Artificial Intelligence In Medicine 142 (2023) 102588

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

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

Artificial intelligence and statistical methods for stratification and prediction of progression in amyotrophic lateral sclerosis: A systematic review

Erica Tavazzi [a,1], Enrico Longato [a,1], Martina Vettoretti [a,1], Helena Aidos [b,1], Isotta Trescato [a,2], Chiara Roversi [a,2], Andreia S. Martins [b,2], Eduardo N. Castanho [b], Ruben Branco [b], Diogo F. Soares [b], Alessandro Guazzo [a], Giovanni Birolo [c], Daniele Pala [d], Pietro Bosoni [d], Adriano Chiò [e], Umberto Manera [e], Mamede de Carvalho [f], Bruno Miranda [f], Marta Gromicho [f], Inês Alves [f], Riccardo Bellazzi [d], Arianna Dagliati [d], Piero Fariselli [c], Sara C. Madeira [b], Barbara Di Camillo [a,g,*]

[a] Department of Information Engineering, University of Padova, Via Gradenigo 6/b, Padua, 35131, Italy
[b] LASIGE and Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisbon, 1749-016, Portugal
[c] Department of Medical Sciences, University of Torino, Corso Dogliotti 14, Turin, 10126, Italy
[d] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 5, Pavia, 27100, Italy
[e] Department of Neurosciences "Rita Levi Montalcini", University of Turin, Via Cherasco 15, Turin, 10126, Italy
[f] Faculdade de Medicina, Instituto de Medicina Molecular João Lobo Antunes, Universidade de Lisboa, Av. Prof. Egas Moniz, Lisbon, 1649-028, Portugal
[g] Department of Comparative Biomedicine and Food Science, University of Padova, Agripolis, Viale dell'Università, 16, Legnaro (PD), 35020, Italy

ARTICLE INFO

Keywords:
Systematic review
Artificial intelligence
Amyotrophic lateral sclerosis
Stratification
Prediction
Disease progression

ABSTRACT

Background: Amyotrophic Lateral Sclerosis (ALS) is a fatal neurodegenerative disorder characterised by the progressive loss of motor neurons in the brain and spinal cord. The fact that ALS's disease course is highly heterogeneous, and its determinants not fully known, combined with ALS's relatively low prevalence, renders the successful application of artificial intelligence (AI) techniques particularly arduous.

Objective: This systematic review aims at identifying areas of agreement and unanswered questions regarding two notable applications of AI in ALS, namely the automatic, data-driven stratification of patients according to their phenotype, and the prediction of ALS progression. Differently from previous works, this review is focused on the methodological landscape of AI in ALS.

Methods: We conducted a systematic search of the Scopus and PubMed databases, looking for studies on data-driven stratification methods based on unsupervised techniques resulting in (A) automatic group discovery or (B) a transformation of the feature space allowing patient subgroups to be identified; and for studies on internally or externally validated methods for the prediction of ALS progression. We described the selected studies according to the following characteristics, when applicable: variables used, methodology, splitting criteria and number of groups, prediction outcomes, validation schemes, and metrics.

Results: Of the starting 1604 unique reports (2837 combined hits between Scopus and PubMed), 239 were selected for thorough screening, leading to the inclusion of 15 studies on patient stratification, 28 on prediction of ALS progression, and 6 on both stratification and prediction. In terms of variables used, most stratification and prediction studies included demographics and features derived from the ALSFRS or ALSFRS-R scores, which were also the main prediction targets. The most represented stratification methods were K-means, and hierarchical and expectation-maximisation clustering; while random forests, logistic regression, the Cox proportional hazard model, and various flavours of deep learning were the most widely used prediction methods. Predictive model validation was, albeit unexpectedly, quite rarely performed in absolute terms (leading to the exclusion of 78 eligible studies), with the overwhelming majority of included studies resorting to internal validation only.

Conclusion: This systematic review highlighted a general agreement in terms of input variable selection for both stratification and prediction of ALS progression, and in terms of prediction targets. A striking lack of

* Corresponding author at: Department of Information Engineering, University of Padova, Via Gradenigo 6/b, Padua, 35131, Italy.
E-mail address: barbara.dicamillo@unipd.it (B. Di Camillo).
[1] These authors contributed equally.
[2] These authors contributed equally.

https://doi.org/10.1016/j.artmed.2023.102588
Received 22 December 2022; Received in revised form 14 April 2023; Accepted 16 May 2023
Available online 20 May 2023

validated models emerged, as well as a general difficulty in reproducing many published studies, mainly due to the absence of the corresponding parameter lists. While deep learning seems promising for prediction applications, its superiority with respect to traditional methods has not been established; there is, instead, ample room for its application in the subfield of patient stratification. Finally, an open question remains on the role of new environmental and behavioural variables collected via novel, real-time sensors.

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a fatal neurodegenerative disorder characterised by the progressive loss of motor neurons in the brain and spinal cord. Its heterogeneity, due to the interaction of many different, mostly unknown genes and pathophysiological processes, is one of the main obstacles to the development of effective therapies [1]. Moreover, the inability to make accurate predictions of disease progression hampers the possibility to effectively plan therapeutic and supportive interventions, and therefore to optimise schedules of medical appointments and evaluations. The possible solutions proposed to address ALS's extreme prognostic variability are substantially two: the search for reliable biomarkers that could reduce diagnostic delay and improve personalised medicine [2], and the development of robust predictive models based on clinical and biological data [3]. At present, the main prognostic model of ALS survival is based on a European multi-centre study (N = 11,475) [4]. Several papers have also been published addressing ALS disease progression prediction and patient stratification, based on machine learning methods and, more in general, on artificial intelligence (AI). Although their reported performance is promising, the adoption of these methods in clinical practice is limited by a series of challenges that are typical of the use of AI in practical scenarios. In particular, a fundamental aspect that needs to be taken into account is the ability of these methods to generalise, i.e., to work well on previously unseen subjects whose data have not been used to train the model. Therefore, it is crucial that models be suitably tested on independent datasets.

In this work, we present a systematic review of the literature focused on models and methods to stratify ALS patients and predict disease progression, highlighting the variables used, the different prediction outcomes or stratification categories, as well as the methodological framework used to validate the models. In total, more than 2800 reports were screened, and 49 studies included in the systematic review: 15 on patient stratification, 28 on progression prediction, and 6 on both stratification and progression prediction. We aim at identifying areas of agreement and unanswered questions, highlighting how AI can address open issues in the field, and at suggesting directions for future research.

2. Methods

2.1. Search strategy and selection criteria

We conducted a systematic review of AI and statistical methods for patient stratification and prediction of ALS progression. Specifically, we focused on data-driven stratification methods based on unsupervised group discovery or on profile-identifying transformations, and on strictly predictive models of ALS progression that had been (internally or externally) validated.

We searched the Scopus and PubMed databases, and applied the following inclusion criteria (for detailed keyword lists, please see the Supplementary Materials, Section S-1).

1. English-language journal articles or conference papers published between January 1st, 2012 and June 25th, 2021.
2. Name of the disease present in its extended form, i.e., "amyotrophic lateral sclerosis", in either the title or abstract.
3. Indication that stratification or prediction of progression had been conducted (e.g., via the presence of keywords such as "prediction", "stratification", or "profiling").
4. Indication that AI, machine learning, or statistical methods had been used (e.g., via keywords such as "model", "risk score", or "deep learning"), or that environmental variables had been considered.
5. Mention of human subjects (e.g., via keywords such as "patient" or "subject") and no terminology related to animal models of ALS (e.g., via keywords such as "animal testing" or "mice").

Then, we excluded all reports that passed the initial selection via query, but had at least one of the following characteristics.

A1. Unavailable full text.
A2. Study on ALS vs. controls.
A3. Study on ALS vs. other diseases or in the context of other diseases.
A4. Animal studies or in-vitro models.
A5. No ALS or no patients with ALS (e.g., focus on caregivers).
A6. Not English language.
A7. Not journal or conference paper.
A8. Review, meta-analysis, or study protocol.
A9. Study on ALS patients, but not focused on stratification or prediction of progression (e.g., onset prediction, lesion segmentation, electroencephalography studies).
A10. Treatment efficacy or side effects study.

Finally, we only retained studies describing the following.

B1. Data-driven stratification methods based on unsupervised techniques resulting in (A) automatic group discovery (e.g., clusters or trajectories); or (B) a transformation of the feature space allowing a patient subset to be characterised according to individual features, patterns, components, or spatial densities.
B2. Prediction methods subject to, at least, (A) internal validation via an estimate of the generalisation error (e.g., cross-validation or bootstrap) or via a hold-out set, i.e., a portion of the initial dataset that was not used for model development but set aside for evaluation; or (B) external validation on a different dataset from the one used for model development.

2.2. Data analysis

All authors contributed to devising the queries and further refinement steps; ET and CR ran the queries and removed duplicate (same DOI) and retracted entries; ET, EL, MV, HA, IT, CR, ASM, ENC, RB, DFS, AG, GB, DP, AD, PF, SM, and BDC read, evaluated, and classified the studies according to the presence of stratification or prediction-of-progression methods. These authors initially read a randomly-assigned subset of reports checking for clear-cut reasons for exclusion and providing a preliminary classification. In case of doubts or ambiguities, at least another author, almost always from another institution, provided his or her opinion. ET, EL, and MV considered all the opinions and resolved all the remaining conflicts. During the data collection process, HA, ASM, ENC, RB, DFS, and SM thoroughly examined the reports classified as describing stratification methods; ET, EL, MV, IT, CR, AG, and BDC those describing prediction-of-progression methods. Furthermore, GB and PF examined the reports specifically focused on deep learning; and DP, PB, and AD those about environmental variables. Overall, at least two authors read each document. We collected the following data from each included study: dataset, dataset availability, sample size, variable categories, use of longitudinal data, prediction outcome, prediction method category, prediction validation

type, prediction validation scheme, prediction validation metric, stratification method category, stratification criteria category, and number of stratified groups.

