Novel insights into amyotrophic lateral sclerosis progression through machine learning: analysis of biomarkers and clinical observations in a large-scale patient database

Berke Yilmaz

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a relentless and devastating neurodegenerative disease characterized by the progressive degeneration of motor neurons in the brain and spinal cord. This study aims to enhance the tracking of ALS progression by identifying key predictors of decline using the ALS Functional Rating Scale (ALSFRS) score. Utilizing the comprehensive Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Database, a diverse array of machine learning algorithms is employed, including logistic and LASSO regressions, support vector machines, random forests, gradient boosted trees, explainable boosted machines, extreme gradient boosted trees, and neural network modeling. After data preprocessing, the study analyzed a clean cohort of approximately 6,000 patients and over 400 features, representing the most extensive dataset used in ALS research within the Pro-ACT framework to date. This dataset includes detailed demographics, medication usage, and blood marker information. The Explainable Boosting Machine (EBM) demonstrated superior performance, achieving an AUC of 0.81, accuracy of 0.74, recall of 0.73, and precision of 0.64, with significant (80%) overlap in key features identified across models. A total of 24 biomarkers were identified as playing a role in ALS progression, with Bicarbonate, Creatine Kinase, Creatinine, Chloride, Calcium, and Phosphorus standing out as the most significant. Both the feature importance scores from the Explainable Boosting Machine (EBM) and the Mann-Whitney Test ($p < 0.001$) confirmed the statistical significance of these key biomarkers, validating their critical roles in the analysis of ALS progression.

Key Words: machine learning, ALS, biomarkers, predictive modeling, explainable boosting machine, neurodegenerative disease, ALS Functional Rating Scale (ALSFRS)

1. Introduction

1.1 Background

Amyotrophic Lateral Sclerosis (ALS) is a severe neurodegenerative disorder characterized by the gradual loss of motor neurons in the brain and spinal cord, leading to muscle atrophy, weakness, and eventually paralysis. Despite its relatively low prevalence compared to other neurological disorders, ALS has a significant impact, with a median survival time of 2 to 5 years post-diagnosis (Ceccanti et al., 2020). Currently, approximately 30,000 people in the United States and 230,000 globally suffer from ALS, with numbers expected to increase to approximately 377,000 by 2040 (Arthur et al., 2016).

Current diagnostic methods for ALS rely primarily on clinical neurological evaluation, electromyography, nerve conduction studies, neuroimaging, and cerebrospinal fluid analysis. However, these approaches often fail to provide a definitive diagnosis in the early stages of the disease, leading to delays in treatment initiation (Brotman et al., 2024).

ALS treatment options are limited and primarily focus on managing symptoms and improving quality of life. Medications such as Riluzole and Edaravone have shown some efficacy in slowing disease progression to some extent (Mandrioli et al., 2018; Huang et al., 2024).

Therefore, identification of relevant biomarkers has critical importance in ALS research, offering invaluable insights that may pave the way for the development of innovative biomarker-based approaches to enhance ALS diagnosis, prognosis, and treatment efficacy in the clinical setting. The development of minimally invasive, cost-effective biomarker assays is imperative to facilitate their widespread adoption in clinical practice.

1.2 Pathophysiology of ALS

A striking contrast exists between the neuro-muscular connections of a healthy individual and those affected by Amyotrophic Lateral Sclerosis (ALS). In a healthy system, robust motor neurons bridge the brain and spinal cord to muscles, ensuring the efficient transmission of signals that govern voluntary movements. This intricate communication enables

coordination and dexterity. In contrast, ALS leads to the gradual degeneration and atrophy of these motor neurons, which become withered and fragmented, disrupting their ability to transmit signals to muscle fibers. As a result, muscles do not receive the necessary stimuli to maintain function and mass, leading to atrophy. This degeneration primarily impacts voluntary muscles, leading to difficulties in speaking, swallowing, and ultimately breathing as the disease advances.

