



Multimodal Speech-Based Biomarkers Outperform the ALS Functional Rating Scale in Predicting Individual Disease Progression in ALS

Hardik Kothare¹, Michael Neumann¹, Vikram Ramanarayanan^{1,2}

¹Modality.AI, Inc., USA

²University of California, San Francisco, USA

hardik.kothare@modality.ai

Abstract

Disease progression in ALS is heterogeneous due to the varying presentation of clinical symptoms. This heterogeneity makes it difficult to accurately quantify longitudinal disease severity in people with ALS (pALS), making it difficult to determine the efficacy of therapeutic interventions. In this work, we explore a Bayesian Logistic Mixed-Effects model that can help predict individual trajectories in pALS. We used metrics extracted from 143 pALS who interacted with a cloud-based multimodal assessment platform comprising standard speaking exercises. We found that multimodal biomarkers can be predicted more accurately than the ALSFRS-R, the clinical gold standard to measure disease state, with dense and sparse training data. Such non-linear models have the potential to help with stratification of pALS into fast and slow progressors and thus inform treatment approaches. Patient stratification is also a key factor in designing clinical trials to test drug efficacy in slowing progression.

Index Terms: Amyotrophic Lateral Sclerosis, speech-based multimodal digital biomarkers, disease progression, individual trajectories, non-linear modelling

1. Introduction

Amyotrophic Lateral Sclerosis (ALS), also known as Lou Gehrig’s disease, is a fast-progressing neurodegenerative disease. It has an estimated prevalence of 4.42 per 100,000 persons [1] and a median survival of 3 to 5 years after disease onset [2]. Approximately 30% of patients with ALS present with bulbar onset, which is marked by a rapid decline in speech and swallowing abilities [3]. The rest exhibit non-bulbar onset, initially showing muscular atrophy in the limbs and trunk [4]. However, most patients with non-bulbar onset ALS eventually develop bulbar symptoms as their disease progresses [2]. While almost all pALS decline functionally over time, the nature of this decline and the clinical presentation of the disease is highly heterogeneous [5]. Furthermore, the clinical gold standard to measure disease progression, the revised ALS Functional Rating Scale (ALSFRS-R) [6], is not granular enough to measure subtle changes in function because it is a subjective questionnaire-based assessment. Several studies have modelled ALSFRS-R score progression in a linear manner [7, 8, 9]. Even clinical trials of potential therapies use the linear slope of the ALSFRS-R score or changes from baseline as primary endpoints [10, 11, 12]. However, there is plenty of evidence suggesting that ALS disease progression is non-linear in nature and that progression rate is a function of disease severity [13, 14].

Some prior work has focused on modelling ALS progression using a two-parameter sigmoid [15, 16]. This work assumes pre-specified trajectory patterns through the parameters.

To overcome this requirement, some work has taken an approach of clustering patients sharing progression patterns using Gaussian processes and a Dirichlet process mixture model [17]. But as mentioned earlier, almost all of this work focuses on the ALSFRS-R score and its domain-specific subscores as the primary endpoint of disease progression modelling.

Over the past few years, objective biomarkers based on speech and facial motor movement have emerged as good proxies for tracking bulbar function decline in pALS [18, 19, 20, 21, 22, 23]. A natural question therefore arises: is there a way to incorporate information from this rich resource of multimodal biomarkers to improve forecasting individual progression trajectories in pALS?

To take a step towards this, we take inspiration from prior work done in projecting disease progression in Alzheimer’s disease using multivariate data [24, 25]. This work used the *leaspy* software package that focuses on reconstructing long-term trajectories from short-term observations using Bayesian mixed-effects models. *leaspy* quantifies patient deviations from group averages, assesses the impact of co-factors, imputes missing data, predicts future observations, and simulates virtual patients. The algorithm is capable of handling scalar and multivariate data, including clinical scores and imaging-derived data. Model fits can be logistic, exponential or linear in nature.

In this work, we ask the following research questions:

1. (RQ1) *Disease trajectory prediction:* Given a set of non-uniformly sampled digital biomarkers and the clinical-standard ALSFRS-R score over an 8-week time period, can individual disease progression trajectories be predicted for the 8-week period that immediately follows?
2. (RQ2) *Unseen trajectory imputation:* Given only a set of non-uniformly sampled digital biomarker values over 8 weeks, can the entire unseen clinical-standard ALSFRS-R score trajectory be predicted for a 16-week period?
3. (RQ3a) *Sparse trajectory prediction with uniform sampling:* How well can we predict disease time-course (as in RQ1) for a desired future timepoint when multimodal digital biomarkers and the ALSFRS-R score are sparsely sampled at two time points that are 8 weeks apart?
4. (RQ3b) *Sparse trajectory prediction with non-uniform sampling:* How well can we predict disease time-course for a future timepoint when multimodal digital biomarkers and other clinically-relevant scores (such as perceptual score of speech impairment and the ALSFRS-R score) are sparsely sampled at two non-uniformly sampled timepoints?

