in this study and the specific preprocessing steps needed to adequately address disease progression problem in ALS patients with BicTric methodology.

5.1. Data

The Lisbon ALS cohort was considered, containing Electronic Health Records from ALS Patients regularly followed at the local ALS clinic since 1995 and last updated in May 2021. Its current version includes a total of 1485 patients, with 998 patients considered eligible as they were consistently assessed using the static and temporal features shown in Table 1 and according to the criteria defined in [7] to compute the Evolution class. Each patient has a set of static features (demographic and clinical data, and genetic information) along with temporal features (measured at the follow-up every 3 months), namely clinical markers of disease progression, as motor function (evaluated by ALSFRS-R) and respiratory function (evaluated by FVC, MIP and MEP).

5.2. Data preprocessing

Data were preprocessed as described by Carreiro et al. [7], and Pires et al. [16] to obtain patients snapshots. Patients are followed up regularly and perform a couple of prescribed tests after each appointment. Since a patient may not be able to perform all the tests in a single day, we have to take the temporal distribution into account when aligning records. In this context, we computed snapshots of the patient's condition with an acceptable time length to comprise all the prescribed tests. The patient's snapshots are sets of assessments obtained by a hierarchical (agglomerative) clustering with some constraints: (1) the evaluations that compose a snapshot cannot belong to the same test, once doctors never prescribe the same test twice; and (2) all the evaluations considered in the same snapshot should be consistent according to the NIV state (the patient should have the same NIV condition in all the assessments). The defined cutting point to form the snapshots was 100 days given the typical periodicity of the assessments as explained in [7], which means that the time windows between snapshots are never higher than 100 days for the monitored tests in this cohort study.

Next, we computed the evolution (E) class for each snapshot according to the NIV administration date. If a patient received NIV within 90 days of the snapshot's date, the learning example is labelled as `Y` or is labelled as `N`, otherwise.

Then, we processed data as explained in Section 4.1 to transform the original dataset into three-way data as input data to BicTric. In addition, we created 3 datasets having examples composed of 3, 4 and 5 consecutive snapshots (CS). The number of initial examples is documented in the 'TOTAL' column of the Table 2 discriminated by classes.

Class imbalance. Due to the short time window considered (90 days) in our case study, the number of non-evolutions examples (class `N`) is far superior to that of evolutions (class `Y` - patients requiring NIV within 90 days), which are key for the learning task. To minimize the impact of this class imbalance, and grounded on empirical evidence against alternatives, we performed Random Undersample (RU) to reduce the majority class examples, obtaining a proportion of 2/3 within the dataset, followed by SMOTE [30], to oversample the minority class examples, minimizing class imbalance. Table 2 shows class distribution after these steps.

6. Results

BicTric was applied with (1) the proposed TCTriCluster approach extended from Zhao and Zaki [24]; (2) the underlying biclustering algorithm used within this triclustering search [24], (3) using the similarity matrices as learning examples in an appropriate classifier. In our experiments, we evaluated four models using Naive Bayes,

Table 2
Class distribution.

	Total		RU		SMOTE	
	N	Y	N	Y	N	Y
3 CS	2019	510	1020	510	1020	1020
4 CS	1669	501	1002	501	1002	1002
5 CS	1401	497	994	497	994	994

Table 3
Baseline 1 results: IPM and SPM based approach by Martins et al. [18].

Clf	Sensitivity	Specificity	AUC	Accuracy
NB	57.7 ± 7.6	60.6 ± 6.8	63.6 ± 4.4	59.0 ± 4.1
SVM	72.3 ± 4.8	65.9 ± 8.7	76.5 ± 4.2	70.5 ± 3.4
XGB	71.0 ± 3.9	54.6 ± 4.9	68.7 ± 3.5	63.8 ± 3.5
RF	72.4 ± 3.8	53.6 ± 6.0	69.5 ± 3.6	64.2 ± 3.3

SVM with Gaussian kernel, XGBoost, and Random Forests due to their state-of-the-art performance for the target predictive task [17,18].

In what follows, we present BicTric's results obtained following the main workflow proposed (Section 6.1) and additional specialized models for stratified patient populations (Section 6.2). We explained some alternative decisions made within BicTric methodology. We compared the results with a baseline obtained by a state-of-the-art pattern-based classifier which is, at the best of our knowledge, a pioneering approach in using discovered patterns as features in predictive tasks. The results from the baseline, published in [18], were obtained from a previous version of the dataset in the study. We reproduced their method with our updated version dataset and used it to evaluate the proposed approach's effectiveness.

6.1. Predictive models without patient stratification

Baselines. Martins et al. [18] proposed a method using the same ALS data to learn predictive models to predict the same target (need for NIV) using an approach that starts by discovering patterns in static and temporal data using Itemset mining in static features and sequential pattern mining in temporal ones. These patterns were then fed to the classification task as features. They used a simple binary approach to compare the learning instances with the patterns and Euclidean distance. The latter showed to be the best approach. In this context, and since it is the closest approach to BicTric found in the literature, we decided to use their results as a baseline (Baseline 1) for our evaluation experiments. Table 3 depicts the results.

Furthermore, we consider an additional static baseline classifier reliant on the static features and temporal data only from the last snapshots from each set of consecutive snapshots. We assess the behaviour of different classifiers learned over these features as our prognostic predictors (Baseline 2). Table 4 depicts the results.

BicTric. We evaluated the approach using 4 different classifiers, as explained before. In the model evaluation step, we used 5-fold cross-validation, balancing the training folds with the same approach used before triclustering step: random undersample of majority class until obtain the proportion of 2/3 followed by SMOTE of minority class achieving equal number of both classes examples. We also ensured that the examples belonging to a specific patient were included in the same fold within cross-validation. Table A.10 depicts BicTric's results.

6.2. Specialized predictive models using patient stratification

Using the same data, we stratified the patients in three groups according to their disease progression profile, following the approach used by Pires et al. [16]. The patients were stratified in Slow, Neutral and Fast progressors according to a Progression Rate (PR) value, computed by:

$$PR = \frac{48 - \text{ALSFRS-R}_{\text{1st Visit}}}{\Delta t_{\text{1st Symptoms; 1st Visit}}}, \quad (3)$$

where 48 is the maximum score for ALSFRS-R feature, $\text{ALSFRS-R}_{\text{1st Visit}}$ is the ALSFRS-R score in the first appointment (diagnosis) and $\Delta t_{\text{1st Symptoms; 1st Visit}}$ is the time in months between the dates of first symptoms and the first appointment [16]. Progression rates were computed for each patient. Considering these values, patients were divided into three groups based on the distribution of PR values as suggested by clinicians: 25% of patients with lower and higher values are stratified as Slow and Fast progressors, respectively. The remaining 50% were considered Neutral progressors.

After stratifying the patients into three groups of progressors, we computed the sets of snapshots for each group as in previous settings. Table 6 shows class distribution after patient stratification. Note that some patients considered in the general model (with all patients) do not have all the information needed to compute the progression rate, so we dismiss their snapshots at this phase.

Again, we considered the same two baselines as was done for the dataset with all patients. We applied the IPM and SPM based approach [18] and used static features together with only the last

Table 4
Baseline 2 results: Models learnt using data from the last point of each set of consecutive snapshots.

	Clf	Sensitivity	Specificity	AUC	Accuracy
3 CS	NB	43.2 ± 4.8	81.4 ± 2.7	70.0 ± 3.1	62.3 ± 2.6
	SVM	43.0 ± 4.9	81.4 ± 2.7	73.7 ± 3.1	62.2 ± 2.6
	XGB	72.9 ± 3.3	72.1 ± 3.4	80.4 ± 2.1	72.5 ± 2.3
	RF	73.5 ± 3.3	72.7 ± 3.9	80.9 ± 1.8	73.1 ± 2.5
4 CS	NB	54.2 ± 4.8	77.3 ± 4.3	73.4 ± 3.3	65.8 ± 3.3
	SVM	55.2 ± 4.5	75.7 ± 4.3	71.4 ± 2.9	65.5 ± 3.3
	XGB	73.5 ± 3.8	71.1 ± 3.6	80.2 ± 2.5	72.3 ± 2.8
	RF	74.3 ± 3.0	71.3 ± 3.6	80.4 ± 2.3	72.8 ± 2.4
5 CS	NB	60.6 ± 4.2	72.7 ± 3.6	73.0 ± 3.1	66.7 ± 3.1
	SVM	62.5 ± 4.3	69.2 ± 3.8	70.6 ± 3.5	65.8 ± 3.3
	XGB	75.3 ± 3.6	69.7 ± 3.9	80.1 ± 2.4	72.5 ± 2.7
	RF	75.0 ± 3.8	68.7 ± 4.0	80.3 ± 2.3	71.9 ± 2.8

Table 5
BicTric results: Summary of the best obtained results. Detailed results are documented in Tables A.10, A.12, A.13, A.14.