The central characteristic of ALS pathology lies in the damage to both upper and lower motor neurons, which are essential for movement. Upper motor neurons, which originate in the brain and travel down to the spinal cord, initiate commands for voluntary movements, while lower motor neurons, which connect the spinal cord to muscles, execute these commands by causing muscle contraction. In ALS, the degeneration of upper motor neurons disrupts movement signals, leading to stiffness (spasticity) and overly brisk reflexes (hyperreflexia). Simultaneously, the deterioration of lower motor neurons weakens muscles, causing shrinkage (atrophy) and involuntary twitches (fasciculations). As more motor neurons lose function and die, muscles progressively weaken, ultimately resulting in paralysis and the loss of voluntary movement. Eventually, the inability to control the muscles needed for breathing leads to respiratory failure, which is the primary cause of death in most ALS patients.

1.3 Current Diagnostic Methods and Treatment Options

Current diagnostic methods for ALS primarily rely on clinical neurological evaluations and electrophysiological testing, such as electromyography, nerve conduction studies, neuroimaging, and cerebrospinal fluid analysis. However, these approaches often struggle to provide a definitive diagnosis in the early stages of the disease, leading to delays in treatment initiation (Iłżecka, 2003). The heterogeneity of ALS phenotypes further complicates diagnosis and prognosis, highlighting the pressing need for more sensitive and specific biomarkers.

In terms of treatment, options for ALS are limited and mainly focus on symptom management and supportive care. While medications like Riluzole and Edaravone have been approved for ALS treatment, their efficacy is modest, and they do not offer a cure. Thakore et al. (2022b) reported that Riluzole improved median survival by 2 months in their study. In a broader analysis, Andrews et al. (2020) reviewed 15 studies and discussed that Riluzole extended median survival by 9 to 16 months in studies with large patient samples, whereas smaller studies showed insignificant results. Similarly, studies by Park et al. (2019), Abe et al.

(2017), and Cho and Shukla (2020) demonstrated that Edaravone can slow ALS progression, although its overall impact remains limited.

1.4 Past Research

In understanding ALS, researchers have highlighted the importance of blood biomarkers as crucial for early detection, monitoring progression, and predicting outcomes. While genomic markers and neuroimaging are valuable, this review specifically focuses on seminal studies of blood biomarkers, aligning with the scope of this research.

Yang et al. (2023) identified eosinophils as potential biomarkers in ALS, demonstrating an inverse correlation with disease progression. Liu et al. (2013) highlighted eosinophil-derived neurotoxin as a potential ALS biomarker, finding significantly higher levels in patients compared to controls. Ong et al. (2017) expanded the biomarker studies by employing predictive models to evaluate functional decline and survival, associating clinical features like weight, alkaline phosphatase, albumin, and creatine kinase levels with disease progression.

The role of creatinine and creatine kinase became the focus of subsequent investigations. Chiò et al. (2014) shed light on the prognostic implications of these biomarkers, particularly creatinine and albumin. Furthermore, Ceccanti et al. (2020) identified a correlation between higher creatine kinase levels and a slower disease progression in a small patient cohort. However, Gao et al. (2022) later discovered that although elevated creatine kinase levels were associated with lower motor neuron denervation, they did not independently predict survival during symptomatic phases.

Hertel et al. (2022) demonstrated a correlation between albumin levels and ALSFRS scores, while Sun et al. (2020) and Gentile et al. (2023) observed that lower albumin levels were associated with faster disease progression.

Turabieh et al. (2023) analyzed the ALS Functional Rating Scale-Revised (FRS-R) slope, highlighting its utility as a robust benchmark for tracking disease progression in ALS. Their models also demonstrated that factors such as days since disease onset and subcomponents of the ALSFRS were predictive of changes in ALSFRS slope.

This overview underscores the complexity of ALS research and the ongoing need to refine methodologies to ensure the reliability and applicability of findings.

1.5 Research Goals

The primary objective of this research is to identify biomarkers predicting progression speed of sporadic ALS using machine learning models. This study evaluates biomarkers' efficacy for predictive and diagnostic purposes to improve early intervention strategies.

The data analyzed for the research is compiled from Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Database. Founded in 2012, PRO-ACT offers the largest anonymized patient data set for ALS research, containing approximately 11,700 patient records and 10 million data points from 29 clinical trials. The PRO-ACT database includes 16 longitudinal data tables with a total of 213 columns, providing a comprehensive foundation for subsequent analyses. The data tables available in the PRO-ACT database include Adverse Events, Forced Vital Capacity (FVC), Slow Vital Capacity (SVC), Laboratory Tests (Labs), ALS History, Vital Signs, Death Data, Demographics, Riluzole, ALS Functional Rating Scale (ALSFRS), El Escorial Criteria, Family History, Hand Grip Strength, Muscle Strength, and Concomitant Medications.