We collected the aforementioned information in a main table (Table 1) sorted by *Aim*, with stratification-only studies at the beginning, studies dealing with both stratification and prediction of progression in the middle, and prediction-only studies at the end.

Starting from the main table (Table 1), we produced a number of pivot tables to further describe the selected studies from multiple relevant points of view: the variables used for stratification (Table S-2, reported in the Supplementary Materials) and prediction of progression (Table S-4, reported in the Supplementary Materials); the specific methods used (Tables S-1 and S-3, respectively, reported in the Supplementary Materials); the splitting criteria and number of groups found in stratification studies (Table 2); the outcomes (Table 3), validation schemes (Table 4), and metrics (Table 5) found in prediction studies. Each pivot table expands the categories reported in the main table into more detailed subcategories.

2.3. Risk of bias

We identified the following main sources of risk of bias: the limited number of available datasets (correlated to the low prevalence of ALS), the types of laboratory and other tests that are performed in routine care vs. in the context of exploratory studies, and the overwhelming tendency to publish meaningful rather than inconclusive results. These risks, while non-negligible, are mitigated by the fact that this systematic review does not aim at quantifying a clinically-relevant endpoint, but at describing the current state of the art of the methodology, which, by definition, includes all the above-mentioned aspects.

2.4. Registration information and study protocol

Details of the protocol for this systematic review were registered on PROSPERO (ID: CRD42021288026) and can be accessed at www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42021288026 [53].

2.5. Role of the funding source

The study was supported by the Horizon 2020 project BRAINTEASER (Bringing Artificial Intelligence home for a better care of amyotrophic lateral sclerosis and multiple sclerosis). BRAINTEASER has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. GA101017598 (start date 01/01/2021).

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

3. Results

3.1. Report screening and filtering

Report screening and filtering are summarised in Fig. 1. The queries selected 2837 reports in total (1402 from PubMed and 1435 from Scopus). After removing duplicates (1230 reports) and retracted articles (3 reports), 1604 reports were retained for the first screening phase. During the first screening phase, 13 reports were excluded because it was not possible to retrieve their full text. The remaining 1591 reports were assessed for eligibility to be included in the analysis. During this step, 1352 reports were excluded because they met one of the exclusion criteria reported in Section 2.1 (A2–A10). In total, 239 studies were selected in this initial screening step: 127 studies on ALS patient stratification with no prediction of progression, 63 studies with prediction of ALS progression but no patient stratification, and 49 studies with both patient stratification and prediction of progression. The selected studies were then filtered according to criteria B1 and B2. 112 studies were excluded from patient stratification with no prediction of progression because they did not meet criterion B1. Of the studies initially assigned to both stratification and prediction of progression, 7 studies were moved to the set of on prediction of progression without stratification based on B1, and other 36 studies were removed from the analysis because neither B1 nor B2 were satisfied. Finally, 42 studies were excluded from those on prediction of ALS progression without patient stratification because they did not meet criterion B2. In total, 49 studies were included in the systematic review: 15 on patient stratification without prediction of progression, 6 on both stratification and prediction of progression, and 28 on prediction of progression without patient stratification (including the aforementioned 7 studies with prediction of progression but non-data-driven stratification; see criterion B1). The characteristics of each study included in the systematic review are summarised in Table 1. Fig. 2 reports the identified works stratified by publication year and aim.

3.2. Data-driven stratification methods

In this section, we summarise the current trends in data-driven stratification of ALS patients. Specifically, we, first, focused on the splitting criteria most frequently used in the literature, and the resulting number of groups, which may serve as a useful starting point for further studies. Then, we compiled an exhaustive list of the methodologies used to identify or characterise the groups, as well as the variables upon which the clustering was based.

In doing so, a total of 21 studies on ALS stratification were selected, with 15 of them meeting only stratification criteria, and 6 including both stratification and prediction. As mentioned above, the main topics of analysis were the splitting criteria for stratification, the number of groups discovered, the data-driven methods applied, and the input variables.

A summary of the splitting criteria is presented in Table 2. Eight studies (38%) identified between 2 and 6 groups based on functional decline measures, such as the decrease in ALSFRS in a given time period, the rate of bulbar decline, and the ALSFRS-R trajectory. Two studies (10%) found 3 patient groups according to SNIP respiratory measurements and FVC trajectories. Two studies (10%) also performed stratification according to survival measures, namely survival rate, time until death, tracheostomy, or NIV. Stratification according to clinical profiles was performed in three studies (14%), with two of them encountering 3 groups based on cognitive and behavioural profile, as well as one study performing a 3-way stratification through prognostic, respiratory, and functional profiling. Additionally, one study (5%) found 4 patient groups according to a stimuli response measure, and another (5%) found 3 groups based on biomarker pathology. Finally, 4 studies (19%) had composite splitting criteria, typically combining ALSFRS-R values with other clinical measures to discover 3 or 4 groups.

The methods used for group discovery are referenced in Table S-1 reported in the Supplementary Materials and in Fig. 3(a). Clustering techniques, specifically K-means, and hierarchical and expectation-maximisation clustering, were the most commonly employed (7 studies, 33%), followed by dimensionality reduction techniques (4 studies, 19%) and survival models (4 studies, 19%). UMAP and principal component analysis were used to find spatial density and profile-based groups, respectively. The ENCALS survival model is a tool for survival time prediction, while RECPAM is a tree-based model that identifies survival risk groups. Mixed effects models (2 studies, 10%) and mixture models (2 studies, 10%) were applied for trajectory group identification. Trajectory model D50 (2 studies, 10%) maps ALSFRS decline and identifies disease phases at a given point in time.

Input variables used in the stratification approaches are presented in Fig. 4(a) and in Table S-2 of the Supplementary Materials. Most stratification studies (16, 76%) made use of ALS-specific variables such as the ALSFRS or ALSFRS-R scores, subscores, slopes, and individual questions, as well as onset site. Diagnostic delay and the revised El Escorial criteria were used in 4 studies (19%), while the remaining features were typically study-specific. Demographic features, namely

Table 1
List of papers included in the systematic review. For each paper, reported data include: the reference, the aim of the paper (S = stratification; P = prediction of progression; S, P = stratification and prediction), the name or the source of the dataset used, the type of data availability, the dataset's sample size, the type of variables used, and if longitudinal variables were used as input features. For papers on prediction of progression the table further reports the outcome of the model, the method category, as well as validation's type, scheme, and metric. For papers on stratification, instead, the method category, the criteria category, and the number of groups are reported. For the definition of acronyms, please see the List of acronyms section.

Ref	Aim	Dataset	Data Availability	Sample Size	Variable Categories	Longitudinal Data	Prediction of progression					Stratification		
							Outcome	Method Category	Validation Type	Validation Scheme	Validation Metric	Method Category	Criteria Category	N groups
[5]	S	Data of patients followed at King's College MND Clinic (London, UK)	Upon request/ private	31	Demographics, ALS-specific variables, Other clinical measurements, Previous pathologies or comorbidities	No						Survival models	Survival	2
[6]	S	Data of patients followed at Jena University Hospital (Jena, Germany)	Upon request/ private	145	ALS-specific variables	Yes						Trajectory models	Functional decline	2
[7]	S	Data of patients followed at the Carlo Besta Neurological Institute (Milan, Italy)	NA	102	Previous pathologies or comorbidities	No						Dimensionality reduction	Clinical profile	3
[8]	S	Prospective population-based registry of Puglia, Italy	NA	94	Demographics, ALS-specific variables, Previous pathologies or comorbidities	No						Survival models	Multiple criteria	3
[9]	S	Data of patients followed at Beaumont Hospital (Dublin, Ireland)	NA	89	Clinical Measurements	No						Clustering	Stimuli response	4
[10]	S	Data of patients autopsied at Houston Methodist Hospital (Houston, TX, USA)	NA	57	ALS-specific variables	No						Clustering	Biomarker pathology	3
[11]	S	Data of patients followed at the Carlo Besta Neurological Institute (Milan, Italy)	NA	71	Previous pathologies or comorbidities	No						Dimensionality reduction	Clinical profile	3
[12]	S	Japanese Consortium for Amyotrophic Lateral Sclerosis Reasearch (JaCALS) registry	NA	465	ALS-specific variables	Yes						Mixed Effects Models	Functional decline	4
[13]	S	French register of ALS (FRALim)	NA	322	Demographics, ALS-specific variables	No						Survival models	Multiple criteria	4
[14]	S	Data of patients followed at the University of Bari Aldo Moro MND Centre (Bari, Italy)	NA	100	Demographics, ALS-specific variables, Clinical Measurements	No						Survival models	Respiratory function	3

(*continued on next page*)

Table 1 (*continued*).