	Approach	Clf	Sensitivity	Specificity	AUC	Accuracy
All	5/D/D	SVM	76.3 ± 2.9	74.7 ± 3.0	83.3 ± 1.9	75.2 ± 2.4
Slow	3/D/D	RF	61.1 ± 11.8	81.7 ± 3.5	82.2 ± 4.1	80.0 ± 3.0
Neutral	4/D/C	SVM	67.9 ± 6.1	73.4 ± 4.5	78.6 ± 3.7	71.6 ± 3.8
Fast	3/D/C	SVM	80.2 ± 7.1	72.2 ± 6.2	82.4 ± 5.4	75.3 ± 4.9

Table 6
Distribution of classes with patient stratification.

		TOTAL		RU		SMOTE	
		N	Y	N	Y	N	Y
3CS	Slow	1006	91	182	91	182	182
	Neutral	730	278	556	278	556	556
	Fast	165	101	165	101	165	165
4CS	Slow	864	91	182	91	182	182
	Neutral	570	271	542	271	542	542
	Fast	135	99	135	99	135	135
5CS	Slow	742	90	180	90	180	180
	Neutral	455	269	455	269	455	455
	Fast	116	98	116	98	116	116

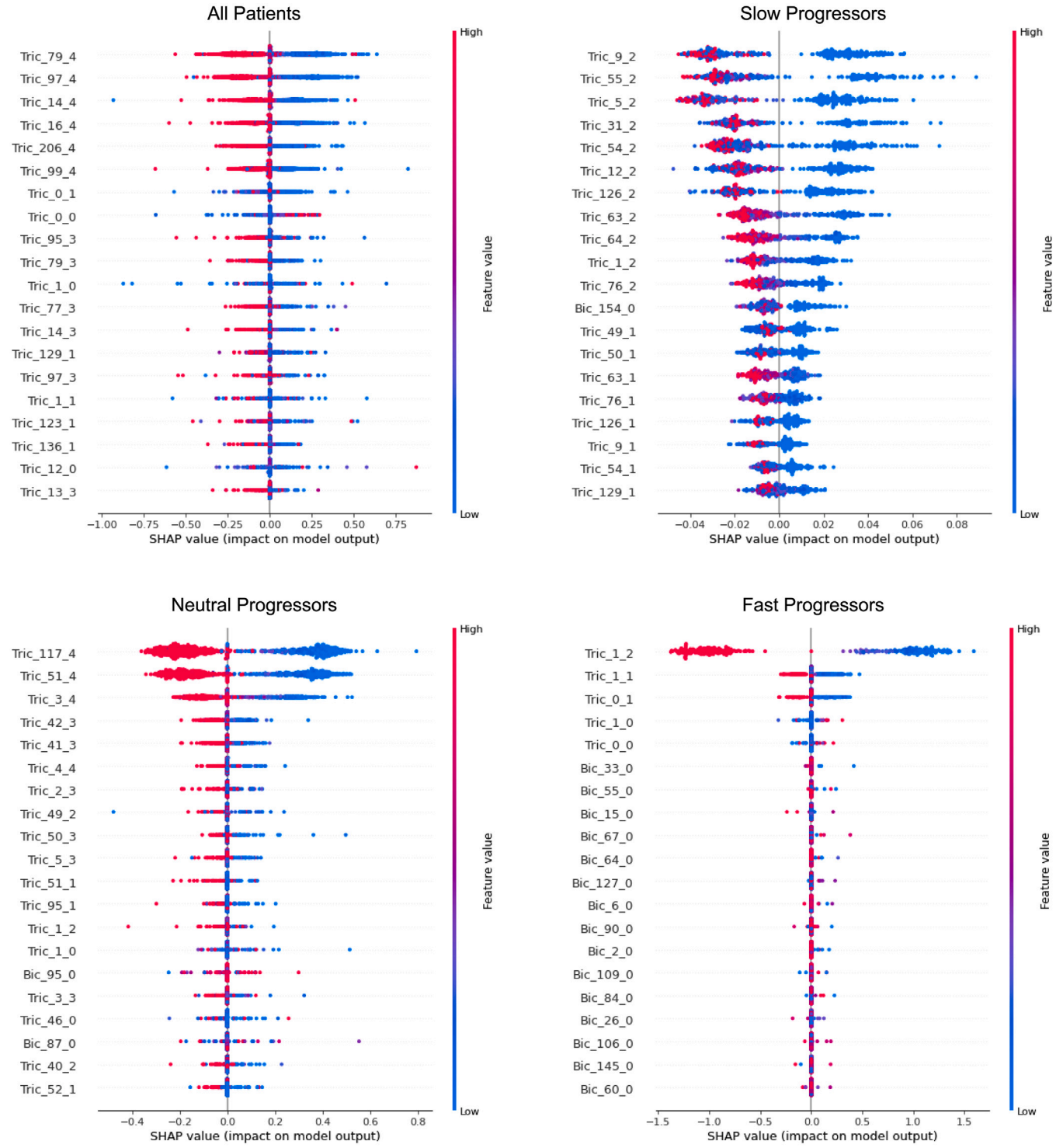

Fig. 5. Impact on the model of the top 20 most relevant patterns used by the classifiers. The terminology used is the following: patterns (documented in Table 9) obtained from triclusters (temporal data) are identified as 'Tric', and those obtained from biclusters (static data) are identified as 'Bic', followed by an identification number and finally, the snapshot position in the set of snapshots, in which 0 is the first position. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

temporal point of each set of consecutive snapshots. Tables 8 and A.11 provide the achieved results for each baseline, respectively.

As was done for the dataset with all the patients, we performed different experiments with the BicTric approach, using Euclidean and/or Correlation matrices as input to the classifiers: Naive Bayes, SVM with Gaussian kernel, XGBoost and Random Forests. Table 5 presents the summary for the best results obtained by these specialized models for slow, neutral and fast progressors. The (Tables A.12, A.13 and A.14) with all the detailed results are documented in Appendix.