After merging datasets across PRO-ACT database, and applying novel data preparation techniques, the dataset with about 6,000 patients was ready for analysis, one of the most comprehensive ALS datasets in literature.

Due to its comprehensive nature and availability, the ALS Functional Rating Scale (ALSFRS) was chosen as the optimal dependent variable to represent the progression of ALS disease in the machine learning models. The ALSFRS is a tool used by healthcare providers to monitor the progression of disability in ALS patients over time. This scale assesses various functional abilities of the patient, including speech, salivation, swallowing, handwriting, utensil handling, dressing, hygiene, bed mobility, walking, stair climbing, and breathing. It comprises 12 questions, each rated from 0 to 4, where higher scores denote better function. The maximum possible score is 48, indicating full function.

A nominal decrease of 10 points in the ALSFRS score is selected as the threshold for defining 'Fast Progression,' as it reflects a substantial deterioration in a patient's health and functional abilities. An important consideration in defining this decline threshold is also to ensure a balanced representation of both slow and fast progression patients in the dataset, enabling more accurate modeling and analysis.

2. Methods

Figure 1 provides a detailed overview of the research methodology, tracing the systematic steps from initial data collection to complex model optimization. Starting with raw data collection from the PRO-ACT database, it outlines the stages of data preparation, such as data cleaning and defining outcomes, followed by exploratory data analysis. The figure further illustrates the development of a propensity model through data segmentation, feature selection, application of various algorithms, and model selection.

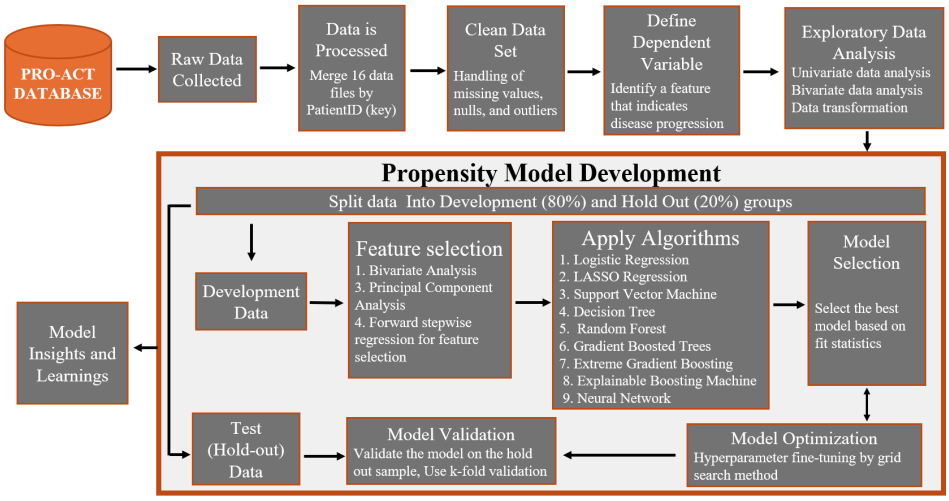

FIGURE 1. A summary of the research procedure broken down into steps.

2.1 Data Pre-Processing

As depicted in Figure 2, the initial data source comprised 16 longitudinal data tables from 11,675 unique patients. Each table was transformed to a wide data format to streamline analysis and modeling tasks. Feature engineering was conducted to create new features, such as tracking blood markers over time, resulting in 428 features.