RQ1 is important for applications such as ALS randomized clinical trials (RCTs), where one would like to understand and model the disease progression trajectory of a treated patient if they hadn’t received the treatment intervention. Note that such

longitudinal prediction models have to be robust to non-uniform sampling of data points. Another application of RQ1 is in clinical care where clinicians need to decide next steps of action depending on the predicted trajectory.

RQ2 extends the above question to the case of missing data points, an often-encountered case in pharmaceutical RCTs [26]. Specifically, we explore if a data trajectory missing in its entirety (in this case, the clinical-standard ALSFRS-R total score) can be fully reconstructed and predicted for future timepoints, given other objective biomarker trajectories.

Finally, RQ3a and RQ3b extend this case even further to the edge case of just 3 observable time points. It is often the case that certain biomarkers and clinical scales can only be sampled at limited, specified times over the duration of clinical trials for a variety of reasons, including but not limited to cost, the need for manual intervention, or patient burden. For this reason, it can be extremely valuable to be able to robustly predict such cost-inefficient biomarkers based on cost-efficient biomarkers that are more accessible.

To our knowledge, this work is the first to explore non-linear trajectory modelling of remotely-collected and automatically-extracted multimodal biomarkers to predict disease progression in ALS. The cloud-based platform and the completely remote nature of data collection makes this approach more accessible to individuals with ALS.

2. Data

Audiovisual data was collected remotely from 143 pALS (70 female, mean age \pm standard deviation = 60.4 ± 10.2 years) using the Modality platform, a cloud-based multimodal dialogue system for neurological and mental health assessments [27]. Of these 143 pALS, the site of onset was non-bulbar (limb, spinal, etc.) for 107 participants and bulbar for 36 participants. This data collection is part of an ongoing collaboration with EverythingALS and the Peter Cohen Foundation. The study protocol was approved by an external Institutional Review Board and all participants signed an informed consent form prior to participating in the study.

Participants receive a weblink that can be opened in an internet browser and device of their choice. A virtual guide, Tina, walks participants through a structured set of tasks designed to elicit speech and facial behaviours. Tasks include, among others, (a) read speech (Sentence Intelligibility Test (SIT), 5-15 words; Reading Passage (RP), the Bamboo passage 99 words), (b) oral diadochokinesis (DDK, repeating the syllables /pa/, /ta/ and /ka/ rapidly) and (c) free speech in the form of a picture description task (PD). The audiovisual stream captured during participants turns is uploaded to the cloud and segmented in real-time on a turn-by-turn basis. Analytics modules extract relevant speech acoustic, facial motoric and linguistic metrics from each task. 3414 sessions were considered for this analysis across all 143 pALS (2833 non-bulbar, 581 bulbar).

3. Methods

Audiovisual metrics extracted from all speech tasks result in a large number of features. In previous work [23], we have identified 17 objective biomarkers that can distinguish pALS with bulbar onset from those with non-bulbar onset using the following steps. Hierarchical clustering using Spearman's rank-order correlations was applied to healthy controls' data to manage multicollinear features [28]. Ward's method was used for clustering, producing 27 clusters, with a representative feature chosen for each. ROC curve analysis with 5-fold cross-validation

evaluated the features for distinguishing bulbar from non-bulbar onset participants. 17 features with an $AUC \geq 0.65$ and a minimal clinically-important difference [29, 30] larger than the standard error were identified (see Table 1). In addition to these 17 multimodal biomarkers, for a subset of 357 sessions from 119 participants (3 time points per participant, 92 non-bulbar / 27 bulbar), we derived ratings of perceived speech impairment (PSI) from three human raters (two speech scientists and one computer science student) using a visual analogue scale (VAS). Previous research has shown that such perceptual ratings align with severity ratings of speech impairment as rated by clinicians [31, 32]. The VAS was presented to raters as a vertical slider whose position converted to an integer score from 0 (not impaired) to 100 (very impaired). Raters listened to the first 15 seconds of the reading passage to make a judgement.