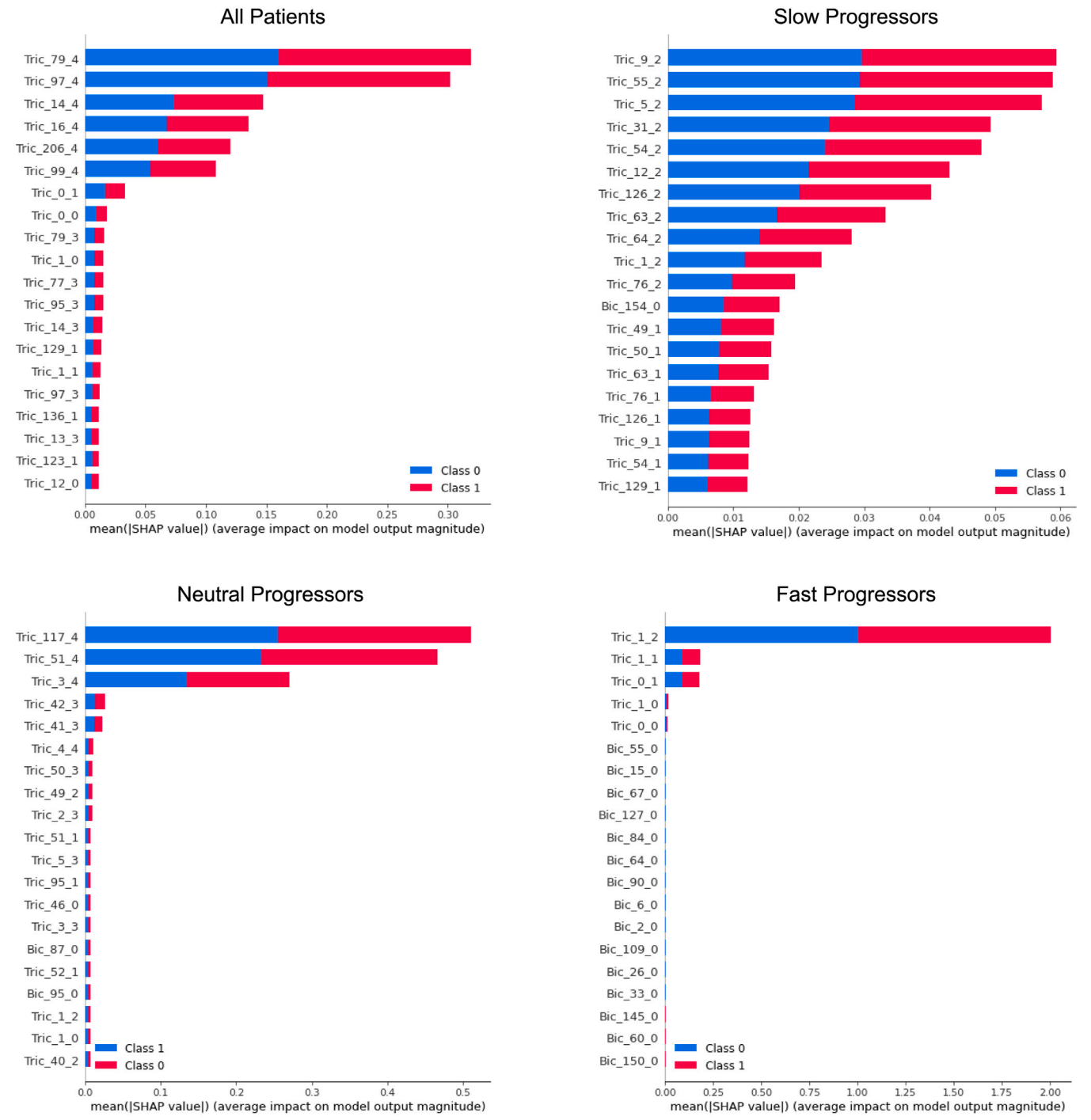

Fig. 6. Top 20 most important patterns used by the classifiers. The terminology used is the following: patterns (documented in Table 9) obtained from triclusters (temporal data) are identified as 'Tric', and those obtained from biclusters (static data) are identified as 'Bic', followed by an identification number and finally, the snapshot position in the set of snapshots, in which 0 is the first position. Class 0 represents 'non-evolutions' and Class 1 represents 'evolutions'.

7. Discussion

7.1. General models (without patient stratification)

When comparing the gathered results in Tables 3, 4 and A.10, we observe that BicTric approach shows significant improvements in predictive performance against all baselines, including the competitive baseline approach followed in [18] for this case study. The SVM classifier yields the best results in our case study. When applied under the BicTric approach, it yields an approximate 83 AUC using an embed Euclidean distance matrix obtained from 5 CS data. The baselines results obtained (with the same classifier) show 76.5 and 70.6 for AUC in baselines 1 and 2, respectively. For baseline 2, SVM was not the best classifier, although the best result (RF) did not surpass BicTric.

Moreover, BicTric (with SVM classifier) achieved Sensitivity higher than 75%, meaning that this classifier correctly predicts a positive evolution for NIV administration in more than 75% of the positive situations.

On the other hand, we observe that Specificity, the true negative rate, is generally higher than 75%, being good in predicting the negative class, in this case, non-evolution to NIV. We also note that this

Table 7
Results obtained with BicTric learnt from the whole dataset per disease progression group.

Group	Sensitivity	Specificity	Accuracy
Slow	51.0 ± 10.2	87.8 ± 2.9	83.8 ± 3.3
Neutral	76.4 ± 4.3	65.7 ± 5.6	69.7 ± 3.6
Fast	92.8 ± 4.7	47.3 ± 7.0	65.7 ± 5.8

Table 8
Baseline results for each ALS progression group.

Clf	Sensitivity	Specificity	AUC	Accuracy
Slow progressors				
NB	59.2 ± 11.3	66.3 ± 10.4	66.8 ± 7.4	63.4 ± 5.4
SVM	70.2 ± 11.1	69.2 ± 8.2	75.9 ± 8.1	69.6 ± 4.4
RF	69.1 ± 8.8	75.7 ± 8.1	80.9 ± 5.4	73.0 ± 4.8
XGB	70.1 ± 7.7	78.1 ± 9.1	81.2 ± 6.9	74.8 ± 6.3
Neutral progressors				
NB	39.4 ± 8.9	78.9 ± 6.8	64.8 ± 5.5	54.1 ± 4.8
SVM	64.8 ± 7.8	61.6 ± 6.7	70.7 ± 4.9	63.6 ± 5.3
RF	79.0 ± 6.2	38.1 ± 6.4	65.7 ± 4.4	63.8 ± 3.8
XGB	73.8 ± 5.4	46.7 ± 7.4	66.3 ± 4.8	63.7 ± 4.3
Fast progressors				
NB	36.3 ± 18.0	75.3 ± 11.9	59.0 ± 11.0	65.3 ± 8.0
SVM	58.7 ± 9.1	66.0 ± 12.4	66.8 ± 7.6	61.4 ± 7.1
RF	78.0 ± 9.5	33.6 ± 11.0	60.2 ± 9.5	61.3 ± 7.3
XGB	71.6 ± 10.1	40.6 ± 12.8	59.2 ± 8.5	59.9 ± 5.5

is an outstanding achievement since the superior result in predicting respiratory insufficiency in patients with ALS. Moreover, although SVM achieved the best results, we can note that other classifiers also achieved high values in Specificity, being harmed by the class imbalance of the original dataset, having test sets in cross-validation with a low number of positive examples.

It is important to note that the high patient heterogeneity, together with the difficulty of learning models without patient stratification as reported in [16], makes these results even more promising, since BicTric's application in specific subgroups to learn specialized models according to disease progression groups, has the potential to improve our prognostic results.

The common use of biclustering and triclustering to compute pattern-centric features in the proposed BicTric methodology yielded relevant insights and promotes the model's interpretability, as shown below.

7.2. Specialized models (stratified patients)

Unexpectedly, the results for these specialized models did not outperform the results obtained when learning from all patients. This led us to think that the general model was biased by the examples belonging to the neutral group patients, which was good in predicting the need for NIV in these neutral patients but not so good for slow and fast progressors. We recovered the general model and tested it with the data from slow, neutral and fast progressors to verify this detail. The results in Table 7 confirm this hypothesis as evidenced by the discrepancy in the Sensitivity and Specificity observed for the Slow and Fast groups.

Looking in detail at the results obtained by the specialized BicTric models, we can verify that the SVM classifier achieved using the dataset with examples composed of 3 consecutive snapshots (3/D/C) for slow progressors 71.2% and 73.7% in Sensitivity and Specificity, respectively. Despite these results being less discrepant in these two metrics, all the other classifiers surpassed 80% in Specificity. The high values of standard deviation in Sensitivity reveal the difficulty of predicting the positive class. Values for Specificity were generally high with low standard deviations meaning that it is easier to predict non-evolutions. Comparing these results with the baseline, in Table 8, we can see that BicTric was better in Sensitivity and largely overpassed the baseline in Specificity.

Table 9
The most relevant patterns used by the different learnt BicTric models: All Patients, Slow, Neutral and Fast progressors. The terminology used is the following: patterns obtained from triclusters (temporal data) are identified as 'Tric', and those obtained from biclusters (static data) are identified as 'Bic', followed by an identification number and finally, the snapshot position in the set of snapshots, in which 0 is the first position.

All patients	
Tric_79_4	[ALS-FRSb = 12, ALS-FRSr = 4, R = 12]
Tric_97_4	[ALS-FRSb = 12, R = 12]
Tric_14_4	[ALS-FRSb = 12, ALS-FRSr = 4]
Tric_16_4	[ALS-FRSb = 12]
Tric_206_4	[ALS-FRSr = 4, R = 12]
Tric_99_4	[R = 12]
Tric_0_1	[ALS-FRS = 37, ALS-FRS-R = 45]
Tric_0_0	[ALS-FRS = 38, ALS-FRS-R = 46]
Slow progressors	
Tric_9_2	[R = 12]
Tric_55_2	[ALS-FRSb = 12, R = 12]
Tric_5_2	[ALS-FRSr = 4]
Tric_31_2	[ALS-FRSb = 12, ALS-FRSr = 4]
Tric_54_2	[ALS-FRSb = 12, ALS-FRSr = 4, R = 12]
Tric_12_2	[ALS-FRSb = 12]
Tric_126_2	[ALS-FRSr = 4, R = 12]
Tric_63_2	[ALS-FRSsUL = 12, ALS-FRSr = 4]
Neutral progressors	
Tric_117_4	[ALS-FRSr = 4, R = 12]
Tric_51_4	[R = 12]
Tric_3_4	[ALS-FRSb = 12, ALS-FRSr = 4]
Tric_42_3	[ALS-FRSb = 12, R = 12]
Tric_41_3	[ALS-FRSb = 12, ALS-FRSr = 4, R = 12]
Fast progressors	
Tric_1_2	[ALS-FRSr = 4, R = 12]
Tric_1_1	[ALS-FRSr = 4, R = 12]
Tric_0_1	[ALS-FRSr = 4, R = 12]

The results for neutral progressors were more similar to the general model. The best model was obtained using the dataset with five consecutive snapshots (4/D/C). The same happened for the fast progressors, needing five consecutive snapshots to achieve high results in Sensitivity. As happened for the slow progressors, BicTric exceeded the baseline results in Neutral and Fast groups, highlighting the superior result for the fast progressors (82.4 AUC) with an SVM classifier.