[15]	S	NA	NA	54	Clinical Measurements	Yes						Mixed Effects Models	Functional decline	2
[16]	S	Data of patients followed at Jena University Hospital (Jena, Germany)	Open	156	ALS-specific variables	Yes						Trajectory models	Functional decline	3
[17]	S	Trophos, Exonhit and the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) datasets, data of patients followed at Pitié Salpètrière Hospital Assistance Publique des Hôpitaux de Paris (Paris, France)	Open	3756	Demographics, ALS-specific variables, Anthroprometrics	No						Dimensionality reduction	Functional decline	3
[18]	S	ONWebDUALS dataset (Lisbon patients only)	NA	473	ALS-specific variables	Yes						Clustering	Functional decline	3
[19]	S	PRO-ACT, Penn Integrated Neurodegenerative Disease Database	Open	7461 (PRO-ACT), 837 (Penn)	ALS-specific variables, Clinical Measurements	Yes						Mixture models	Respiratory function	3
[20]	S, P	Database of the ALS Clinic in Lisbon, Medical Academic Centre of Lisbon (Lisbon, Portugal).	Upon request/ private	1360	Demographics, Clinical measurements, Anthropometrics, ALS-specific variables	No	FVC	Classification	Internal	Repeated K-fold cross-validation	Sensitivity, Specificity, AUROC	Clustering	Clinical profile	2-4
[21]	S, P	The Trophos and the Exonhit datasets from the Northeast ALS Consortium. PRO-ACT. Data of patients followed at Pitié Salpètrière Hospital Assistance Publique des Hôpitaux de Paris (Paris, France)	Open	5220	Demographics, Clinical measurements, Anthropometrics (Prediction only), ALS-specific variables	No	Survival	Classification	External	External dataset	Accuracy, Sensitivity, Specificity, Precision, F1, Balanced accuracy, AUROC	Dimensionality reduction	Survival	3
[22]	S, P	PRO-ACT.	Open	2424	Demographics, Clinical measurements, Quantitative laboratory measurements, Anthropometrics (Stratification only), ALS-specific variables	Yes	ALSFRS or ALSFRS-R, FVC	Regression	Internal	Holdout set	R-squared, RMSE, Expected/ observed comparison	Clustering	Multiple criteria	4

(*continued on next page*)

gender and age at onset or baseline, were also commonly used (9 studies, 43%). Nine studies (43%) include other clinical measurements, most commonly FVC (7 studies, 33%), pulse and blood pressure (3 studies, 14%), with the remaining features being study-specific. Five studies (24%) used anthropometric variables, 4 (19%) of which including weight. Laboratory measurements and presence of previous pathologies and comorbidities were included in 2 (10%) and 4 (19%) studies, respectively.

Table 1 (*continued*).

[23]	S, P	PRO-ACT. The Irish National ALS Register. The Piemonte and Valle d'Aosta Amyotrophic Lateral Sclerosis (PARALS) register.	Open	10723 (PRO-ACT), 1479 (Irish National ALS Register and PAR-ALS)	Demographics, Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	Yes	ALSFRS or ALSFRS-R, Survival	Regres-sion, Survival	Internal and External	Holdout set, External dataset	C-index, RMSE, Pearson's correlation	Cluster-ing	Multiple criteria	4
[24]	S, P	PRO-ACT.	Open	338	ALS-specific variables	Yes	ALSFRS or ALSFRS-R	Classifi-cation	Internal	Holdout set	AUROC	Mixture models	Functional decline	2
[25]	S, P	PRO-ACT.	Open	2590	ALS-specific variables	Yes	ALSFRS or ALSFRS-R	Classifi-cation	Internal	Holdout set	Accuracy, F1, MAE	Cluster-ing	Functional decline	4-6
[26]	P	Database of the ALS Clinic in Lisbon, Medical Academic Centre of Lisbon (Lisbon, Portugal).	Upon request/ private	1375	Demographics, Clinical measurements, Anthropometrics, ALS-specific variables	Yes	FVC	Classifi-cation	Internal	Repeated K-fold cross-validation	Sensitivity, Specificity, AUROC			
[27]	P	Database of the ALS Clinic in Lisbon, Medical Academic Centre of Lisbon (Lisbon, Portugal).	Upon request/ private	1214	Demographics, Clinical measurements, Anthropometrics, ALS-specific variables	Yes	ALSFRS or ALSFRS-R, FVC	Proba-bilistic graph	Internal	K-fold cross-validation	Accuracy, Sensitivity, AUROC			
[28]	P	The Piemonte and Valle d'Aosta Amyotrophic Lateral Sclerosis (PARALS) register.	Upon request/ private	683	Demographics, Clinical measurements, Anthropometrics, ALS-specific variables	Yes	Survival	Classifi-cation	Internal	Holdout set	AUROC, AUPRC			
[29]	P	PRO-ACT.	Upon request/ private	3772	Demographics, Clinical measurements, Quantitative laboratory measurements, ALS-specific variables	Yes	ALSFRS or ALSFRS-R, FVC	Classifi-cation, Proba-bilistic graph	Internal	K-fold cross-validation	Accuracy			
[30]	P	Database of the ALS Clinic in Lisbon, Medical Academic Centre of Lisbon (Lisbon, Portugal)	Upon request/ private	1220	Demographics, Clinical measurements, Anthropometrics, ALS-specific variables	Yes	FVC	Classifi-cation	Internal	Repeated K-fold cross-validation	Accuracy, Sensitivity, Specificity, AUROC			
[31]	P	PRO-ACT.	Open	4957	Demographics, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	Yes	ALSFRS or ALSFRS-R, FVC, Survival	Proba-bilistic graph	Internal	K-fold cross-validation	AUROC, Expected/ observed comparison			

(*continued on next page*)

3.3. Predictive models of ALS progression

In this section, we present an overview of the literature landscape of predictive models of ALS progression with the double aim of describing the current trends in this field, and providing a synthesis of the main coordinates needed for the development of a new model building upon the existing state of the art. Hence, first we focused on the clinical outcomes that are most frequently predicted, both to summarise the most relevant aspects of ALS' prognosis, and to highlight potential gaps. Then, we analysed the used methodologies, identifying the most relevant frameworks (e.g., classification, survival analysis) for each outcome, as well as the variables selected as predictors. Finally, given

Table 1 (*continued*).

[32]	P	Data of patients treated at the University of Pennsylvania Comprehensive ALS Center (Philadelphia, PA, United States).	Upon request/ private	765	Demographics, ALS-specific variable	No	Survival	Classifi-cation	Internal and External	K-fold cross-validation, External dataset	Sensitivity, Specificity, Precision, NPV, AUROC, Chi-squared test, Hosmer–Lemeshow test
[33]	P	Data from the Gene Expression Omnibus (GEO) repository (GSE112676 and GSE112680).	Open	396	Demographics, ALS-specific variable	No	Survival	Survival	Internal	Repeated K-fold cross-validation	C-index
[34]	P	Data of patients treated at the San Raffaele Hospital in Milan (Milano, Italy).	Upon request/ private	149	Demographics, Clinical measurements, Anthropometrics, ALS-specific variables	No	Survival	Survival	Internal	K-fold cross-validation	AUROC
[4]	P	Data of patients treated in 14 specialised ALS centres in Belgium, France, the Netherlands, Germany, Ireland, Italy, Portugal, Switzerland, and the UK.	Upon request/ private	11475	Demographics, ALS-specific variables, Previous pathologies or comorbidities	No	Survival	Survival	Internal and External	Leave-one-dataset-out cross-validation	AUROC, Calibration plot, Calibration slope
[35]	P	Data of patients treated at the Sunnybrook Research Institute (Toronto, Canada).	Upon request/ private	64	Clinical measurements	No	ALSFRS or ALSFRS-R	Classifi-cation	Internal	Repeated K-fold cross-validation	Accuracy
[36]	P	Data of patients treated at the Emory ALS Clinic (Emory University Hospital, Atlanta, GA, United States).	Upon request/ private	801	Demographics, Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	No	Survival	Regres-sion, Classifi-cation	Internal	K-fold cross-validation	AUROC, R-squared, RMSE
[37]	P	Data of patients treated at Department of Neurology, West China Hospital of Sichuan University (Chengdu, Sichuan province, China).	Upon request/ private	553	Demographics, Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	No	Survival	Classifi-cation	Internal	Holdout set	AUROC
[38]	P	PRO-ACT.	Open	4346	Demographics, Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	Yes	FVC	Regres-sion	Internal and External	K-fold cross-validation, External database	R-squared, RMSE

(*continued on next page*)

Table 1 (*continued*).

[39]	P	PRO-ACT.	Open	9924	Demographics, Quantitative laboratory measurements, ALS-specific variables	Yes	ALSFRS or ALSFRS-R, FVC	Probabilistic graph	Internal	Holdout set	MSE, Expected/observed comparison
[40]	P	PRO-ACT.	Open	8635	Demographics, Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	Yes	ALSFRS or ALSFRS-R, Survival	Regression, Classification	Internal	K-fold cross-validation	AUROC, Normalised RMSE
[41]	P	PRO-ACT.	Open	3742	Demographics, Clinical measurements, Anthropometrics, ALS-specific variables	Yes	ALSFRS or ALSFRS-R	Regression	Internal and External	Holdout set, External dataset	RMSE
[42]	P	Data of patients treated at the NEMO (NEuroMuscolar Omnicentre) Clinical Centre.	Upon request/private	428	Demographics, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	No	Survival	Survival	Internal and External	Holdout set, External dataset	AUROC, Expected/observed comparison
[43]	P	The Irish National ALS Register.	Upon request/private	204	ALS-specific variables	No	Survival	Survival	Internal and External	Holdout set, External dataset	Precision, NPV, Expected/observed comparison
[44]	P	PRO-ACT.	Open	1822	Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	Yes	ALSFRS or ALSFRS-R	Regression	Internal	Holdout set	RMSE, Pearson's correlation
[45]	P	Database of the ALS Clinic in Lisbon, Medical Academic Centre of Lisbon (Lisbon, Portugal).	Upon request/private	517	Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables	Yes	FVC	Classification	Internal	Repeated K-fold cross-validation, Holdout set	AUROC
[46]	P	PRO-ACT.	Open	1822	Demographics, Clinical measurements, Quantitative laboratory measurements, Anthropometrics, ALS-specific variables, Previous pathologies or comorbidities	Yes	ALSFRS or ALSFRS-R	Regression	Internal	Holdout set	RMSE, Pearson's correlation
[47]	P	PRO-ACT.	Open	1197	Anthropometrics, ALS-specific variables	Yes	ALSFRS or ALSFRS-R	Classification	Internal	Holdout set	Accuracy

(*continued on next page*)

the importance of validating new predictive models and comparing them to existing benchmarks, we focused on the validation techniques and metrics used to measure model performance in the literature.