Participants answered the ALSFRS-R survey at the end of their session. The ALSFRS-R score evaluates physical function of patients across 4 domains - bulbar, fine motor, gross motor and respiratory - through 12 questions. Each question can be scored from 0 (complete loss of function) to 4 (normal function). The ALSFRS-R total score (possible scores ranging from 0 to 48), the ALSFRS-R speech score (scores ranging from 0 to 4), the ALSFRS-R bulbar subscore (comprising responses to three questions about speech, salivation and swallowing; scores ranging from 0 to 12) were also considered for all sessions.

To model trajectories of disease progression, we used Bayesian Logistic Mixed-Effects (BLME) models to predict average and individual trajectories of multivariate data based on repeated observations [25]. The models require all features to be normalized between 0 and 1 with an increase in feature values temporally. The *leaspy* package utilises a Markov Chain Monte Carlo - Stochastic Approximation Expectation Maximization (MCMC-SAEM) algorithm to estimate parameters for the logistic fit. Source dimension was defined as the square root of the number of features rounded off to the closest integer and a Gaussian diagonal noise model was used.

To answer the first research question (RQ1), we took data from all participants who had 16 weeks worth of data from their first session. This resulted in data from 71 pALS (33 female, 56 non-bulbar/15 bulbar) and 851 sessions (average number of sessions per participant \pm standard deviation = 11.99 ± 3.11 sessions). We used a leave-one-out cross validation approach to test the performance of the models in predicting the values in the subsequent 8 weeks. Model performance was evaluated using normalized mean absolute error (nMAE) that accounts for the differences in feature ranges. We evaluated model performance by predicting values of the 17 multimodal biomarkers and ALSFRS-R scores over weeks 8 to 16 given the first 8 weeks worth of data.

To answer the second research question (RQ2), we used values of the 17 multimodal biomarkers from the 851 sessions used in RQ1 to predict the ALSFRS-R scores over all 16 weeks. Note that the model did not see any part of the ALSFRS-R score trajectory during testing.

To answer the third research question (RQ3a), we took data from participants who had completed a session 8 weeks after their first session (56 ± 3 days) and a session at 16 weeks (112 ± 3 days). We only considered these three sessions per participant. This resulted in data from 24 pALS (72 sessions, 12 female, 6 bulbar).

To answer the fourth research question (RQ3b), we used the 17 multimodal digital biomarkers, PSI and the three ALSFRS-R scores from the subset of 357 sessions (for which PSI was rated) as features.

Feature name	Description
Speaking duration (RP)	Time taken to read the passage
cTV (DDK)	Cycle-to-cycle temporal variability during diadochokinesis
CTA (RP)	Canonical Timing Alignment of the reading passage, a number between 0% (non-alignment) and 100% (perfect alignment), measured as the normalised inverse Levenshtein edit distance between words and silence boundaries
Word count (PD)	Count of words used to describe the picture
Max. eyebrow displacement (SIT)	Maximum vertical displacement of eyebrows while reading sentences
PPT (SIT)	Percentage pause time while reading sentences
Max. lip width (RP)	Maximum lip width while reading the passage
HNR (SIT)	Harmonics-to-noise ratio while reading sentences
CPP (RP)	Cepstral peak prominence while reading the passage
HNR (DDK)	Harmonics-to-noise ratio during diadochokinesis
Mean lip aperture (SIT)	Mean aperture of the lips while reading sentences
Max. eye opening (SIT)	Maximum eye opening while reading sentences
Closed class word ratio (PD)	Ratio of closed class words to open class words while describing the picture
Mean F0 (RP)	Mean fundamental frequency while reading the passage
Max. jaw velocity down (SIT)	Maximum downward jaw velocity while reading sentences
DDK syllable count	Number of syllables during diadochokinesis
Max. jaw velocity up (RP)	Maximum upward jaw velocity while reading the passage

Table 1: *Features*; *RP* = reading passage, *DDK* = diadochokinesis, *PD* = picture description, *SIT* = sentence intelligibility test

To benchmark BLME model performance for all four research questions, we fit two naive models - a linear regression model and a multi-layer perceptron (MLP) regressor with three hidden layers (128, 64, 32 neurons), ReLU activation, Adam optimizer, an L2 regularization parameter of 0.0001 and a learning rate of 0.001 [33]. Note that the baseline models for RQ2 were the same as that for RQ1 because the naive regression models would not converge for unseen trajectories.