In general, we observed contrasts between Sensitivity and Specificity caused by the different patterns unveiled by the biclustering/triclustering phase. The obtained patterns were more discriminative (and more in number) for one of the classes, making the classifier disregard the evidence for the less discriminating class. Despite this fact, the results show Sensitivity levels above 75%, making these results very promising in ALS prognostic for respiratory insufficiency.

7.3. Model interpretability

Complementarily to predictive accuracy, we are interested in assessing the interpretability of the pattern-centric feature space produced by BicTric by studying the top patterns highlighted by the classifiers. To this aim, we chose to analyse the top discriminative patterns discovered when applying BicTric, by the predictive models (identified previously) yielding the best results. The goal is to identify the most relevant features, what features appear together, and whether the more relevant temporal patterns (considering putative patterns from the average patient) can help physicians in their prognostics.

To perform the model explainability and analyse the most important patterns used by the model, we applied the unified SHAP

Table A.10
BicTric results for the NIV predictive task. CS—Number of considered consecutive snapshots/similarity criteria used in matrix T/similarity criteria used in matrix S; C—Pearson correlation, D—Euclidean distance.

CS	Clf	Sensitivity	Specificity	AUC	Accuracy
3/C/C	NB	50.2 ± 4.9	79.0 ± 3.0	71.9 ± 2.2	73.2 ± 2.0
	SVM	65.2 ± 4.9	78.7 ± 3.6	79.2 ± 1.9	76.0 ± 2.4
	XGB	59.0 ± 5.8	80.9 ± 1.9	78.7 ± 2.1	76.5 ± 1.5
	RF	59.1 ± 4.6	82.2 ± 2.0	79.9 ± 1.9	77.5 ± 1.4
3/C/D	NB	50.8 ± 4.3	79.7 ± 1.9	72.2 ± 2.1	73.9 ± 1.6
	SVM	63.8 ± 3.5	75.2 ± 2.5	77.0 ± 1.9	72.9 ± 2.0
	XGB	60.0 ± 4.9	81.2 ± 1.8	79.1 ± 2.1	76.9 ± 1.6
	RF	59.0 ± 5.0	82.6 ± 2.0	80.0 ± 1.8	77.9 ± 1.4
3/D/C	NB	56.5 ± 4.2	75.8 ± 2.7	73.4 ± 2.1	71.9 ± 2.1
	SVM	72.0 ± 4.0	77.3 ± 1.7	82.5 ± 1.6	76.2 ± 1.5
	XGB	58.0 ± 4.1	82.1 ± 1.8	79.4 ± 1.6	77.2 ± 1.3
	RF	60.1 ± 4.9	82.7 ± 1.8	80.6 ± 1.9	78.2 ± 1.4
3/D/D	NB	56.6 ± 4.2	77.6 ± 1.7	74.1 ± 2.0	73.3 ± 1.5
	SVM	72.5 ± 3.8	76.7 ± 1.9	81.8 ± 1.7	75.9 ± 1.6
	XGB	59.1 ± 5.3	82.2 ± 1.5	79.6 ± 2.0	77.5 ± 1.3
	RF	59.1 ± 4.7	83.0 ± 1.9	80.9 ± 1.7	78.2 ± 1.4
4/C/C	NB	57.5 ± 6.0	78.4 ± 2.5	74.6 ± 2.4	73.6 ± 1.9
	SVM	69.5 ± 4.3	79.5 ± 2.6	81.0 ± 2.6	77.2 ± 2.1
	XGB	58.3 ± 4.9	82.2 ± 2.0	80.3 ± 1.7	76.7 ± 1.8
	RF	60.0 ± 4.4	83.3 ± 1.8	81.3 ± 1.9	77.9 ± 1.7
4/C/D	NB	57.1 ± 4.6	79.0 ± 2.6	75.2 ± 2.2	73.9 ± 1.8
	SVM	71.0 ± 4.1	75.1 ± 2.5	79.6 ± 2.2	74.2 ± 2.1
	XGB	57.7 ± 4.0	82.2 ± 1.9	80.0 ± 1.8	76.5 ± 1.7
	RF	59.4 ± 3.7	83.7 ± 2.0	81.3 ± 2.0	78.1 ± 1.8
4/D/C	NB	63.2 ± 4.8	76.3 ± 2.4	75.6 ± 2.2	73.3 ± 2.0
	SVM	69.0 ± 4.6	79.9 ± 2.2	83.3 ± 2.4	77.4 ± 1.8
	XGB	58.3 ± 4.1	81.9 ± 1.8	80.1 ± 2.2	76.5 ± 1.6
	RF	59.9 ± 4.2	83.5 ± 1.8	81.7 ± 2.2	78.1 ± 1.7
4/D/D	NB	61.8 ± 3.9	77.2 ± 2.5	76.5 ± 2.1	73.6 ± 2.1
	SVM	69.6 ± 4.1	78.8 ± 2.6	83.1 ± 2.4	76.7 ± 2.1
	XGB	58.8 ± 4.6	82.5 ± 1.8	80.3 ± 2.0	77.0 ± 1.5
	RF	60.0 ± 4.0	83.8 ± 1.8	81.7 ± 2.1	78.3 ± 1.6
5/C/C	NB	61.2 ± 4.1	78.3 ± 2.1	75.9 ± 2.2	73.8 ± 2.0
	SVM	75.7 ± 3.1	75.8 ± 2.8	81.3 ± 2.3	75.8 ± 2.2
	XGB	62.7 ± 4.5	81.1 ± 2.7	81.1 ± 1.9	76.3 ± 2.2
	RF	63.5 ± 4.2	82.2 ± 1.8	82.0 ± 1.5	77.3 ± 1.5
5/C/D	NB	61.0 ± 4.8	78.4 ± 2.6	76.9 ± 2.2	73.8 ± 2.2
	SVM	76.7 ± 3.4	72.8 ± 2.6	80.2 ± 2.0	73.8 ± 2.1
	XGB	62.1 ± 4.4	81.1 ± 2.0	81.1 ± 1.8	76.1 ± 1.7
	RF	62.9 ± 3.8	82.7 ± 2.5	82.0 ± 1.5	77.5 ± 1.8
5/D/C	NB	65.1 ± 3.8	76.6 ± 2.5	76.7 ± 2.1	73.6 ± 2.2
	SVM	76.1 ± 2.9	75.3 ± 2.6	83.2 ± 1.9	75.5 ± 2.3
	XGB	62.5 ± 4.6	81.3 ± 2.1	81.1 ± 1.1	76.4 ± 1.7
	RF	62.4 ± 4.0	83.2 ± 1.9	82.4 ± 1.3	77.8 ± 1.3
5/D/D	NB	65.2 ± 3.6	77.3 ± 2.7	77.5 ± 2.1	74.1 ± 2.3
	SVM	76.3 ± 2.9	74.7 ± 3.0	**83.3 ± 1.9**	75.2 ± 2.4
	XGB	60.9 ± 2.8	81.9 ± 2.4	81.3 ± 1.2	76.4 ± 1.8
	RF	62.3 ± 4.2	83.3 ± 1.8	82.5 ± 1.2	77.8 ± 1.3

approach [31]. In particular, we selected the KernelExplainer and TreeExplainer methods, which introduce the possibility of directly measuring local feature interaction effects [32].

Fig. 6 and Table 9 show the classifier's most relevant patterns for each group of patients, ranked by their feature importance output by SHAP, with the top ones showing more predictive power than those at the bottom. An overall analysis of the most important patterns discovered is a good descriptor of patients who are not significantly affected by the disease since the values of the features are generally high (a maximum score means the functions are not affected). We noted that these patterns identify clearly the patients who will not need NIV within 90 days (non-evolutions) and are thus highly discriminative. These patterns are also relevant for further clinical analysis since these features can be essential to identify disease progression patterns leading to the need for NIV for a given time window.

Table A.11
Baseline 1 results for each ALS disease progression group: Models learnt using data from the last point of each set of consecutive snapshots.