The selected studies on ALS progression prediction were in total 34 (28 prediction of progression only, and 6 stratification and prediction). As previously stated, these studies were analysed with a focus on

Table 1 (*continued*).

[48]	P	The South-East England Amyotrophic Lateral Sclerosis (SEALS) population register.	Upon request/ private	713	Demographics, ALS-specific variables	No	Survival	Survival	Internal	Holdout set	Expected/ observed comparison
[49]	P	Database of the ALS Clinic in Lisbon, Medical Academic Centre of Lisbon (Lisbon, Portugal).	Upon request/ private	1110	NA	Yes	FVC	Classifi-cation	Internal	Holdout set	AUROC
[50]	P	Patients treated at the clinic for motor neuron diseases of the University Medical Center Utrecht (UMCU)	Upon request/ private	268	Demographics, Clinical measurements, ALS-specific variables, Previous pathologies or comorbidities	No	Survival	Classifi-cation	Internal	Holdout set, External dataset	Accuracy
[51]	P	PRO-ACT.	Open	1936	Demographics, Clinical measurements, Quantitative laboratory measurements, ALS-specific variables	Yes	Survival	Classifi-cation	Internal	K-fold cross-validation	Accuracy, AUROC, C-index, MAE
[52]	P	Data of patients treated at the outpatient clinic for motor neuron diseases of the University Medical Center Utrecht (Utrecht, the Netherlands).	Upon request/ private	135	Demographics, Clinical measurements, ALS-specific variables, Previous pathologies or comorbidities	No	Survival	Classifi-cation	Internal	Holdout set	Accuracy

the considered outcome, the method used, the input variables, the validation method and the performance metrics.

The considered outcomes are summarised and referenced in Table 3. Twelve studies (35%) focused on the prediction of ALSFRS or ALSFRS-R score, subscores, or slope, with prediction horizons ranging between 3 and 36 months. Twelve studies (35%) considered some outcomes related to the respiratory function and interventions, e.g., FVC, need of NIV, respiratory insufficiency, or tracheostomy. A few studies focused on predicting progression trajectories (2 studies, 6%) or a slow/fast progression class (1 study, 3%). Death was considered as a target outcome in 16 studies (47%).

Concerning the methodology used, 8 studies (24%) approached the problem as a regression task (prediction of a numerical outcome). The approaches more commonly used for regression problems were linear regression and GLM (3 studies, 38%), random forest (8 studies, 100%), and other tree-based models (5 studies, 63%). SVM was used only in 2 studies (25%). Nineteen studies (56%) focused on a classification problem (prediction of a categorical outcome). The methodologies most commonly used for tackling classification problems were logistic regression (8 studies, 42%), naïve Bayes (5 studies, 26%), SVM (4 studies, 21%), random forest (9 studies, 47%), and other tree-based methods (6 studies, 32%). One study considered an ensemble classifier including naïve Bayes, logistic regression, and decision trees. Five studies (26%) investigated deep learning approaches based on deep neural networks (NN), e.g., LSTM, recurrent NN, convolutional NN, and dense multilayer NN. In 7 studies (21%), the prediction task was framed as a survival analysis task, with most of the studies using the Cox proportional hazard model (5 studies, 71%). Finally, 3 studies (9%) proposed a probabilistic graphical model based on dynamic Bayesian networks. The methods used for predicting progression are referenced in Table S-3 of the Supplementary Materials and in Fig. 3(b).

Table S-4 of the Supplementary Materials and Fig. 4(b) show the full list of variables that were used as inputs to at least one of the analysed predictive models. These variables included demographics (e.g., age, ethnicity, gender), clinical measurements collected during routine visits (e.g., pulse, blood pressure, respiratory rate), quantitative laboratory measurements (e.g., white blood cell count, albumin, cholesterol, triglycerides levels), anthropometric measurements (e.g., height, weight), ALS-specific variables related to disease onset and progression (e.g., ALSFRS scores, need of NIV, FVC), and previous pathologies or comorbidities (e.g., diabetes, frontotemporal dementia, cardiovascular diseases). The variables most frequently used were age at onset, ALSFRS (or ALSFRS-R) score, and onset site, used by 74%, 71%, and 68% of the models, respectively. The frequency of use of the other variables was below 50%, with most variables used by just one or two models.

Internal or external validation was one of the inclusion criteria of this systematic review. As summarised in Table 4, 25 studies, i.e., 74% of the total, included only an internal validation, performed on a holdout dataset (12 studies, 48%), via K-fold cross validation (6 studies, 24%), or via a repeated K-fold cross-validation scheme (7 studies, 28%). One study (3%) included only an external validation. Just 8 studies, i.e., the 24% of the total, included both internal and external validation.

The performance metrics used for model assessment are reported in Table 5. For regression models, model performance was most frequently assessed by the root mean squared error (100%), followed by Pearson's correlation (50%) and R-squared (38%). For classification models, all the studies assessed discrimination ability, most commonly using the area under the receiver-operating characteristic curve (AUROC; 68%).

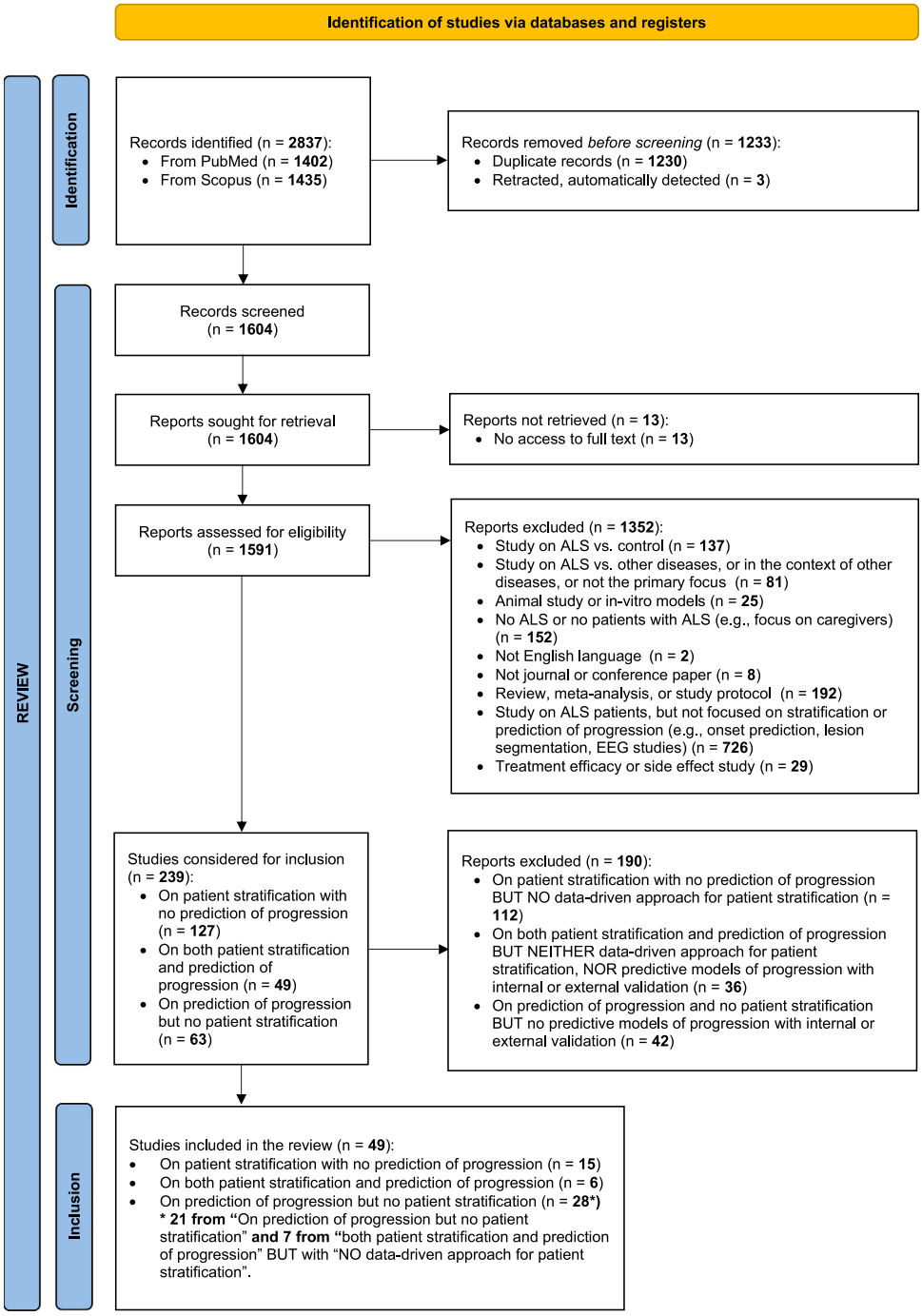


Fig. 1. PRISMA flowchart summarising the steps of the screening process with details about the number of reports selected by the initial query, the number of reports excluded at each screening step, and the number of studies finally selected for the review.

Other discrimination metrics were: accuracy (47%), sensitivity and specificity (26%), mean absolute error (16%), precision (11%), F1-score (11%), negative predictive value (NPV, 5%), area under the precision–recall curve (5%) and C-index (5%). The calibration of the models was assessed only by one study using the Hosmer-Lemeshow test. Survival models were assessed using C-index (29%), AUROC (43%), precision (14%), and NPV (14%) as discrimination metrics. Four survival studies (57%) also assessed model calibration using expected/observed comparison or calibration plot and slope. Finally, the performance metrics used for probabilistic graphical models were: AUROC (67%), expected/observed comparison (67%), accuracy and sensitivity (33%), and mean squared error (33%).