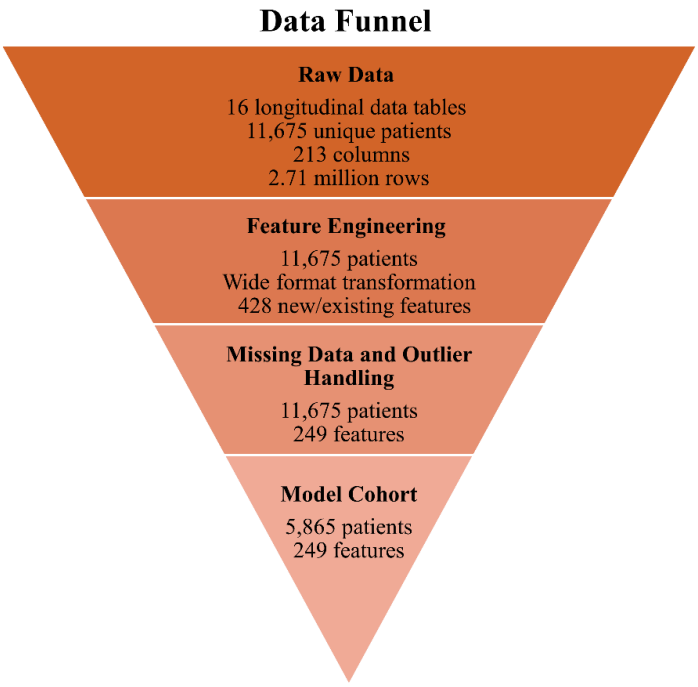

FIGURE 2. Data funnel reduction steps, explaining how the data is prepared for the analysis.

Addressing missing data and outliers was essential to ensure the dataset's integrity. Features with missing values exceeding 30% were systematically pruned to mitigate potential biases and inaccuracies in the analysis. To address missing data within the dataset, an imputation strategy was employed, utilizing methods such as frequency, mode, or median imputation. These techniques were chosen for their effectiveness in preserving data completeness while minimizing the impact of missing values on subsequent analyses. Outliers were managed using percentile-based caps (97th percentile) and floors (3rd percentile) to reduce skewed results and enhance robustness.

2.2 Data Analysis

The data analysis process followed four key steps: univariate analysis, defining the dependent (outcome) variable, bivariate analysis, and dimensionality reduction.

2.2.1 Univariate Analysis

Histograms and boxplots are utilized to visualize the distribution of each independent feature. Additionally, descriptive statistics are computed for all independent features, such as mean, median, standard deviation, range,

minimum, maximum, skewness, and quartiles. These statistics offer a comprehensive overview of the data's central tendency, spread, and variability, facilitating the identification of outliers or anomalies. Moreover, the analysis includes the assessment of the nulls, and missing values for each feature, aiding in data cleaning efforts. Descriptive analytics for a sample of features is presented in Table 1.

Descriptive Statistics									
Feature	Count	Mean	Standard Deviation	Min	25th Percentile	50th Percentile	75th Percentile	Max	Skewness
Abs Eosinophil	3,958	0.48	0.92	0.00	0.10	0.16	0.27	10.08	3.53
Alkaline Phosphatase	4,327	76.0	22.3	22.3	61.2	73	86.8	346.1	1.88
ALSFRS (starting)	5,864	31	4.78	20	28	32	35	41	-0.35
Bicarbonate	4,725	26.2	2.83	17.20	24.25	26.00	28.00	43.00	0.37
Calcium	4,738	2.36	0.10	1.81	2.29	2.36	2.43	3.57	0.39
Creatine Kinase	4,075	296	277	12	119	216	374	4,368	2.53
Eosinophils	3,873	2.33	1.46	0.00	1.37	2.00	2.92	24.00	2.73
Glucose	5,214	5.61	1.61	0.78	4.77	5.27	5.98	28.80	3.96
Hematocrit	5,208	38.8	13.39	0.35	39.77	42.75	45.29	57.00	-2.33
Hematocrit Change	5,208	0.04	2.64	-32.33	-1.30	0.00	1.40	25.20	-0.10
Lymphocyte	4,479	26.2	6.58	7.35	21.64	25.70	30.18	67.00	0.44
Days since onset	5,749	640	401	-4745	-814	-544	-356	-16	0.14
Phosphorus	4,312	1.2	0.13	0.74	1.12	1.20	1.28	1.72	-0.15
Platelets	4,961	242	75	0	205	243	284	622	-0.60
Potassium % Change	5,468	0.44	9.01	-91.33	-4.88	0.00	5.13	70.45	0.02
Red Blood Cells	4,674	2.2E+07	3.2E+08	2.1E-03	4.4E+03	4.7E+03	5.0E+03	5.7E+09	9.59
White Blood Cell Change	4,674	0.5	2.48	-12.47	-0.70	0.10	1.17	37.00	0.91

TABLE 1. Descriptive analytics for univariate analysis for a sample of features.