	Clf	Sensitivity	Specificity	AUC	Accuracy
Slow progressors					
3 CS	NB	21.7 ± 11.4	89.5 ± 2.9	66.7 ± 7.7	55.6 ± 6.0
	SVM	21.2 ± 11.5	89.4 ± 2.7	71.3 ± 7.4	55.3 ± 6.1
	XGB	73.1 ± 8.6	76.4 ± 9.3	81.5 ± 4.2	74.7 ± 5.3
	RF	74.4 ± 8.2	75.3 ± 9.3	82.3 ± 4.4	74.8 ± 6.0
4 CS	NB	37.1 ± 11.7	79.4 ± 6.8	69.1 ± 6.5	58.2 ± 6.1
	SVM	29.5 ± 11.2	86.6 ± 5.4	75.3 ± 6.6	58.0 ± 6.0
	XGB	70.5 ± 8.4	69.6 ± 10.4	77.9 ± 6.1	70.0 ± 6.5
	RF	73.1 ± 7.7	69.6 ± 10.9	80.4 ± 5.3	71.3 ± 6.9
5 CS	NB	46.6 ± 11.2	74.2 ± 8.5	69.7 ± 7.3	60.4 ± 7.0
	SVM	39.9 ± 11.3	79.9 ± 7.6	69.9 ± 7.5	59.9 ± 6.9
	XGB	71.3 ± 9.4	68.0 ± 8.4	77.0 ± 6.3	69.7 ± 6.1
	RF	75.0 ± 8.8	69.1 ± 10.6	79.8 ± 6.1	72.1 ± 7.1
Neutral progressors					
3 CS	NB	52.2 ± 7.9	76.5 ± 3.9	64.7 ± 5.2	59.9 ± 4.5
	SVM	51.7 ± 7.9	77.2 ± 3.7	66.8 ± 4.6	60.0 ± 4.5
	XGB	80.1 ± 4.4	69.8 ± 4.6	77.3 ± 2.7	70.5 ± 2.5
	RF	80.5 ± 5.4	69.7 ± 5.4	78.1 ± 3.1	70.6 ± 3.4
4 CS	NB	64.0 ± 5.4	70.8 ± 5.6	67.4 ± 3.8	62.9 ± 3.8
	SVM	63.9 ± 5.3	70.3 ± 5.4	65.1 ± 4.7	62.6 ± 3.7
	XGB	79.1 ± 4.7	67.8 ± 5.0	77.1 ± 3.1	69.0 ± 3.6
	RF	80.7 ± 4.6	65.8 ± 5.6	77.5 ± 3.1	68.7 ± 3.6
5 CS	NB	71.4 ± 6.7	64.4 ± 6.6	67.2 ± 5.3	63.4 ± 4.8
	SVM	73.7 ± 6.0	61.2 ± 6.3	64.8 ± 5.3	63.0 ± 4.5
	XGB	80.7 ± 4.8	64.1 ± 6.8	75.0 ± 4.3	67.9 ± 4.0
	RF	80.1 ± 5.8	63.8 ± 6.8	75.1 ± 4.1	67.5 ± 4.2
Fast progressors					
3 CS	NB	65.0 ± 11.2	67.0 ± 11.1	71.1 ± 6.6	66.0 ± 6.4
	SVM	69.8 ± 10.8	64.0 ± 10.8	67.8 ± 6.6	66.9 ± 6.3
	XGB	68.4 ± 11.6	65.9 ± 11.3	75.2 ± 6.1	67.1 ± 7.1
	RF	70.4 ± 10.1	65.9 ± 10.7	77.7 ± 4.9	68.2 ± 5.4
4 CS	NB	69.3 ± 11.0	58.6 ± 10.6	67.6 ± 8.3	64.0 ± 7.5
	SVM	76.8 ± 8.7	52.5 ± 12.1	64.7 ± 9.6	64.7 ± 7.9
	XGB	65.5 ± 10.8	59.8 ± 13.4	68.8 ± 7.6	62.6 ± 8.5
	RF	67.5 ± 10.6	56.0 ± 14.3	72.3 ± 7.5	61.8 ± 8.9
5 CS	NB	74.2 ± 9.5	51.4 ± 17.3	67.1 ± 10.9	62.8 ± 9.2
	SVM	84.1 ± 6.6	42.0 ± 14.7	62.7 ± 8.9	63.0 ± 8.0
	XGB	60.6 ± 14.1	57.0 ± 14.4	67.4 ± 9.1	58.8 ± 9.6
	RF	68.1 ± 11.7	52.5 ± 15.5	69.6 ± 9.2	60.3 ± 9.5

Looking in detail at the results obtained with the whole dataset, considering all the patients, we can pinpoint the patterns `[ALS-FRSb=12, ALS-FRSr=4, R=12]` and `[ALS-FRSb= 12, R=12]` from the last snapshot of the set as the patterns that more influence the classifier's decision. To better understand this influence, we used the summary plot from SHAP [31] to visualize how the similarity between the patients and the patterns lead to a positive or negative evolution for respiratory insufficiency. Fig. 5 depicts the impact of each pattern in the model output. Concerning these two patterns, which presented the higher importance and the next four, we can see a good separation between the red and the blue point, which led us to conclude that a low level of similarity with these patterns means an evolution to an NIV state. On the other hand, a high similarity leads to a non-evolution to an NIV state.

The rationale for the clinical decision towards NIV indication incorporates objective assessments (respiratory tests, progression rate and respiratory symptoms) and subjective markers (previous neurologist experience and patient interaction). It would be essential to attain more objective predictive models for the NIV timing, potentially applied in different centres, improving standardization of the clinical care. Regarding clinical interpretability based on features of temporal progression provided by functional status, it is straightforward to understand the critical role of the respiratory symptoms, as given by ALS-FRSr and R. However, our study shows the important role of

Table A.12
BicTric results for the NIV predictive task specialized for **slow progressors**. CS—Number of considered consecutive snapshots/similarity criteria used in matrix T/similarity criteria used in matrix S; C—Pearson correlation, D—Euclidean distance.