4. Discussion

Neurodegenerative diseases such as ALS are characterised by multimorbidity and progressive impairment involving all neurological functions. Patients have to manage an increasing need for care at home and alternated periods in hospitals, leading to feelings of uncertainty regarding their future (e.g., in terms of timing of disease phases), often exacerbated by increasing psychological and economic burdens. At the same time, clinicians would benefit from additional tools aimed at supporting their patients through the phases of the disease, e.g., as an aid to personalise treatment regimens or get alerts on urgent interventions, and aid clinical trial design and analyses.

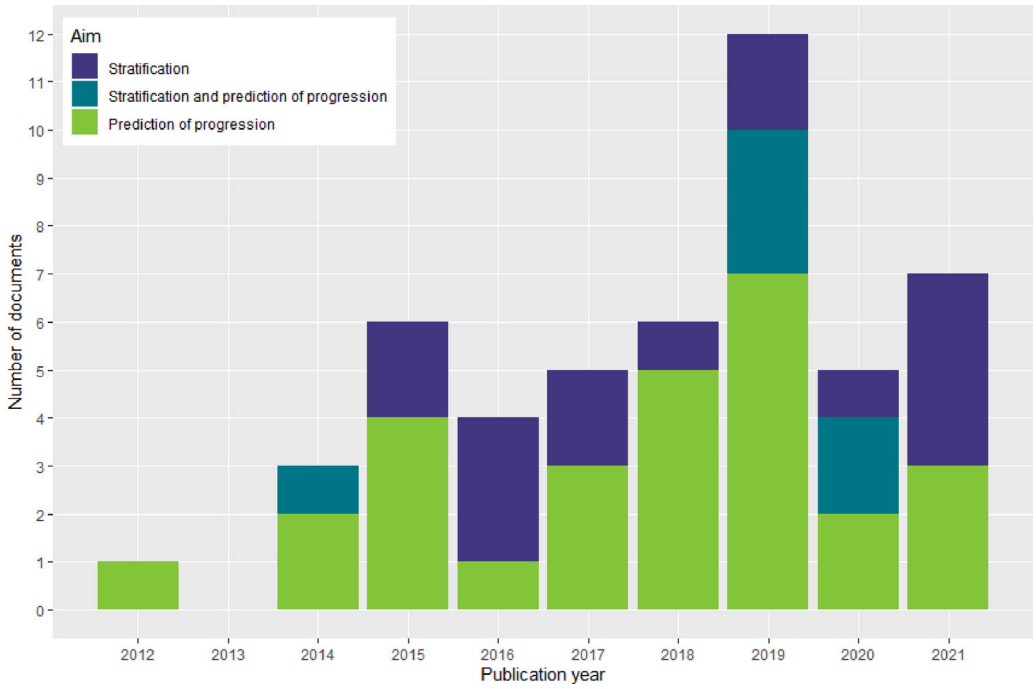


Fig. 2. Number of identified studies by publication year. Studies on prediction of progression only are displayed in light green, studies on stratification and prediction of progression in dark green, and studies on stratification only in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

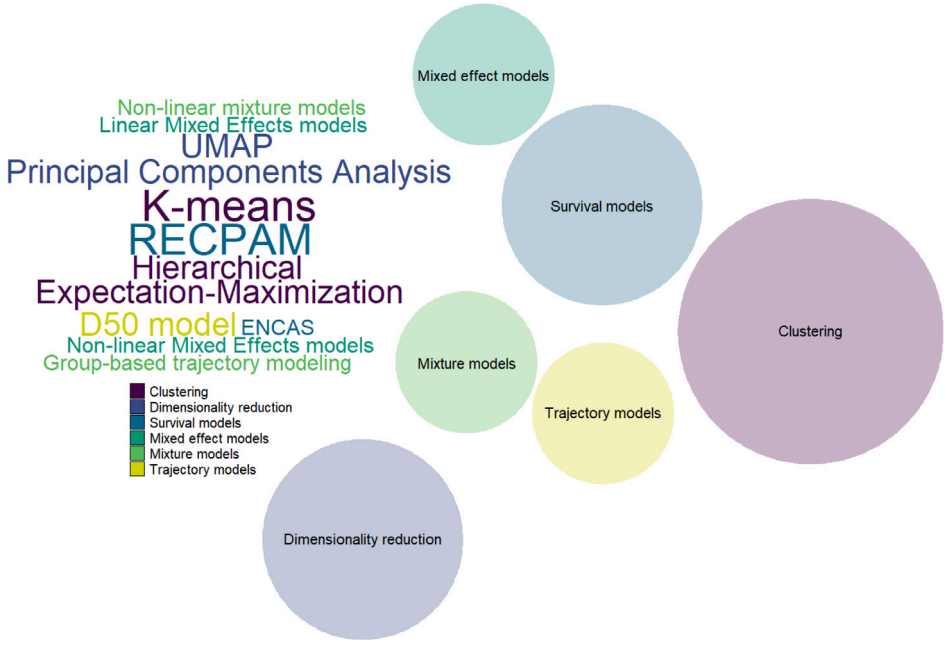
Table 2
Summary of the splitting criteria used in the identified stratification models. In particular, splitting criteria are grouped in more general splitting criteria categories (first column). Then, for each specific criterion, the number of groups considered is reported as well as the references to the papers using that criterion. Finally, the last column reports the number of papers using each criterion and, in brackets, the percentage of papers using each criterion, relative to the total number of stratification papers (i.e., 21). For the definition of acronyms, please see the List of acronyms section.

Splitting criteria category	Splitting criteria	Nr. of groups	Ref.	Counts (%)
Functional decline	ALSFRS value decline	2–6	[17], [18], [24], [25]	4 (19%)
	Bulbar decline	2	[15]	1 (5%)
	ALSFRS-R trajectory	2–4	[6], [12], [16]	3 (14%)
Respiratory function	SNIP value	3	[14]	1 (5%)
	FVC modelled trajectory	3	[19]	1 (5%)
Survival	1-yr survival rate	3	[21]	1 (5%)
	Time to composite outcome (death, tracheostomy, NIV >23h)	5	[5]	1 (5%)
Clinical profile	Cognitive/Behavioural profile	3	[7], [11]	2 (10%)
	Prognostic profile	2	[20]	1 (5%)
	Functional profile	2	[20]	1 (5%)
	Respiratory profile	4	[20]	1 (2%)
Stimuli response	Average auditory mismatch negativity (MMN) delay	4	[9]	1 (5%)
Biomarker pathology	TDP-43 pathology	3	[10],	1 (5%)
Multiple criteria	Multiple criteria, such as: ALSFRS/ALSFRS-R, FVC.	3–4	[8], [13], [22], [23]	4 (19%)

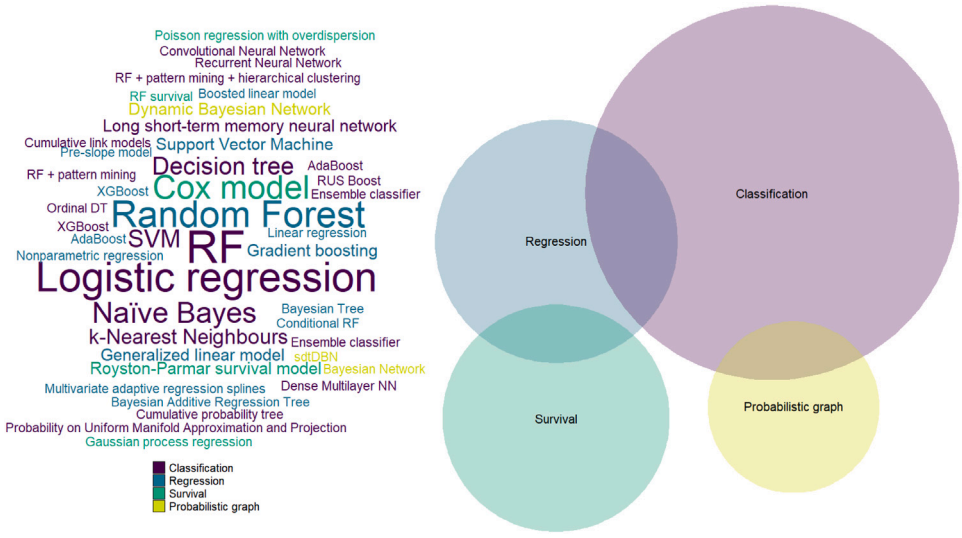
Thus, being able to predict the progression path of ALS patients is extremely important to improve prognostication and intervention-timing in routine clinical practice. Moreover, clinical trials could be more effectively designed, e.g., by ensuring the allocation of equivalent expected-outcomes populations to the various intervention arms of a trial, or by allowing a more accurate estimation of the required sample size as well as particular needs for the follow-up. Similarly, stratification of patients allows for analysing differences in disease management and medical complications between groups and provides relevant information to optimise disease management strategies, including end-of-life decisions.

4.1. Main results and future directions

In this work, we systematically reviewed the literature on models to stratify ALS patients and predict disease progression. In total, more than 2800 reports were screened, and 49 studies were included in the systematic review: 15 on patient stratification without progression prediction, 28 on progression prediction without patient stratification, and 6 on both stratification and progression prediction. Differently from previous works, we specifically focused on data-driven stratification methods based on unsupervised group discovery or on



(a) Stratification methods.



(b) Prediction of progression methods.