2.2.2 Model Outcome Definition

Outcome (label, dependent variable) for the machine learning model needs to be a benchmark for the progression of ALS diseases. Based on the Pro-ACT data, several metrics can be utilized to evaluate this progression, including ALSFRS, ALSFRS-R, FVC, SVC, BMI, and Muscle Strength. Selecting the appropriate outcome is crucial for accurately modeling the disease trajectory and ensuring the model's predictive performance.

Using univariate analysis, ALSFRS is identified as the optimal outcome variable due to its comprehensive nature in assessing overall

health and its widespread availability in the dataset. This makes it a robust choice for evaluating ALS progression in the model. Cohorting analysis, shown in Figure 3, which examines the starting FRS score and its decline over 6, 9, and 12 months, reveals heterogeneous data, suggesting varying disease progression rates among patients. Two methodologies for defining the outcome, nominal point decrease vs. percent of score decrease, yield similar results in terms of machine learning model performance and bivariate analysis. Ultimately, a nominal decrease of 10 points in ALSFRS score is chosen as it signifies a significant change in the patient's health and capabilities. An important consideration in defining this decline threshold is to maintain a balanced representation of the classes within the dataset. By employing this specific criterion, approximately 40% of the sample population can be categorized as experiencing "Fast Progression," as illustrated in Figure 4.

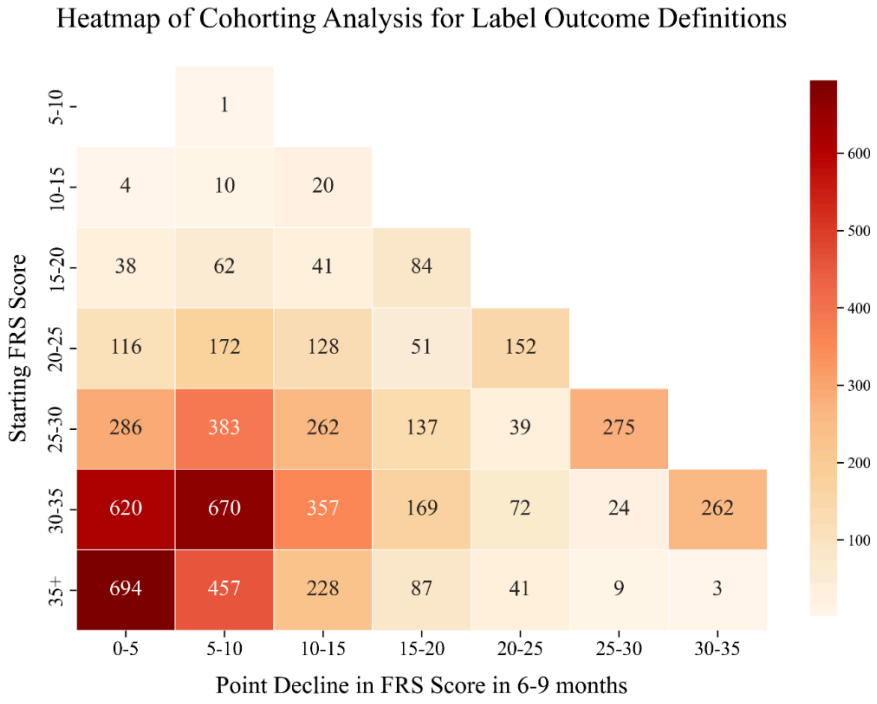

FIGURE 3. Cohort analysis for label outcome definitions.

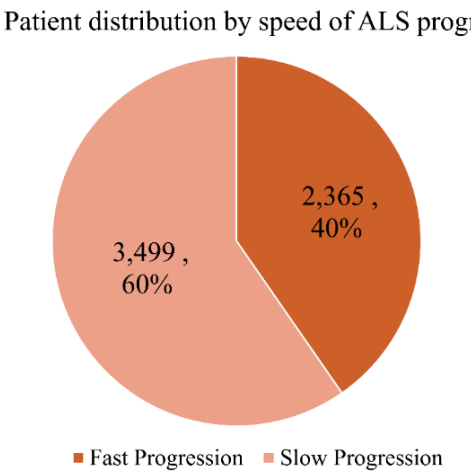

FIGURE 4. Fast vs. Slow Progression sample based on 10-point nominal decline.