4. Results

When it came to predicting values of multimodal digital biomarkers and ALSFRS-R scores for an 8-week time period based on a preceding 8-week period (RQ1), the BLME model performed better than or as good as the linear regression and MLP regressor model for all features and scores except closed class word ratio (PD) in bulbar pALS. nMAE was < 0.2 for all features and scores (see Table 2). The best performance was for speaking duration of RP in non-bulbar pALS and mean lip aperture for SIT in bulbar pALS (nMAE = 0.05). Also performing well were predictions for RP CTA, PD word count, maximum vertical eyebrow displacement and HNR during SIT in non-bulbar pALS (nMAE for all four metrics = 0.07). The worst performance was for cycle-to-cycle temporal variability during oral diadochokinesis, maximum upward jaw velocity for the reading passage and the ALSFRS-R total score in bulbar pALS (nMAE = 0.17).

Predicting ALSFRS-R scores for a 16-week period unseen trajectory (RQ2) base on 8 weeks worth of multimodal biomarker data had the least nMAE for the ALSFRS-R speech score in bulbar pALS (0.24). nMAE for the ALSFRS-R total score was higher (see Table 2) than the benchmark models.

When multimodal digital biomarkers and ALSFRS-R scores were sparsely sampled 8 weeks apart, predicting values at a third time point 8 weeks in the future (RQ3a) was more accurate with the BLME than the linear regression model and the MLP regressor. Unsurprisingly, the predictions were not as good as those in RQ1 (denser sampling) but the model did a great job of predicting speaking duration of RP in non-bulbar

pALS (nMAE = 0.06) and PD word count in bulbar pALS (nMAE = 0.05). Notably, prediction of most digital biomarkers was much more accurate than that of the ALSFRS-R scores.

When it came to predicting values of multimodal digital biomarkers, perceived speech impairment and ALSFRS-R scores for the third time point for the 119 participants (RQ3b), the BLME logistic model performed better across the board except for the ALSFRS-R speech score. For RQ3 BLME model, nMAE was < 0.2 for all automatically-extracted multimodal metrics (except PD closed class word ratio in bulbar pALS), PSI and for the ALSFRS-R total score and ALSFRS-R bulbar subscore (see Table 2). We observed the best performance for speaking duration in non-bulbar pALS while reading a passage (nMAE = 0.05). Model performance was the worst in predicting the ALSFRS-R speech score (nMAE = 0.27 for bulbar pALS and 0.26 for non-bulbar pALS).

5. Discussion

The aim of this work was to evaluate whether individual disease progression trajectories in people with ALS can be predicted using non-linear Bayesian logistic mixed effects models. Data was collected remotely, thus making the system more accessible to people with ALS who have severe mobility issues due to loss of motor function.

We observed that when a set of non-uniformly sampled digital biomarkers and ALSFRS-R scores are available (RQ1), individual disease progression trajectories over the next 8 weeks can be predicted well using the BLME model (see Table 2). The ALSFRS-R speech score (range of 4) can be predicted with an nMAE of 0.16 in bulbar pALS which is approximately just over half a point off the observed score. Thus, the BLME approach is useful in predicting disease trajectory in pALS with non-uniformly sampled digital biomarkers. The model performance worsens slightly when the ALSFRS-R scores are unseen and need to be predicted over the course of 16 weeks (RQ2). Thus, the trajectory of ALSFRS-R scores is hard to predict when the data is unseen. This may indicate the subjective non-granular nature of ALSFRS-R scores. When the training data