Fig. 3. Summary of the identified stratification (panel a) and prediction of progression methods (panel b). The coloured bubbles in the right part of each panel represent the different method categories; the size of each bubble is proportional to the number of studies using a method of that category; intersections between bubbles represent studies using methods of more than one category. The word clouds to the left of each panel represent the specific methods used by the identified studies; the colour of the word corresponds to the method category; the size of the word is proportional to the number of studies using that method. For the definition of acronyms, please see the List of acronyms section. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

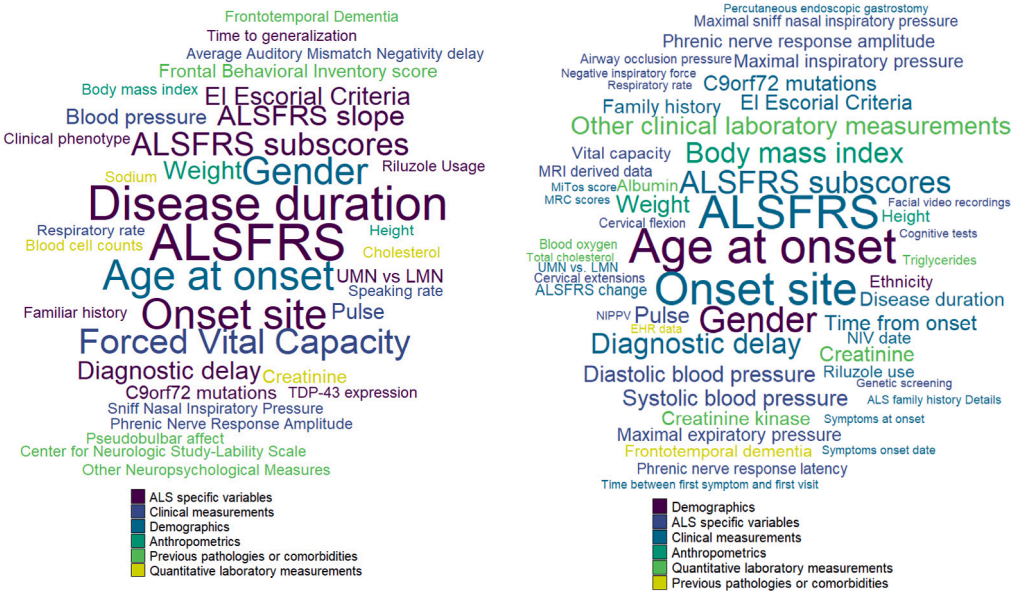
profile-identifying transformations, and strictly predictive models of ALS progression.

4.1.1. Patient stratification: summary and perspectives

As for stratification models, we analysed the input variables, the splitting criteria, and the modelling methodology. The most commonly used variables for stratification found in the literature were: demographics (age at onset or baseline, gender) and anthropometrics (weight, height, BMI); ALS-specific variables, such as functional scores, affected regions at onset and throughout the disease course, family history, treatments, and genetic markers; clinical measurements, mostly related to cardiac and respiratory functions, as well as laboratory measurements; and neuropsychological measures assessing cognition and behavioural changes throughout the disease course.

Splitting criteria were most commonly based on functional decline measures, such as the decrease in ALSFRS score in a given time period, rate of bulbar decline and FVC trajectory. Few studies also performed stratification according to survival measures such as time until death, tracheostomy, or NIV support.

Methods used for stratification included unsupervised clustering for automatic group discovery, dimensionality reduction for uncovering patient profiles and spatial-density-based groups, statistical models



(a) Input variables of stratification models.

(b) Input variables of prediction of progression models.

Fig. 4. Summary of the input variables used by the identified stratification (panel a) and prediction of progression (panel b) models. In each word cloud, the colour of the word corresponds to different categories of input variables, while the size of the word is proportional to the number of studies using that particular input variable. For the definition of acronyms, please see the List of acronyms section. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Summary of the outcomes considered in the identified prediction of progression models. For each specific outcome the table reports the references, the number of papers and the percentage of papers (relative to the total number of prediction of progression papers, i.e., 34) using that specific outcome. For the definition of acronyms, please see the List of acronyms section.

Outcome category	Outcome	Ref.	Counts (%)
ALSFRS or ALSFRS-R	Value at given time points (6, 12, 36 months)	[27], [29], [25], [41]	4 (12%)
	12 months slope of change	[22], [40], [47]	3 (9%)
	3 to 12 months slope	[23], [44], [46]	3 (9%)
ALSFRS subscores (e.g. bulbar, respiratory, motor)	Value at given time points (6, 12, 36 months)	[20], [27], [35]	3 (9%)
Death	Yes/No	[32], [33], [4], [36], [42]	5 (14%)
FVC	Value	[38]	1 (3%)
	Change between 3 to 12 months	[22]	1 (3%)
Need of NIV	After 90 days	[26], [20], [30], [49], [45]	5 (14%)
	After 180 days	[26], [30], [45]	3 (9%)
	After 365 days	[26], [30], [45]	3 (9%)
	Without window	[33], [4]	2 (6%)
Patient score trajectories over time	Evaluated at 12, 24, 36, 42 months	[31], [39]	2 (6%)
Respiratory insufficiency	Defined as initiation of NIV, FVC < 50% of predicted, tracheostomy placement, or death	[32]	1 (3%)
Slow/Fast disease progression	Evaluated at 6 or 12 months	[24]	1 (3%)
Survival	12 months	[21], [23]	2 (6%)
	18 months	[23]	1 (3%)
	24 months	[23]	1 (3%)
	36 months	[37]	1 (3%)
	No window	[40], [48], [50], [52], [34], [43], [28], [51]	9 (26%)
Tracheostomy	Yes/No	[33], [4], [42], [43]	4 (12%)

(e.g., mixed effects, mixture models) for trajectory clustering, models for survival risk, and trajectory models defining groups according to disease stage. Even though deep learning techniques have been applied to electronic health records to uncover clinical patient stratification or trajectories in other neurodegenerative disorders such as Alzheimer's and Parkinson's disease, their use for patient stratification in ALS seems to have been neglected. As regards temporal data usage, most techniques focus on patient trajectory grouping. However, other strategies, combining temporal and static information to uncover patient groups with similar characteristics and disease evolution, might be explored. For example, biclustering and triclustering techniques could be applied to identify specific patterns from the patients' follow-up data. Rigorous

Table 4
Summary of the validation methods used for assessing the identified prediction of progression models. In particular, for each validation type (internal, external or both) and scheme, the references of the papers adopting it are reported, together with their number and percentage (relative to the total number of prediction of progression papers, i.e., 34). For the definition of acronyms, please see the List of acronyms section.

Validation type	Validation scheme	Ref.	Counts (%)
Internal	Holdout set	[28], [22], [37], [39], [44], [24], [46], [47], [48], [49], [25], [52]	12 (35%)
	K-fold cross-validation	[27], [29], [31], [34], [40], [51], [36]	6 (18%)
	Repeated K-fold cross-validation	[26], [20], [30], [33], [35], [45]	7 (21%)
External	External dataset	[21]	1 (3%)
Both	Holdout set External dataset	[23], [41], [42], [43], [50]	5 (15%)
	K-fold cross-validation External dataset	[32], [38]	2 (6%)
	Leave-one-dataset-out cross-validation	[4]	1 (3%)

Table 5
Summary of the performance metrics used for assessing the identified prediction-of-progression models. In particular, the table lists the performance metrics used for each model category. The specific references of the papers using each performance metric are reported in the third column. Finally, the last column reports the number of papers using each performance metric and, in brackets, the percentage of models in the corresponding model category that use that particular performance metric. For the definition of acronyms, please see the List of acronyms section.

Method category	Performance metric	Reference paper	Counts (%)
Regression	RMSE	[22], [23], [36], [38], [40], [44], [41], [46]	8 (100%)
	Pearson's correlation	[22], [23], [44], [46]	4 (50%)
	R-squared	[22], [36], [38]	3 (38%)
Classification	AUROC	[26], [20], [28], [21], [30], [32], [36], [37], [40], [45], [24], [49], [51]	13 (68%)
	Accuracy	[21], [29], [30], [35], [47], [50], [51], [25], [52]	9 (47%)
	Sensitivity	[26], [20], [21], [30], [32]	5 (26%)
	Specificity	[26], [20], [21], [30], [32]	5 (26%)
	MSE	[51], [25], [29]	3 (16%)
	Precision	[21], [32]	2 (11%)
	F1-score	[21], [25]	2 (11%)
	NPV	[32]	1 (5%)
	AUPRC	[28]	1 (5%)
	C-index	[51]	1 (5%)
	Chi-squared test	[32]	1 (5%)
	Hosmer–Lemeshow test	[32]	1 (5%)
Survival	Expected/observed comparison	[42], [43], [48]	3 (43%)
	AUROC	[34], [4], [42]	3 (43%)
	C-index	[33], [23]	2 (29%)
	Precision	[43]	1 (14%)
	NPV	[43]	1 (14%)
	Calibration plot	[4]	1 (14%)
	Calibration slope	[4]	1 (14%)
Probabilistic graph	AUROC	[27], [31]	2 (67%)
	Expected/observed comparison	[31], [39]	2 (67%)
	Accuracy	[27]	1 (33%)
	Sensitivity	[27]	1 (33%)
	MSE	[39]	1 (33%)

validation of stratification groups is missing from the literature. While clustering-based techniques can make use of internal metrics to evaluate robustness, other approaches mostly rely on assessing statistical differences between the resulting groups. Thus, possible routes towards improving the reliability and interpretability of stratification models would be the development of problem-specific metrics, preferably backed by clinical insight and external model validation, e.g., applying the stratification method to different cohorts and evaluating similarity.