CS	Clf	Sensitivity	Specificity	AUC	Accuracy
3/C/C	NB	39.6 ± 14.4	81.4 ± 5.7	67.7 ± 5.9	77.9 ± 4.5
	SVM	59.7 ± 8.1	73.1 ± 6.0	73.6 ± 4.8	72.0 ± 5.4
	XGB	61.4 ± 11.2	80.6 ± 2.7	81.6 ± 4.1	79.0 ± 2.4
	RF	59.4 ± 9.7	82.0 ± 3.4	81.0 ± 4.3	80.1 ± 3.3
3/C/D	NB	33.9 ± 10.2	87.3 ± 3.5	68.8 ± 5.2	82.9 ± 3.3
	SVM	49.7 ± 11.9	76.5 ± 6.0	70.5 ± 5.6	74.3 ± 5.2
	XGB	62.2 ± 12.5	80.8 ± 2.8	81.2 ± 4.3	79.3 ± 2.6
	RF	58.8 ± 11.3	82.0 ± 3.1	81.1 ± 4.6	80.1 ± 2.8
3/D/C	NB	40.5 ± 10.3	82.3 ± 3.7	73.0 ± 4.1	78.8 ± 3.2
	SVM	71.2 ± 8.5	73.7 ± 2.8	81.1 ± 4.1	73.5 ± 2.6
	XGB	61.5 ± 12.2	80.6 ± 3.7	81.6 ± 4.1	79.0 ± 2.7
	RF	60.9 ± 11.7	82.0 ± 3.3	82.0 ± 4.0	80.3 ± 2.6
3/D/D	NB	37.9 ± 9.3	85.4 ± 3.2	74.3 ± 4.8	81.5 ± 3.1
	SVM	70.8 ± 9.9	73.7 ± 3.3	80.6 ± 4.3	73.5 ± 3.1
	XGB	61.5 ± 12.2	80.9 ± 3.3	81.2 ± 4.4	79.3 ± 2.7
	RF	61.1 ± 11.8	81.7 ± 3.5	82.2 ± 4.1	80.0 ± 3.0
4/C/C	NB	37.8 ± 14.1	83.1 ± 3.9	68.5 ± 8.0	78.8 ± 3.3
	SVM	58.7 ± 10.7	78.0 ± 4.4	73.7 ± 6.4	76.2 ± 4.1
	XGB	59.4 ± 13.5	80.0 ± 2.6	78.7 ± 5.9	78.1 ± 2.5
	RF	54.6 ± 11.6	82.0 ± 2.8	79.3 ± 5.0	79.4 ± 2.3
4/C/D	NB	37.6 ± 15.7	84.9 ± 3.9	67.3 ± 6.9	80.4 ± 3.2
	SVM	49.0 ± 12.7	78.6 ± 4.1	70.4 ± 6.6	75.8 ± 3.5
	XGB	62.4 ± 12.1	80.9 ± 3.0	78.8 ± 6.0	79.2 ± 3.0
	RF	55.0 ± 10.8	82.7 ± 2.8	79.2 ± 5.3	80.0 ± 2.7
4/D/C	NB	41.1 ± 14.3	82.1 ± 3.1	72.3 ± 7.2	78.2 ± 2.8
	SVM	66.9 ± 10.7	72.1 ± 4.5	78.1 ± 4.4	71.6 ± 3.7
	XGB	58.5 ± 12.7	80.2 ± 3.1	78.2 ± 6.1	78.2 ± 2.8
	RF	56.3 ± 11.2	83.3 ± 3.2	79.9 ± 5.2	80.7 ± 2.8
4/D/D	NB	40.7 ± 13.1	83.6 ± 3.2	71.0 ± 6.3	79.5 ± 2.8
	SVM	67.8 ± 10.5	71.4 ± 4.4	77.7 ± 4.2	71.0 ± 3.7
	XGB	60.2 ± 13.5	80.7 ± 2.9	77.4 ± 6.2	78.7 ± 2.7
	RF	58.5 ± 12.4	83.0 ± 3.1	80.6 ± 5.2	80.6 ± 2.9
5/C/C	NB	43.1 ± 11.3	82.6 ± 3.5	70.2 ± 7.4	78.3 ± 3.2
	SVM	61.1 ± 12.3	76.1 ± 5.3	75.1 ± 6.6	74.4 ± 4.6
	XGB	64.9 ± 10.7	80.1 ± 3.9	81.1 ± 3.6	78.5 ± 3.1
	RF	56.7 ± 13.0	82.3 ± 3.1	79.4 ± 4.0	79.5 ± 2.9
5/C/D	NB	41.3 ± 12.2	84.9 ± 3.7	70.8 ± 6.9	80.2 ± 3.4
	SVM	52.9 ± 11.2	79.8 ± 4.2	72.9 ± 6.3	76.9 ± 3.6
	XGB	61.6 ± 12.3	80.4 ± 3.8	80.1 ± 4.0	78.3 ± 3.1
	RF	57.6 ± 11.4	81.3 ± 3.5	79.0 ± 4.5	78.8 ± 3.0
5/D/C	NB	44.0 ± 11.5	81.6 ± 3.4	73.1 ± 6.9	77.5 ± 3.2
	SVM	64.2 ± 11.4	74.1 ± 4.1	77.2 ± 4.4	73.1 ± 3.0
	XGB	61.6 ± 11.3	80.6 ± 2.8	79.5 ± 4.3	78.6 ± 2.3
	RF	58.4 ± 13.3	82.3 ± 3.0	80.1 ± 4.6	79.7 ± 2.7
5/D/D	NB	43.1 ± 12.7	83.6 ± 3.7	73.5 ± 6.0	79.2 ± 3.3
	SVM	64.7 ± 12.1	73.6 ± 4.0	76.9 ± 4.3	72.6 ± 2.8
	XGB	59.8 ± 11.4	80.8 ± 3.9	79.4 ± 4.3	78.5 ± 3.5
	RF	57.8 ± 12.7	82.0 ± 3.0	80.3 ± 4.3	79.4 ± 2.7

Table A.13
BicTric results for the NIV predictive task specialized for **neutral progressors**. CS—Number of considered consecutive snapshots/similarity criteria used in matrix T/similarity criteria used in matrix S; C—Pearson correlation, D—Euclidean distance.

CS	Clf	Sensitivity	Specificity	AUC	Accuracy
3/C/C	NB	53.4 ± 7.6	71.2 ± 5.9	69.3 ± 2.8	66.2 ± 3.5
	SVM	59.8 ± 5.6	76.1 ± 4.7	74.6 ± 3.0	71.6 ± 3.2
	XGB	52.7 ± 7.8	80.4 ± 4.5	74.8 ± 3.1	72.8 ± 3.4
	RF	49.8 ± 7.0	82.3 ± 3.0	75.6 ± 2.8	73.4 ± 2.2
3/C/D	NB	57.5 ± 5.7	70.2 ± 3.5	69.6 ± 3.3	66.7 ± 2.6
	SVM	55.5 ± 6.2	72.4 ± 5.7	71.2 ± 3.3	67.7 ± 4.0
	XGB	52.0 ± 7.2	80.2 ± 4.2	74.2 ± 3.5	72.4 ± 3.0
	RF	49.9 ± 7.6	83.1 ± 3.8	75.4 ± 2.8	74.0 ± 2.8
3/D/C	NB	57.1 ± 5.6	69.6 ± 4.8	70.7 ± 2.6	66.2 ± 3.1
	SVM	60.4 ± 6.9	79.9 ± 3.5	77.6 ± 2.3	74.6 ± 2.0
	XGB	52.2 ± 6.8	80.6 ± 3.3	75.5 ± 2.9	72.8 ± 2.6
	RF	51.4 ± 6.1	83.2 ± 3.2	77.1 ± 2.5	74.4 ± 2.5
3/D/D	NB	58.6 ± 5.1	71.2 ± 4.0	71.9 ± 2.9	67.7 ± 2.8
	SVM	63.2 ± 7.1	77.2 ± 3.7	77.6 ± 2.7	73.3 ± 1.9
	XGB	53.3 ± 7.2	81.0 ± 3.2	75.5 ± 3.2	73.4 ± 2.4
	RF	51.3 ± 7.4	83.3 ± 3.4	77.1 ± 2.6	74.4 ± 2.5
4/C/C	NB	56.0 ± 7.3	72.5 ± 5.0	70.9 ± 3.6	67.2 ± 3.8
	SVM	68.4 ± 6.5	71.9 ± 6.0	76.1 ± 4.6	70.8 ± 3.8
	XGB	55.2 ± 5.5	80.2 ± 4.5	72.8 ± 4.2	72.2 ± 3.5
	RF	56.8 ± 7.6	81.6 ± 4.0	75.2 ± 3.9	73.6 ± 3.2
4/C/D	NB	59.5 ± 6.3	71.8 ± 4.5	71.7 ± 4.0	67.8 ± 3.6
	SVM	67.5 ± 6.3	67.2 ± 4.6	73.1 ± 3.8	67.3 ± 2.8
	XGB	55.2 ± 5.8	80.2 ± 4.1	72.9 ± 4.1	72.2 ± 3.1
	RF	55.7 ± 6.9	82.0 ± 4.8	75.3 ± 4.0	73.5 ± 3.5
4/D/C	NB	62.7 ± 6.5	69.2 ± 5.2	71.9 ± 4.0	67.1 ± 3.8
	SVM	67.9 ± 6.1	73.4 ± 4.5	78.6 ± 3.7	71.6 ± 3.5
	XGB	56.3 ± 6.8	80.0 ± 4.2	73.3 ± 4.0	72.3 ± 3.2
	RF	56.5 ± 6.0	82.4 ± 2.9	76.5 ± 3.4	74.0 ± 2.1
4/D/D	NB	62.4 ± 5.9	70.5 ± 4.5	72.8 ± 4.3	67.9 ± 3.8
	SVM	68.3 ± 6.3	72.5 ± 4.8	78.0 ± 3.6	71.1 ± 3.8
	XGB	57.5 ± 6.8	80.2 ± 4.1	73.7 ± 4.1	72.9 ± 3.4
	RF	55.3 ± 5.6	83.1 ± 3.8	76.4 ± 3.3	74.1 ± 2.6
5/C/C	NB	57.6 ± 5.9	72.3 ± 4.3	71.4 ± 3.6	66.8 ± 3.4
	SVM	72.8 ± 6.7	68.0 ± 6.1	75.7 ± 4.6	69.8 ± 3.7
	XGB	60.2 ± 5.9	76.4 ± 4.4	74.6 ± 3.8	70.4 ± 3.3
	RF	61.2 ± 6.7	78.3 ± 4.9	76.9 ± 3.7	72.0 ± 3.6
5/C/D	NB	59.8 ± 6.4	73.2 ± 4.1	72.5 ± 4.2	68.2 ± 3.9
	SVM	69.0 ± 5.3	66.5 ± 4.9	74.0 ± 3.8	67.5 ± 3.7
	XGB	59.7 ± 5.8	77.1 ± 4.4	74.2 ± 3.6	70.7 ± 3.1
	RF	60.7 ± 5.9	79.4 ± 4.3	76.4 ± 3.9	72.5 ± 3.7
5/D/C	NB	62.6 ± 6.1	71.1 ± 4.3	71.9 ± 3.7	68.0 ± 3.6
	SVM	76.0 ± 6.8	66.5 ± 5.0	78.4 ± 4.6	70.0 ± 4.3
	XGB	61.6 ± 4.9	77.9 ± 4.2	75.9 ± 3.6	71.9 ± 2.9
	RF	62.5 ± 8.4	79.1 ± 4.0	78.3 ± 3.8	73.0 ± 3.0
5/D/D	NB	63.6 ± 6.6	71.7 ± 4.5	73.1 ± 4.2	68.7 ± 4.3
	SVM	74.9 ± 6.0	66.2 ± 5.2	77.8 ± 4.6	69.4 ± 4.1
	XGB	61.3 ± 6.1	77.7 ± 4.0	75.5 ± 3.4	71.6 ± 3.1
	RF	60.7 ± 7.1	79.3 ± 4.3	77.8 ± 3.9	72.3 ± 3.2