2.2.3 Bivariate Analysis

Bivariate analysis is conducted to explore the relationship between the dependent variable and independent features within the dataset. Firstly, data plots such as scatter plots and pair plots were utilized to visually examine the distribution of independent features against the dependent variable. Additionally, cross-tabulation is employed to provide a tabular view of the data pairs, facilitating a comprehensive understanding of the relationship between the dependent feature and independent features.

Furthermore, statistics such as Pearson Correlation, Spearman Correlation, Kendall correlation for continuous features, whereas chi-square and Cramer's V were computed for categorical features. These statistical measures help quantify the strength and significance of the relationships observed. Table 2 and Table 3 display the correlations statistics for a sample of categorical and continuous features.

Correlation Statistics			
Feature	**Pearson**	**Spearman**	**Kendall**
Absolute Basophil Average 6m	-0.038	-0.025	-0.021
Alkaline Phosphatase Level	0.042	0.062	0.050
ALSFRS Score (starting)	-0.204	-0.210	-0.177
Bicarbonate Level	0.137	0.129	0.105

Blood Urea Nitrogen % Change	0.005	0.000	0.000
Chloride Level	-0.131	-0.114	-0.094
Creatine Kinase Level	-0.109	-0.118	-0.097
Creatine Kinase % Change	-0.006	-0.058	-0.048
Creatinine Average 6m	-0.093	-0.124	-0.101
Days since onset	0.263	0.284	0.232
Eosinophils Level	-0.041	-0.039	-0.032
Glucose Level	-0.002	0.034	0.028
Hematocrit Change	0.023	0.033	0.027
Hematocrit Level	0.037	0.037	0.030
Lymphocytes Level	-0.088	-0.085	-0.069
Phosphorus Level	0.175	0.177	0.145
Phosphorus % Change	0.027	0.030	0.025
Platelets Level	0.051	0.049	0.040
Potassium Level	0.008	0.011	0.009
Potassium % Change	-0.008	-0.002	-0.002
Weight	-0.094	-0.099	-0.081
White Blood Cell Change	-0.011	0.004	0.003
White Blood Cell Level	0.040	0.048	0.040

TABLE 2. A sample of bivariate correlation analysis between continuous features and ALSFRS score decline.

Correlation Statistics			
Feature	Chi-Square Test		Cramer's V
	Statistics	p-value	
Bulbar onset	92.40	7.10E-22	0.126
Gastrointestinal issues	77.48	1.34E-18	0.115
Medication - Amitripyline	40.11	2.40E-10	0.083
Medication - Lorazepam	20.38	6.34E-06	0.059
Medication - Tylenol	9.67	1.88E-03	0.041
Renal Issues	8.36	3.84E-03	0.038
Medication - Baclofen	8.05	4.54E-03	0.037
Race Hispanic Latino	4.85	2.77E-02	0.029
Medication - Betacarotene	3.98	4.61E-02	0.026
Medication – Riluzole	5.75	5.64E-02	0.031
Blood Pressure	15.86	3.22E-03	0.052
Diagnosis Bucket	128.96	7.62E-23	0.148
Symptom Onset	428.52	1.82E-88	0.270
Pulse	54.98	6.93E-12	0.097

TABLE 3. A sample of bivariate correlation analysis between categorical features and ALSFRS score decline.

Lastly, correlations between independent features are computed. This step identifies highly correlated features to avoid collinearity, which can negatively impact the performance and interpretability of certain models, such as logistic regression and LASSO regression. Correlation heatmap displaying feature correlations is shown in Figure 5.