Feature / Score	Normalized mean absolute error (nMAE) values (Bulbar / Non-bulbar)									
	LR RQ1/RQ2	MLPR RQ1/RQ2	BLME RQ1	BLME RQ2	LR RQ3a	MLPR RQ3a	BLME RQ3a	LR RQ3b	MLPR RQ3b	BLME RQ3b
Speaking duration (RP)	0.16 / 0.13	0.15 / 0.13	0.10 / 0.05	-	0.36 / 0.18	0.56 / 0.37	0.13 / 0.06	0.19 / 0.12	0.21 / 0.12	0.15 / 0.05
cTV (DDK)	0.25 / 0.18	0.25 / 0.20	0.17 / 0.14	-	0.26 / 0.23	1.68 / 2.25	0.17 / 0.15	0.26 / 0.19	0.92 / 0.78	0.18 / 0.14
CTA (RP)	0.25 / 0.19	0.25 / 0.19	0.10 / 0.07	-	0.40 / 0.22	0.65 / 0.77	0.12 / 0.08	0.26 / 0.18	0.29 / 0.22	0.16 / 0.10
Word count (PD)	0.10 / 0.09	0.10 / 0.10	0.06 / 0.07	-	0.15 / 0.10	0.16 / 0.16	0.05 / 0.11	0.17 / 0.15	0.16 / 0.15	0.10 / 0.09
Max. eyebrow disp. (SIT)	0.14 / 0.13	0.14 / 0.14	0.08 / 0.07	-	0.19 / 0.20	0.18 / 0.21	0.14 / 0.22	0.18 / 0.14	0.22 / 0.18	0.17 / 0.13
PPT (SIT)	0.21 / 0.12	0.21 / 0.11	0.13 / 0.10	-	0.41 / 0.18	0.41 / 0.18	0.31 / 0.11	0.20 / 0.16	0.21 / 0.16	0.19 / 0.12
Max. lip width (RP)	0.17 / 0.16	0.19 / 0.18	0.14 / 0.14	-	0.25 / 0.16	0.37 / 0.25	0.17 / 0.11	0.16 / 0.15	0.24 / 0.21	0.15 / 0.13
HNR (SIT)	0.21 / 0.14	0.21 / 0.14	0.08 / 0.07	-	0.28 / 0.19	0.28 / 0.19	0.13 / 0.09	0.19 / 0.14	0.24 / 0.17	0.12 / 0.09
CPP (RP)	0.15 / 0.12	0.15 / 0.13	0.12 / 0.10	-	0.19 / 0.16	0.91 / 0.93	0.11 / 0.15	0.14 / 0.13	0.20 / 0.19	0.10 / 0.10
HNR (DDK)	0.17 / 0.14	0.17 / 0.14	0.10 / 0.08	-	0.23 / 0.19	0.23 / 0.19	0.09 / 0.14	0.19 / 0.14	0.24 / 0.17	0.16 / 0.10
Mean lip aperture (SIT)	0.12 / 0.15	0.14 / 0.15	0.05 / 0.07	-	0.20 / 0.19	0.34 / 0.51	0.10 / 0.16	0.19 / 0.17	0.21 / 0.17	0.16 / 0.16
Max. eye opening (SIT)	0.15 / 0.15	0.15 / 0.16	0.10 / 0.11	-	0.21 / 0.21	2.18 / 2.30	0.21 / 0.16	0.17 / 0.16	0.32 / 0.34	0.14 / 0.13
Closed class word ratio (PD)	0.12 / 0.07	0.12 / 0.07	0.14 / 0.07	-	0.23 / 0.09	0.55 / 0.67	0.25 / 0.09	0.16 / 0.07	0.22 / 0.11	0.22 / 0.06
Mean F0 (RP)	0.17 / 0.17	0.18 / 0.18	0.09 / 0.09	-	0.23 / 0.23	0.70 / 0.64	0.12 / 0.06	0.18 / 0.18	0.20 / 0.18	0.07 / 0.06
Max. jaw velocity down (SIT)	0.16 / 0.16	0.19 / 0.18	0.14 / 0.13	-	0.20 / 0.21	4.56 / 4.55	0.17 / 0.13	0.18 / 0.16	0.27 / 0.37	0.15 / 0.14
DDK syllable count	0.15 / 0.13	0.15 / 0.13	0.13 / 0.09	-	0.21 / 0.18	0.21 / 0.27	0.09 / 0.14	0.13 / 0.15	0.13 / 0.15	0.10 / 0.11
Max. jaw velocity up (RP)	0.16 / 0.17	0.17 / 0.17	0.17 / 0.15	-	0.19 / 0.24	1.05 / 1.11	0.23 / 0.20	0.20 / 0.18	0.49 / 0.60	0.17 / 0.19
Perceived Speech Impairment	-	-	-	-	-	-	-	0.39 / 0.24	0.44 / 0.22	0.11 / 0.09
ALSFRS-R speech score	0.23 / 0.20	0.25 / 0.20	0.16 / 0.11	0.24 / 0.14	0.37 / 0.36	0.35 / 0.37	0.29 / 0.21	0.21 / 0.17	0.22 / 0.26	0.27 / 0.26
ALSFRS-R bulbar subscore	0.20 / 0.15	0.20 / 0.16	0.14 / 0.10	0.19 / 0.12	0.19 / 0.21	0.20 / 0.20	0.22 / 0.18	0.24 / 0.18	0.28 / 0.30	0.17 / 0.14
ALSFRS-R total score	0.18 / 0.18	0.19 / 0.17	0.17 / 0.14	0.23 / 0.19	0.27 / 0.26	0.82 / 1.04	0.33 / 0.36	0.14 / 0.18	0.22 / 0.22	0.14 / 0.16