After stratification, groups are often compared by means of statistical differences of clinical variables or used to train models of survival time, of time-to-composite-endpoint, or of functional or respiratory decline. An open question is whether stratification may be useful as a preliminary step to train predictive models on separate clusters of subjects, the hypothesis being that focusing on groups of patients with similar characteristics or disease courses could aid in uncovering disease mechanisms and dependencies among variables. This is of particular relevance in the case of diseases with highly heterogeneous presentations such as ALS. However, it must be noted that stratification, in general, leads to a reduction in sample size, which may ultimately render prognostic prediction strategies on some groups unfeasible.

4.1.2. Prediction of progression: summary and perspectives

As for predictive models, we analysed the input variables, the target clinical outcome to be predicted, and the modelling methodology. The

most used input variables included demographics, age at onset, onset site, and ALSFRS (or ALSFRS-R) at baseline. The most common clinical outcomes considered by these models were: the ALSFRS (or ALSFRS-R) score and its decrease over time, the FVC value (crucial to determine the need for NIV), and the occurrence of relevant events related to disease progression, i.e., NIV, FVC < 50%, tracheostomy or death. The most common modelling techniques explored in the literature were: RFs, other tree-based models (e.g., AdaBoost and XGBoost), SVM, LR, the Cox model, and BN models. Only a handful of studies attempted to use deep-learning techniques. The choices of architecture for these neural networks were mainly guided by the nature of the input features, with several models selecting LSTM networks to handle the longitudinal data from patient history, and others combining multiple blocks of fully connected layers to merge features from different sources. Only one model employed convolutional techniques.

The classification framework was the most common modelling choice. Only a few studies framed the problem as a survival analysis to account for time-to-event directly. More in general, even though these models can predict single survival or intervention endpoints, there is a distinct lack of tools able to model the entire disease course over time, considering all the dynamic variables and their relationships. This is also the case for deep-learning models, even though many novel architectures have been proposed in the last few years: from straightforward extensions of the Cox proportional hazard model with non-linear risk [54], to more powerful methods that can model complex non-proportional survival functions and handle multiple adverse events [55,56], which are often present in ALS datasets. Whether this new generation of models can outperform classical methods on ALS survival prediction has yet to be investigated.

Despite the number of published models, few of them are reported together with their parameter values or are externally validated on data not used for training, whereas the validation step would be extremely important to assess model generalisability and, thus, applicability to real-life scenarios. Moreover, since data availability is often an issue within a specific research project given the low prevalence of the disease, models and their parameters should be available so as to allow external validation or recalibration of the models on new populations.

Many of the models proposed in the literature to predict disease progression (as well as to stratify patients) are developed using clinical trial datasets. The development of predictive models using clinical trial data has two important limitations. First, the patients participating in clinical trials are usually selected according to specific inclusion criteria and, thus, are not representative of the general ALS population. Second, clinical trials are generally focused on specific aspects of the disease and, as such, they collect particular variables to test specific hypotheses. The result is that they include predictive variables that are specific to the particular trial's settings and are not commonly collected during routine patient visits. These models are, then, difficult to implement in clinical practice, with real-world data, where access to such complex or specific variables is limited. On the other hand, the inclusion of new environmental and behavioural variables such as pollution data or data from wearables would be of interest. Air quality data should be collected at different granularities, both at a macroscopic level from public stations and at a microscopic level from either portable/wearable sensors or a city grid of low-cost sensors, exploiting a flexible framework that can adapt to the progressive impairments experienced by ALS patients. Integrating continuous and quantitative measurements collected through sensors, both as disease evolution descriptors and as outcomes, might help extending the qualitative monitoring scales currently used in the clinic to assess the functional status of the patients.

5. Conclusion

The inherent promise of the use of AI in medicine and health care is that it will improve the general interpretation of data and related clinical significance with consequent effects on patient care and policy: suggesting diagnosis, prognosis, and treatment; optimising health cost analysis and distribution; defining mitigation policies; and promoting behavioural and societal changes. However, in order to meet these goals, several characteristics that go far beyond the statistical description of a method should be considered [57], describing the key requirements that AI systems should meet in order to be trustworthy. Among others, an appropriate assessment of AI models is perhaps the most important aspect: suitable train/validate/test frameworks should be used to assess model performance and ability to generalise, i.e., the ability to make correct judgments on unseen data in a well-defined domain of validity and in a reproducible way. Even though further considerations about AI applicability go beyond the scope of this work, this review highlighted that this type of information is seldom available in the scientific literature, limiting the impact of the practical application of AI methods in health care.

List of acronyms

AI	artificial intelligence
ALS	amyotrophic lateral sclerosis
ALSFRS	ALS functional rating scale
ALSFRS-R	revised ALS functional rating scale
AUPRC	area under the precision–recall curve
AUROC	area under the receiver operating characteristic curve
BART	Bayesian additive regression tree
BMI	body mass index
BN	Bayesian network
BRAINTEASER	Bringing Artificial Intelligence Home for a Better Care of Amyotrophic Lateral Sclerosis and Multiple Sclerosis
C-index	concordance index
CNN	convolutional neural network
CNS-LS	Center for Neurologic Study-liability scale
DBN	dynamic Bayesian network
DT	decision tree
EEG	electroencephalography
EHR	electronic health record
ENCALS	European network for the cure of ALS
ENPA	neuropsychological examination for aphasia (Italian & esame neuropsicologico per l'afasia)
FBI	frontal behavioural inventory
FTD	frontotemporal dementia
FVC	forced vital capacity
GB	gradient boosting
GLM	generalised linear model
LMN	lower motor neuron
LR	logistic regression
LSTM	long short-term memory
MAE	mean absolute error
MiToS	Milano-Torino ALS staging system
MMN	mismatch negativity
MRC	Medical Research Council
MRI	magnetic resonance imaging
NB	naïve Bayes
NIPPV	nasal intermittent positive pressure ventilation
NIV	non-invasive ventilation
NN	neural networks
NPV	negative predictive value
PBA	pseudobulbar affect
PRO-ACT	pooled resource open-access ALS clinical trials

RAVLT	Rey auditory verbal learning test
RECPAM	recursive partitioning and amalgamation
RF	random forest
RMSE	root mean squared error
RNN	recurrent neural network
SNIP	sniff nasal inspiratory pressure
SVM	support vector machine
UMAP	uniform manifold approximation and projection
UMN	upper motor neuron

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.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.artmed.2023.102588.

References

[1] Van Es MA, Hardiman O, Chiò A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, et al. Amyotrophic lateral sclerosis. Lancet 2017;390(10107):2084–98.

[2] Witzel S, Mayer K, Oeckl P. Biomarkers for amyotrophic lateral sclerosis. Curr Opin Neurol 2022;35(5):699–704.

[3] Xu L, He B, Zhang Y, Chen L, Fan D, Zhan S, et al. Prognostic models for amyotrophic lateral sclerosis: A systematic review. J Neurol 2021;268(9):3361–70.

[4] Westeneng H-J, Debray TP, Visser AE, van Eijk RP, Rooney JP, Calvo A, et al. Prognosis for patients with amyotrophic lateral sclerosis: Development and validation of a personalised prediction model. Lancet Neurol 2018;17(5):423–33.

[5] Wannop K, Bashford J, Wickham A, Iniesta R, Drakakis E, Boutelle M, et al. Fasciculation analysis reveals a novel parameter that correlates with predicted survival in amyotrophic lateral sclerosis. Muscle Nerve 2021;63(3):392–6.

[6] Steinbach R, Gaur N, Roediger A, Mayer TE, Witte OW, Prell T, et al. Disease aggressiveness signatures of amyotrophic lateral sclerosis in white matter tracts revealed by the D50 disease progression model. Hum Brain Mapp 2021;42(3):737–52.

[7] Consonni M, Dalla Bella E, Nigri A, Pinardi C, Demichelis G, Porcu L, et al. Cognitive syndromes and C9orf72 mutation are not related to cerebellar degeneration in amyotrophic lateral sclerosis. Front Neurosci 2019;13:440.

[8] Tortelli R, Copetti M, Ruggieri M, Cortese R, Capozzo R, Leo A, et al. Cerebrospinal fluid neurofilament light chain levels: Marker of progression to generalized amyotrophic lateral sclerosis. Eur J Neurol 2015;22(1):215–8.

[9] Iyer PM, Mohr K, Broderick M, Gavin B, Burke T, Bede P, et al. Mismatch negativity as an indicator of cognitive sub-domain dysfunction in amyotrophic lateral sclerosis. Front Neurol 2017;8:395.

[10] Cykowski MD, Powell SZ, Peterson LE, Appel JW, Rivera AL, Takei H, et al. Clinical significance of TDP-43 neuropathology in amyotrophic lateral sclerosis. J Neuropathol Exper Neurol 2017;76(5):402–13.

[11] Consonni M, Catricala E, Dalla Bella E, Gessa VC, Lauria G, Cappa SF. Beyond the consensus criteria: Multiple cognitive profiles in amyotrophic lateral sclerosis? Cortex 2016;81:162–7.

[12] Watanabe H, Atsuta N, Hirakawa A, Nakamura R, Nakatochi M, Ishigaki S, et al. A rapid functional decline type of amyotrophic lateral sclerosis is linked to low expression of TTN. J Neurol Neurosurg Psychiat 2016;87(8):851–8.

[13] Marin B, Couratier P, Arcuti S, Copetti M, Fontana A, Nicol M, et al. Stratification of ALS patients' survival: A population-based study. J Neurol 2016;263(1):100–11.

[14] Capozzo R, Quaranta VN, Pellegrini F, Fontana A, Copetti M, Carratu P, et al. Sniff nasal inspiratory pressure as a prognostic factor of tracheostomy or death in amyotrophic lateral sclerosis. J Neurol 2015;262(3):593–603.

[15] Rong P, Yunusova Y, Green JR. Speech intelligibility decline in individuals with fast and slow rates of ALS progression. In: Sixteenth annual conference of the international speech communication association. 2015.