respiratory function in determining NIV time, which is congruent with clinical data pointing out a worse prognosis in patients with bulbar disability. Moreover, our study indicates that the relative importance of respiratory symptoms and bulbar dysfunction depends on the specific population categorized by progression rate. For fast progressors, respiratory symptoms are the only relevant predictor. However, concerning neutral progressors, respiratory symptoms and bulbar impairment are important, while each variable has a reduced independent predictor value in slow progressors. Our study dissects the clinical decision rationale and can contribute to finding an objective prognostic model. Thus, clinically, all the most relevant patterns in Table 9 are in complete agreement with the classifier decision of non-evolution to NIV state (interpreted by Fig. 5).

Concerning the fast progressors, triclustering did not output a high number of patterns. As we can see by Fig. 6 and Table 9, the pattern `[ALS-FRSr=4, R=12]` is highly discriminative. We can see a delineate separation between the most similar and dissimilar patient profiles from that pattern in Fig. 5.

8. Conclusions and future work

In this work, we proposed to learn predictive models from longitudinal cohort data by mapping the original heterogeneous data space into a feature space described by discriminative patterns learnt by biclustering static data and triclustering temporal data. The produced feature space yield statistical properties of interest and can be subsequently used by state-of-the-art predictive models to leverage both predictability and explainability. Results in the ALS case study, where prognostic models were learned to classify NIV need within 90 days since the last visit, show promising evidence along with both targets.

Table A.14
BicTric results for the NIV predictive task specialized for **fast progressors**. CS—Number of considered consecutive snapshots/similarity criteria used in matrix T/ similarity criteria used in matrix S; C—Pearson correlation, D—Euclidean distance.

CS	Clf	Sensitivity	Specificity	AUC	Accuracy
3/C/C	NB	59.4 ± 12.2	53.3 ± 14.3	61.1 ± 7.1	55.7 ± 7.1
	SVM	74.7 ± 8.5	60.6 ± 12.0	76.1 ± 5.9	65.9 ± 6.7
	XGB	65.2 ± 9.5	75.6 ± 7.9	79.5 ± 5.7	71.7 ± 6.3
	RF	60.8 ± 9.3	72.0 ± 7.7	76.4 ± 6.0	67.8 ± 5.7
3/C/D	NB	52.3 ± 11.6	59.0 ± 10.8	61.5 ± 6.7	56.5 ± 7.0
	SVM	62.6 ± 14.4	57.8 ± 13.7	65.3 ± 8.1	59.6 ± 7.5
	XGB	63.3 ± 10.3	73.7 ± 8.0	77.7 ± 5.7	69.7 ± 6.5
	RF	60.3 ± 10.8	72.8 ± 8.2	75.8 ± 6.3	68.0 ± 6.7
3/D/C	NB	60.0 ± 11.9	52.2 ± 13.3	61.7 ± 7.0	55.2 ± 6.5
	SVM	80.2 ± 7.1	72.2 ± 6.2	82.4 ± 5.4	75.3 ± 4.9
	XGB	66.0 ± 8.5	76.6 ± 6.5	80.4 ± 5.5	72.6 ± 5.2
	RF	61.2 ± 8.6	73.2 ± 6.3	76.8 ± 6.3	68.7 ± 5.2
3/D/D	NB	53.7 ± 12.0	60.6 ± 10.4	62.2 ± 7.1	58.0 ± 7.4
	SVM	77.7 ± 10.2	73.2 ± 6.9	81.0 ± 6.0	74.9 ± 5.6
	XGB	64.4 ± 7.8	75.2 ± 7.2	78.8 ± 6.1	71.0 ± 4.7
	RF	60.0 ± 9.1	73.1 ± 6.9	76.3 ± 5.9	68.1 ± 5.5
4/C/C	NB	68.9 ± 10.0	49.3 ± 9.1	62.9 ± 6.8	57.6 ± 5.5
	SVM	81.7 ± 12.3	54.8 ± 9.1	74.0 ± 7.0	66.2 ± 5.9
	XGB	63.5 ± 10.1	74.7 ± 9.1	76.4 ± 6.2	69.9 ± 6.3
	RF	66.1 ± 9.2	72.4 ± 9.6	76.6 ± 6.2	69.8 ± 6.5
4/C/D	NB	55.6 ± 8.9	58.5 ± 11.7	64.6 ± 8.0	57.3 ± 6.5
	SVM	61.4 ± 14.5	61.3 ± 11.7	69.5 ± 7.2	61.4 ± 6.4
	XGB	62.0 ± 10.6	73.5 ± 9.7	75.5 ± 7.0	68.6 ± 6.2
	RF	64.9 ± 9.0	73.5 ± 8.2	76.5 ± 6.5	69.9 ± 5.7
4/D/C	NB	71.3 ± 9.8	49.5 ± 9.6	65.0 ± 6.7	58.7 ± 5.4
	SVM	80.8 ± 6.8	66.1 ± 10.9	79.0 ± 5.7	72.3 ± 7.1
	XGB	62.0 ± 9.5	75.7 ± 9.1	78.1 ± 5.4	69.9 ± 5.9
	RF	63.6 ± 10.5	72.9 ± 9.2	77.7 ± 6.4	69.0 ± 5.8
4/D/D	NB	57.2 ± 11.0	59.9 ± 11.9	66.0 ± 8.1	58.7 ± 6.9
	SVM	77.8 ± 7.4	67.9 ± 8.9	77.8 ± 6.4	72.1 ± 5.9
	XGB	62.6 ± 9.4	74.4 ± 8.9	76.7 ± 6.4	69.4 ± 5.9
	RF	64.4 ± 9.6	72.9 ± 9.5	77.3 ± 6.3	69.3 ± 5.6
5/C/C	NB	66.4 ± 10.6	49.5 ± 14.5	60.0 ± 8.6	57.3 ± 7.9
	SVM	83.1 ± 9.9	47.8 ± 12.6	70.3 ± 8.6	64.0 ± 7.8
	XGB	66.3 ± 12.2	67.5 ± 10.5	74.9 ± 6.2	67.0 ± 5.9
	RF	68.2 ± 10.6	65.8 ± 9.3	76.4 ± 6.7	66.9 ± 6.1
5/C/D	NB	60.0 ± 10.2	57.6 ± 13.1	65.5 ± 8.2	58.7 ± 7.6
	SVM	60.0 ± 10.9	59.1 ± 14.5	66.6 ± 8.8	59.6 ± 7.8
	XGB	64.7 ± 14.2	66.5 ± 9.2	72.7 ± 6.7	65.7 ± 5.4
	RF	68.1 ± 13.4	66.2 ± 10.8	74.7 ± 7.2	67.1 ± 4.9
5/D/C	NB	67.4 ± 12.1	49.5 ± 15.0	61.5 ± 8.8	57.8 ± 7.8
	SVM	80.4 ± 8.7	61.9 ± 8.7	77.4 ± 6.0	70.4 ± 6.2
	XGB	65.5 ± 12.2	65.3 ± 9.4	74.5 ± 7.2	65.4 ± 5.6
	RF	68.8 ± 10.8	64.8 ± 10.5	76.7 ± 6.7	66.6 ± 5.7
5/D/D	NB	60.5 ± 10.2	58.0 ± 12.9	65.7 ± 8.3	59.1 ± 7.4
	SVM	72.0 ± 13.5	64.6 ± 12.4	75.1 ± 6.1	68.0 ± 6.3
	XGB	65.3 ± 13.4	67.7 ± 10.6	73.6 ± 7.0	66.6 ± 7.2
	RF	70.0 ± 10.8	65.2 ± 10.5	74.9 ± 7.4	67.4 ± 6.1

We further highlight relevant future directions, including BicTric's validation for alternative prognostication end-points in ALS and other neurodegenerative diseases, as well as the enhancement of the underlying biclustering and triclustering searches in accordance with the gathered behavioural limitations. Nevertheless, this work shows the potentialities of putting together biclustering and triclustering stances to learn discriminative prognostic patterns, which can be used not only to learn more accurate prognostic models from high dimensional, heterogeneous and temporal data but also to provide clinicians with novel insights concerning disease progression.