Table 2: Normalized mean absolute error (nMAE) values; a ‘-’ means the feature was not part of model testing; LR = linear regression model, MLPR = Multi-Layer Perceptron Regressor, BLME = Bayesian Logistic Mixed-Effects model, RP = reading passage, DDK = diadochokinesis, PD = picture description, disp. = displacement, SIT = sentence intelligibility test, cTV = cycle-to-cycle temporal variability, CTA = Canonical Timing Alignment, PPT = percentage pause time, HNR = harmonics-to-noise ratio, CPP = Cepstral Peak Prominence, F0 = Fundamental frequency of voice

for the model is sparse (as in RQ3a and RQ3b), model performance is good for digital biomarkers but worse in predicting the ALSFRS-R speech score as compared to disease trajectory prediction with denser data. Thus, automatically-extracted multimodal biomarkers may be more robust to data sparsity than ALSFRS-R scores.

Speech timing features like speaking duration perform well across the board, i.e. RQ1, RQ3a and RQ3b. This is consistent with past research that shows that speech measures are more responsive and sensitive to change in disease state in pALS [22]. However, the results suggest that orofacial motor movement can be better predicted when the training dataset is dense (8 weeks worth of data).

Predicting speaking duration and word count values with great accuracy in pALS indicates that logistic models are capable of tracking individual trajectories of muscle weakness and the resulting dysarthria due to bulbar degeneration. Such non-linear models have the potential to help with stratification of pALS into fast and slow progressors and thus inform treatment approaches. Patient stratification is also a key factor in designing clinical trials to test the efficacy of drugs in slowing progression. Future work will explore the use of these non-linear models in patient stratification.

We used simple benchmark models in this work to maintain interpretability of their performance evaluation. In future work, we plan to use recurrent neural networks and other deep learning methods to test disease trajectory prediction. Although the BLME algorithm was capable of handling multivariate medical information, we did not include any neuroimaging data or data on neurofilament protein concentrations in participant’s cerebrospinal fluid or blood. Inclusion of such complementary data points may improve model performance but gathering this information requires frequent visits to the clinic and invasive procedures. Distance to clinic has been shown to be a limiter in recruitment of pALS for clinical trials [34]. Future work could explore whether models including remotely-collected multimodal digital biomarkers, which can either be collected non-uniformly and sparsely or uniformly and densely, are capable of predict-

ing disease progression as efficiently as models that combine this information with imaging or lab biomarkers. If the answer to this question is yes, it would greatly improve clinical trial operations and reduce patient burden as well as costs.

6. Conclusion

In conclusion, individual trajectories of disease progression can be predicted with great accuracy in individuals with Amyotrophic Lateral Sclerosis using remotely-collected automatically-extracted multimodal biomarkers and non-linear modelling methods. This work builds on prior work in non-linear modelling of ALSFRS-R decline by demonstrating that in certain cases, changes in automatically-extracted speech acoustic, facial motor and linguistic objective biomarkers could track longitudinal physiological changes more sensitively than changes in categorical scales.

7. Acknowledgement

This work was supported by the National Institutes of Health grant R42DC019877. We thank our collaborators at EverythingALS and the Peter Cohen Foundation for participant recruitment and data collection.

8. References

- [1] L. Xu, T. Liu, L. Liu, X. Yao, L. Chen, D. Fan, S. Zhan, and S. Wang, “Global variation in prevalence and incidence of amyotrophic lateral sclerosis: A systematic review and meta-analysis,” *Journal of Neurology*, vol. 267, pp. 944–953, 2020.
- [2] L. J. Haverkamp, V. Appel, and S. H. Appel, “Natural history of amyotrophic lateral sclerosis in a database population validation of a scoring system and a model for survival prediction,” *Brain*, vol. 118, no. 3, pp. 707–719, 1995.
- [3] J. R. Green, Y. Yunusova, M. S. Kuruville, J. Wang, G. L. Pattee, L. Synhorst, L. Zinman, and J. D. Berry, “Bulbar and speech motor assessment in ALS: Challenges and future directions,” *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14, no. 7-8, pp. 494–500, 2013.