[16] Dreger M, Steinbach R, Gaur N, Metzner K, Stubendorff B, Witte OW, et al. Cerebrospinal fluid neurofilament light chain (NFL) predicts disease aggressiveness in amyotrophic lateral sclerosis: An application of the D50 disease progression model. Front Neurosci 2021;15:264.

[17] Grollemund V, Le Chat G, Secchi-Buhour M-S, Delbot F, Pradat-Peyre J-F, Bede P, et al. Manifold learning for amyotrophic lateral sclerosis functional loss assessment. J Neurol 2021;268(3):825–50.

[18] Matos J, Pires S, Aidos H, Gromicho M, Pinto S, Carvalho Md, et al. Unravelling disease presentation patterns in ALS using biclustering for discriminative meta-features discovery. In: International work-conference on bioinformatics and biomedical engineering. Springer; 2020, p. 517–28.

[19] Ackrivo J, Hansen-Flaschen J, Jones BL, Wileyto EP, Schwab RJ, Elman L, et al. Classifying patients with amyotrophic lateral sclerosis by changes in FVC. a group-based trajectory analysis. Am J Respir Crit Care Med 2019;200(12):1513–21.

[20] Pires S, Gromicho M, Pinto S, Carvalho Md, Madeira SC. Patient stratification using clinical and patient profiles: Targeting personalized prognostic prediction in ALS. In: International work-conference on bioinformatics and biomedical engineering. Springer; 2020, p. 529–41.

[21] Grollemund V, Chat GL, Secchi-Buhour M-S, Delbot F, Pradat-Peyre J-F, Bede P, et al. Development and validation of a 1-year survival prognosis estimation model for amyotrophic lateral sclerosis using manifold learning algorithm UMAP. Sci Rep 2020;10(1):1–12.

[22] Tang M, Gao C, Goutman SA, Kalinin A, Mukherjee B, Guan Y, et al. Model-based and model-free techniques for amyotrophic lateral sclerosis diagnostic prediction and patient clustering. Neuroinformatics 2019;17(3):407–21.

[23] Kueffner R, Zach N, Bronfeld M, Norel R, Atassi N, Balagurusamy V, et al. Stratification of amyotrophic lateral sclerosis patients: A crowdsourcing approach. Sci Rep 2019;9(1):1–14.

[24] Gomeni R, Fava M. PRO-ACT consortium, amyotrophic lateral sclerosis disease progression model. Amyotroph Lateral Scler Frontotemporal Degener 2014;15(1–2):119–29.

[25] Halbersberg D, Lerner B. Temporal modeling of deterioration patterns and clustering for disease prediction of ALS patients. In: 2019 18th IEEE international conference on machine learning and applications. ICMLA, IEEE; 2019, p. 62–8.

[26] Martins AS, Gromicho M, Pinto S, de Carvalho M, Madeira SC. Learning prognostic models using disease progression patterns: Predicting the need for non-invasive ventilation in amyotrophic lateral sclerosis. IEEE/ACM Trans Comput Biol Bioinform 2021.

[27] Leão T, Madeira SC, Gromicho M, de Carvalho M, Carvalho AM. Learning dynamic bayesian networks from time-dependent and time-independent data: Unraveling disease progression in amyotrophic lateral sclerosis. J Biomed Inform 2021;117:103730.

[28] Tavazzi E, Daberdaku S, Vasta R, Calvo A, Chiò A, Di Camillo B. Exploiting mutual information for the imputation of static and dynamic mixed-type clinical data with an adaptive k-nearest neighbours approach. BMC Med Inform Decis Mak 2020;20(5):1–23.

[29] Gordon J, Lerner B. Insights into amyotrophic lateral sclerosis from a machine learning perspective. J Clin Med 2019;8(10):1578.

[30] Pires S, Gromicho M, Pinto S, Carvalho M, Madeira SC. Predicting non-invasive ventilation in ALS patients using stratified disease progression groups. In: 2018 IEEE international conference on data mining workshops. ICDMW, IEEE; 2018, p. 748–57.

[31] Zandonà A, Vasta R, Chiò A, Di Camillo B. A dynamic Bayesian network model for the simulation of amyotrophic lateral sclerosis progression. BMC Bioinformatics 2019;20(4):1–11.

[32] Ackrivo J, Hansen-Flaschen J, Wileyto EP, Schwab RJ, Elman L, Kawut SM. Development of a prognostic model of respiratory insufficiency or death in amyotrophic lateral sclerosis. Eur Respir J 2019;53(4).

[33] Swindell WR, Kruse CP, List EO, Berryman DE, Kopchick JJ. ALS blood expression profiling identifies new biomarkers, patient subgroups, and evidence for neutrophilia and hypoxia. J Transl Med 2019;17(1):1–33.

[34] Agosta F, Spinelli E, Riva N, Fontana A, Basaia S, Canu E, et al. Survival prediction models in motor neuron disease. Eur J Neurol 2019;26(9):1143–52.

[35] Bandini A, Green JR, Wang J, Campbell TF, Zinman L, Yunusova Y. Kinematic features of jaw and lips distinguish symptomatic from presymptomatic stages of bulbar decline in amyotrophic lateral sclerosis. J Speech Lang Hearing Res 2018;61(5):1118–29.

[36] Pfohl SR, Kim RB, Coan GS, Mitchell CS. Unraveling the complexity of amyotrophic lateral sclerosis survival prediction. Front Neuroinform 2018;12:36.

[37] Wei Q-Q, Chen Y, Chen X, Cao B, Ou R, Zhang L, et al. Prognostic nomogram associated with longer survival in amyotrophic lateral sclerosis patients. Aging Dis 2018;9(6):965.

[38] Jahandideh S, Taylor AA, Beaulieu D, Keymer M, Meng L, Bian A, et al. Longitudinal modeling to predict vital capacity in amyotrophic lateral sclerosis. Amyotrop Lateral Scler Frontotemporal Degener 2018;19(3–4):294–302.

[39] Zandonà A, Francescon M, Bronfeld M, Calvo A, Chiò A, Di Camillo B. A dynamic bayesian network model for simulation of disease progression in amyotrophic lateral sclerosis patients. PeerJ Preprints; 2017.

[40] Ong M-L, Tan PF, Holbrook JD. Predicting functional decline and survival in amyotrophic lateral sclerosis. PLoS One 2017;12(4):e0174925.

[41] Taylor AA, Fournier C, Polak M, Wang L, Zach N, Keymer M, et al. PRO-AACT consortium, predicting disease progression in amyotrophic lateral sclerosis. Ann Clin Transl Neurol 2016;3(11):866–75.

[42] Lunetta C, Lizio A, Melazzini MG, Maestri E, Sansone VA. Amyotrophic lateral sclerosis survival score (ALS-SS): A simple scoring system for early prediction of patient survival. Amyotroph Lateral Scleros Frontotemporal Degener 2016;17(1–2):93–100.

[43] Elamin M, Bede P, Montuschi A, Pender N, Chio A, Hardiman O. Predicting prognosis in amyotrophic lateral sclerosis: A simple algorithm. J Neurol 2015;262(6):1447–54.
[44] Küffner R, Zach N, Norel R, Hawe J, Schoenfeld D, Wang L, et al. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. Nature Biotechnol 2015;33(1):51–7.
[45] Carreiro AV, Amaral PM, Pinto S, Tomás P, de Carvalho M, Madeira SC. Prognostic models based on patient snapshots and time windows: Predicting disease progression to assisted ventilation in amyotrophic lateral sclerosis. J Biomed Inform 2015;58:133–44.
[46] Hothorn T, Jung HH. RandomForest4Life: A random forest for predicting ALS disease progression. Amyotrop Lateral Scleros Frontotemporal Degener 2014;15(5–6):444–52.
[47] Ko KD, El-Ghazawi T, Kim D, Morizono H. Predicting the severity of motor neuron disease progression using electronic health record data with a cloud computing big data approach. In: 2014 IEEE conference on computational intelligence in bioinformatics and computational biology. IEEE; 2014, p. 1–6.
[48] Scotton WJ, Scott KM, Moore DH, Almedom L, Wijesekera LC, Janssen A, et al. Prognostic categories for amyotrophic lateral sclerosis. Amyotrop Lateral Scleros 2012;13(6):502–8.
[49] Ferreira A, Madeira SC, Gromicho M, Carvalho Md, Vinga S, Carvalho AM. Predictive medicine using interpretable recurrent neural networks. In: International conference on pattern recognition. Springer; 2021, p. 187–202.
[50] Meier JM, van der Burgh HK, Nitert AD, Bede P, de Lange SC, Hardiman O, et al. Connectome-based propagation model in amyotrophic lateral sclerosis. Ann Neurol 2020;87(5):725–38.
[51] Grisan E, Zandonà A, Di Camillo B. Deep convolutional neural network for survival estimation of amyotrophic lateral sclerosis patients. In: ESANN. 2019.
[52] van der Burgh HK, Schmidt R, Westeneng H-J, de Reus MA, van den Berg LH, van den Heuvel MP. Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. NeuroImage: Clin 2017;13:361–9.
[53] Di Camillo B, Vettoretti M, Tavazzi E, Longato E, Roversi C, Trescato I, et al. Artificial intelligence and statistical methods for stratification and prediction of amyotrophic lateral sclerosis. In: PROSPERO: international prospective register of systematic reviews. 2021, https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021288026.
[54] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: Personalized treatment recommender system using a cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18(1).
[55] Lee C, Zame W, Yoon J, der Schaar MV. DeepHit: A deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI conference on artificial intelligence, Vol. 32, no. 1. 2018.
[56] Wang Z, Sun J. SurvTRACE: Transformers for survival analysis with competing events. 2021, arXiv:2110.00855.
[57] Ethics guidelines for trustworthy AI. European commission, futurium. 2018, https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html.