CRediT authorship contribution statement

Diogo F. Soares: Methodology, Implemented the approach, Analysed the data and results, Writing – original draft, Writing – review & editing. **Rui Henriques:** Methodology, Writing – review & editing, Supervision, Revised the results critically. **Marta Gromicho:** Data collection and preprocessing, Definition of ALS case study, Critically analysed the results from a clinical point of view. **Mamede de Carvalho:** Performed the clinical follow-up of the patients' cohort in Lisbon, Data collection and preprocessing, Definition of ALS case study, Critically analysed the results from a clinical point of view. **Sara C. Madeira:** Methodology, Writing – review, Supervision, Revised the results critically.

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.

Acknowledgements

This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, the Portuguese public agency for science, technology and innovation, funding to projects AIpALS (PTDC/CCI-CIF/4613/2020), LASIGE (UIDB/ 00408/2020 and UIDP/00408/2020) and INESC-ID (UIDB/ 50021/2020) Research Units, and PhD research scholarship (2020.05100.BD) to DFS; and by the BRAINTEASER project which has received funding from the European Union's Horizon 2020 research and innovation programme, under the grant agreement No 101017598. All authors approved the final version of the manuscript.

Appendix. Table results

See Tables A.10–A.14.

References

[1] L. Parsons, E. Haque, H. Liu, Subspace clustering for high dimensional data: a review, ACM SIGKDD Explor. Newsl. 6 (1) (2004) 90–105.
[2] S.C. Madeira, A.L. Oliveira, Biclustering algorithms for biological data analysis: a survey, IEEE/ACM Trans. Comput. Biol. Bioinform. 1 (1) (2004) 24–45.
[3] R. Henriques, S.C. Madeira, Triclustering algorithms for three-dimensional data analysis: A comprehensive survey, ACM Comput. Surv. 51 (5) (2019) 95.
[4] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection: A data perspective, ACM Comput. Surv. 50 (6) (2017) 1–45.
[5] A. Bibal, B. Frénay, Measuring quality and interpretability of dimensionality reduction visualizations, in: Safe Machine Learning Workshop at ICLR, 2019.
[6] R. Henriques, S.C. Madeira, FleBic: Learning classifiers from high-dimensional biomedical data using discriminative biclusters with non-constant patterns, Pattern Recognit. 115 (2021) 107900.
[7] A.V. Carreiro, P.M. Amaral, S. Pinto, P. Tomás, M. de Carvalho, S.C. Madeira, Prognostic models based on patient snapshots and time windows: Predicting disease progression to assisted ventilation in amyotrophic lateral sclerosis, J. Biomed. Inform. 58 (2015) 133–144.
[8] L. Ma, J. Gao, Y. Wang, C. Zhang, J. Wang, W. Ruan, W. Tang, X. Gao, X. Ma, Adacare: Explainable clinical health status representation learning via scale-adaptive feature extraction and recalibration, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 825–832.
[9] L. Alexandre, R.S. Costa, L.L. Santos, R. Henriques, Mining pre-surgical patterns able to discriminate post-surgical outcomes in the oncological domain, IEEE J. Biomed. Health Inf. (2021).
[10] S.A. Andersen, G.D. Borasioc, M. de Carvalho, A. Chioe, P. Van Dammef, O. Hardimang, K. Kolleweh, K.E. Morrisoni, et al., EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)–revised report of an EFNS task force, Eur. J. Neurol. 19 (2011) 360–375.
[11] S.C. Bourke, M. Tomlinson, T.L. Williams, R.E. Bullock, P.J. Shaw, G.J. Gibson, Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial, Lancet Neurol. 5 (2) (2006) 140–147.
[12] H.K. van der Burgh, R. Schmidt, H.-J. Westeneng, M.A. de Reus, L.H. van den Berg, M.P. van den Heuvel, Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis, NeuroImage: Clin. 13 (2017) 361–369.
[13] S.R. Pfohl, R.B. Kim, G.S. Coan, C.S. Mitchell, Unraveling the complexity of amyotrophic lateral sclerosis survival prediction, Front. Neuroinform. 12 (2018) 36.

[14] V. Grollemund, P.-F. Pradat, G. Querin, F. Delbot, G. Le Chat, J.-F. Pradat-Peyre, P. Bede, Machine learning in amyotrophic lateral sclerosis: achievements, pitfalls, and future directions, Front. Neurosci. 13 (2019) 135.

[15] A. Zandonà, R. Vasta, A. Chiò, B. Di Camillo, A dynamic Bayesian network model for the simulation of amyotrophic lateral sclerosis progression, BMC Bioinformatics 20 (4) (2019) 118.

[16] S. Pires, M. Gromicho, S. Pinto, M. Carvalho, S.C. Madeira, Predicting non-invasive ventilation in ALS patients using stratified disease progression groups, in: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE, 2018, pp. 748–757.

[17] S. Pires, M. Gromicho, S. Pinto, M. de Carvalho, S.C. Madeira, Patient stratification using clinical and patient profiles: Targeting personalized prognostic prediction in ALS, in: International Work-Conference on Bioinformatics and Biomedical Engineering, Springer, 2020, pp. 529–541.

[18] A.S. Martins, M. Gromicho, S. Pinto, M. de Carvalho, S.C. Madeira, Learning prognostic models using DiseaseProgression patterns: Predicting the need fornon-invasive ventilation in amyotrophic LateralSclerosis, IEEE/ACM Trans. Comput. Biol. Bioinform. (2021).

[19] J. Matos, S. Pires, H. Aidos, M. Gromicho, S. Pinto, M. de Carvalho, S.C. Madeira, Unravelling disease presentation patterns in ALS using biclustering for discriminative meta-features discovery, in: International Work-Conference on Bioinformatics and Biomedical Engineering, Springer, 2020, pp. 517–528.

[20] D. Soares, R. Henriques, M. Gromicho, S. Pinto, M. de Carvalho, S.C. Madeira, Towards triclustering-based classification of three-way clinical data: A case study on predicting non-invasive ventilation in ALS, in: International Conference on Practical Applications of Computational Biology & Bioinformatics, Springer, 2020, pp. 112–122.

[21] H. Das, B. Naik, H. Behera, Medical disease analysis using neuro-fuzzy with feature extraction model for classification, Inform. Med. Unlocked 18 (2020) 100288.

[22] R. Henriques, C. Antunes, S.C. Madeira, A structured view on pattern mining-based biclustering, Pattern Recognit. 48 (12) (2015) 3941–3958.

[23] R. Henriques, S.C. Madeira, BSig: evaluating the statistical significance of biclustering solutions, Data Min. Knowl. Discov. 32 (1) (2018) 124–161.

[24] L. Zhao, M.J. Zaki, TRICLUSTER: An effective algorithm for mining coherent clusters in 3D microarray data, in: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, SIGMOD '05, ACM, New York, NY, USA, 2005, pp. 694–705.

[25] S.C. Madeira, M.C. Teixeira, I. Sa-Correia, A.L. Oliveira, Identification of regulatory modules in time series gene expression data using a linear time biclustering algorithm, IEEE/ACM Trans. Comput. Biol. Bioinform. 7 (1) (2008) 153–165.

[26] F. Divina, B. Pontes, R. Giráldez, J.S. Aguilar-Ruiz, An effective measure for assessing the quality of biclusters, Comput. Biol. Med. 42 (2) (2012) 245–256.

[27] P. Flach, Machine Learning: The Art and Science of Algorithms that Make Sense of Data, Cambridge University Press, 2012.

[28] C. Heffernan, C. Jenkinson, T. Holmes, H. Macleod, W. Kinnear, D. Oliver, N. Leigh, M.-A. Ampong, Management of respiration in MND/ALS patients: An evidence based review, Amyotroph. Lateral Scler. 7 (1) (2006) 5–15.

[29] J.M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, A. Nakanishi, B.A.S. Group, A. complete listing of the BDNF Study Group, et al., The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function, J. Neurol. Sci. 169 (1–2) (1999) 13–21.

[30] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artificial Intelligence Res. 16 (2002) 321–357.

[31] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 4768–4777.

[32] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 2522–5839.