- [4] L. C. Wijesekera and P. Nigel Leigh, "Amyotrophic lateral sclerosis," *Orphanet Journal of Rare Diseases*, vol. 4, pp. 1–22, 2009.
- [5] S. R. Pfohl, R. B. Kim, G. S. Coan, and C. S. Mitchell, "Unraveling the complexity of amyotrophic lateral sclerosis survival prediction," *Frontiers in Neuroinformatics*, vol. 12, p. 36, 2018.
- [6] 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," *Journal of the Neurological Sciences*, vol. 169, no. 1-2, pp. 13–21, 1999.
- [7] M. C. Kiernan, S. Vucic, K. Talbot, C. J. McDermott, O. Hardiman, J. M. Shefner, A. Al-Chalabi, W. Huynh, M. Cudkovicz, P. Talman *et al.*, "Improving clinical trial outcomes in amyotrophic lateral sclerosis," *Nature Reviews Neurology*, vol. 17, no. 2, pp. 104–118, 2021.
- [8] C. Armon, M. C. Graves, D. Moses, D. K. Forté, L. Sepulveda, S. M. Darby, and R. A. Smith, "Linear estimates of disease progression predict survival in patients with amyotrophic lateral sclerosis," *Muscle & Nerve*, vol. 23, no. 6, pp. 874–882, 2000.
- [9] M. Elamin, P. Bede, A. Montuschi, N. Pender, A. Chio, and O. Hardiman, "Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm," *Journal of Neurology*, vol. 262, pp. 1447–1454, 2015.
- [10] S. Paganoni, E. A. Macklin, S. Hendrix, J. D. Berry, M. A. Elliott, S. Maiser, C. Karam, J. B. Caress, M. A. Owegi, A. Quick *et al.*, "Trial of sodium phenylbutyrate–taurursodiol for amyotrophic lateral sclerosis," *New England Journal of Medicine*, vol. 383, no. 10, pp. 919–930, 2020.
- [11] M. E. Cudkovicz, S. Titus, M. Kearney, H. Yu, A. Sherman, D. Schoenfeld, D. Hayden, A. Shui, B. Brooks, R. Conwit *et al.*, "Safety and efficacy of ceftriaxone for amyotrophic lateral sclerosis: a multi-stage, randomised, double-blind, placebo-controlled trial," *The Lancet Neurology*, vol. 13, no. 11, pp. 1083–1091, 2014.
- [12] K. Abe, M. Aoki, S. Tsuji, Y. Itoyama, G. Sobue, M. Togo, C. Hamada, M. Tanaka, M. Akimoto, K. Nakamura *et al.*, "Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial," *The Lancet Neurology*, vol. 16, no. 7, pp. 505–512, 2017.
- [13] P. H. Gordon, B. Cheng, F. Salachas, P.-F. Pradat, G. Bruneteau, P. Corcia, L. Lacomblez, and V. Meininger, "Progression in als is not linear but is curvilinear," *Journal of Neurology*, vol. 257, pp. 1713–1717, 2010.
- [14] N. J. Thakore, B. R. Lapin, E. P. Pioro, and P. R. O.-A. A. C. T. Consortium, "Trajectories of impairment in amyotrophic lateral sclerosis: insights from the pooled resource open-access als clinical trials cohort," *Muscle & Nerve*, vol. 57, no. 6, pp. 937–945, 2018.
- [15] K. Poesen, M. De Schaepdryver, B. Stubendorff, B. Gille, P. Muckova, S. Wandler, T. Prell, T. M. Ringer, H. Rhode, O. Stevens *et al.*, "Neurofilament markers for als correlate with extent of upper and lower motor neuron disease," *Neurology*, vol. 88, no. 24, pp. 2302–2309, 2017.
- [16] R. Steinbach, M. Batyrbekova, N. Gaur, A. Voss, B. Stubendorff, T. E. Mayer, C. Gaser, O. W. Witte, T. Prell, and J. Grosskreutz, "Applying the d50 disease progression model to gray and white matter pathology in amyotrophic lateral sclerosis," *NeuroImage: Clinical*, vol. 25, p. 102094, 2020.
- [17] D. Ramamoorthy, K. Severson, S. Ghosh, K. Sachs, J. D. Glass, C. N. Fournier, P. R. O.-A. A. C. T. C. S. A. 5, T. M. Herrington *et al.*, "Identifying patterns in amyotrophic lateral sclerosis progression from sparse longitudinal data," *Nature Computational Science*, vol. 2, no. 9, pp. 605–616, 2022.
- [18] V. Ramanarayanan, A. C. Lammert, H. P. Rowe, T. F. Quatieri, and J. R. Green, "Speech as a biomarker: Opportunities, interpretability, and challenges," *Perspectives of the ASHA Special Interest Groups*, pp. 1–8, 2022.
- [19] Y. Yunusova, J. R. Green, M. J. Lindstrom, L. J. Ball, G. L. Pattee, and L. Zinman, "Kinematics of disease progression in bulbar ALS," *Journal of Communication Disorders*, vol. 43, no. 1, pp. 6–20, 2010.
- [20] G. M. Stegmann, S. Hahn, J. Liss, J. Shefner, S. Rutkove, K. Shelton, C. J. Duncan, and V. Berisha, "Early Detection and Tracking of Bulbar Changes in Als via Frequent and Remote Speech Analysis," *NPJ Digital Medicine*, vol. 3, no. 1, pp. 1–5, 2020.
- [21] M. Eshghi, Y. Yunusova, K. P. Connaghan, B. J. Perry, M. F. Maffei, J. D. Berry, L. Zinman, S. Kalra, L. Korngut, A. Genge *et al.*, "Rate of speech decline in individuals with amyotrophic lateral sclerosis," *Scientific Reports*, vol. 12, no. 1, p. 15713, 2022.
- [22] H. Kothare, M. Neumann, J. Liscombe, J. Green, and V. Ramanarayanan, "Responsiveness, Sensitivity and Clinical Utility of Timing-Related Speech Biomarkers for Remote Monitoring of ALS Disease Progression," in *Proc. Interspeech*, 2023, pp. 2323–2327.
- [23] M. Neumann, H. Kothare, and V. Ramanarayanan, "Multimodal speech biomarkers for remote monitoring of als disease progression," *Computers in Biology and Medicine*, vol. 180, 2024.
- [24] J.-B. Schiratti, S. Allasonnière, O. Colliot, and S. Durrleman, "A bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations," *Journal of Machine Learning Research*, vol. 18, no. 133, pp. 1–33, 2017.
- [25] E. Maheux, I. Koval, J. Ortholand, C. Birkenbihl, D. Archetti, V. Bouteloup, S. Epelbaum, C. Dufouil, M. Hofmann-Apitius, and S. Durrleman, "Forecasting individual progression trajectories in alzheimer's disease," *Nature Communications*, vol. 14, no. 1, p. 761, 2023.
- [26] A. M. Wood, I. R. White, and S. G. Thompson, "Are missing outcome data adequately handled? a review of published randomized controlled trials in major medical journals," *Clinical trials*, vol. 1, no. 4, pp. 368–376, 2004.
- [27] V. Ramanarayanan, "Multimodal technologies for remote assessment of neurological and mental health," *Journal of Speech, Language, and Hearing Research*, vol. 67, no. 11, pp. 4233–4245, 2024.
- [28] D. Ienco and R. Meo, "Exploration and Reduction of the Feature Space by Hierarchical Clustering," in *Proceedings of the 2008 Siam International Conference on Data Mining*. SIAM, 2008, pp. 577–587.
- [29] K. L. Stipanovic, Y. Yunusova, J. D. Berry, and J. R. Green, "Minimally detectable change and minimal clinically important difference of a decline in sentence intelligibility and speaking rate for individuals with amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 61, no. 11, pp. 2757–2771, 2018.
- [30] H. Kothare, M. Neumann, J. Liscombe, O. Roesler, W. Burke, A. Exner, S. Snyder, A. Cornish, D. Habberstad, D. Pautler *et al.*, "Statistical and clinical utility of multimodal dialogue-based speech and facial metrics for parkinson's disease assessment," in *INTERSPEECH*, 2022, pp. 3658–3662.
- [31] J. E. Sussman and K. Tjaden, "Perceptual Measures of Speech From Individuals With Parkinson's Disease and Multiple Sclerosis: Intelligibility and Beyond," *Journal of Speech, Language, and Hearing Research*, vol. 55, no. 4, pp. 1208–1219, 2012.
- [32] K. L. Stipanovic, K. M. Palmer, H. P. Rowe, Y. Yunusova, J. D. Berry, and J. R. Green, "'you say severe, i say mild': Toward an empirical classification of dysarthria severity," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 12, pp. 4718–4735, 2021.
- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [34] M. Collet, "How much does distance limit the pool of potential clinical trial participants in the united states," *F1000Research*, vol. 10, p. f1000research, 2017.