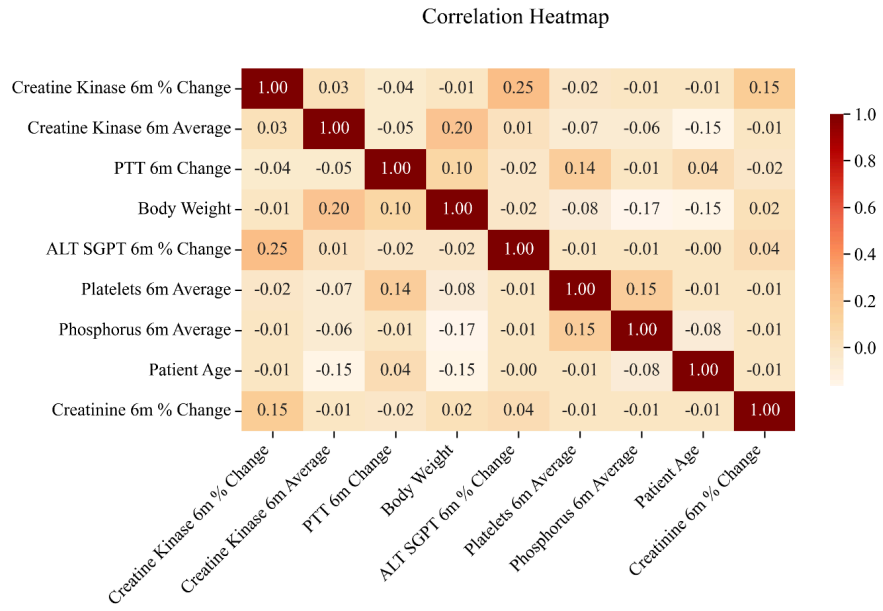

FIGURE 5. Pairwise correlation heatmap for a sample of features studied for the analysis.

2.2.4 Dimensionality Reduction

Dimensionality reduction is essential for enhancing the efficiency of machine learning models by simplifying complex datasets. Principal Component Analysis (PCA) is conducted to reduce dimensionality while retaining essential information. In this study, absolute PCA loadings are used and the top features within each principal component are picked to represent the most diverse set of features. This approach manages model complexity and minimizes collinearity, addressing common issues in model building related to high correlation among predictors. At the conclusion of the data analysis step, the feature set is streamlined to approximately 100 features.

2.3 Machine Learning Algorithms for propensity modeling

In this study, the observation period spans 6 months, while the projection period extends to 9 months, allowing for a comprehensive analysis of disease progression over time. The prediction gap of 3 months helps avoid overlearning from the training data. ALSFRS decline by 10 points is considered as the positive class in the model, with patients having less than 20 ALSFRS points are removed from the cohort to ensure data integrity.

The features identified from PCA and bivariate analysis were further refined using "Forward Stepwise Logistic Regression" to enhance feature selection, resulting in a refined set comprising about 65 features. Age, weight, height, and initial ALSFRS scores serve as control features in the model. Modeling approaches were constructed on these refined features, including Logistic Regression, LASSO Regression, Support Vector Machine, Decision Tree, Random Forest, Gradient Boosted Trees, Extreme Gradient Boosting Trees, Explainable Boosting Machine, and Neural Networks Models.

Feature normalization and scaling were specifically applied to Logistic Regression and Support Vector Machine algorithms to ensure uniformity and optimal performance. Model performance is evaluated using ROC Curve, AUC (Area under the Curve), Accuracy, Precision, and Recall metrics. The models are developed on a randomly selected sample comprising 80% of the dataset, with performance metrics calculated and reported based on the remaining 20% used as test data.

Lastly, model performance optimized through a combination of feature sets and hyperparameter finetuning, particularly using grid search.

3. Results

After the model selection process, performance across different algorithms appears notably similar, with the Explainable Boosting Machine (EBM) showing a distinct performance edge. The test statistics for all employed algorithms are summarized in Table 4. The EBM is selected as the best model, achieving an AUC (Area Under the Curve) of 0.81, an accuracy of 0.74, a precision of 0.64, and a recall of 0.73, outperforming the other models on all metrics. Moreover, EBM is an additive model which allows for a more detailed analysis of feature importance at both aggregate and individual patient levels, enhancing the interpretability and applicability of the results (Nori et al., 2021).

Model Performance Comparison				
Machine Learning Algorithm	**AUC**	**Accuracy**	**Precision**	**Recall**
Logistic Regression	0.79	0.71	0.61	0.72
LASSO Logistic Regression	0.79	0.72	0.61	0.71
Support Vector Machine (SVM)	0.79	0.72	0.62	0.70
Decision Trees	0.60	0.62	0.50	0.50
Random Forest	0.77	0.70	0.59	0.72
Gradient Boosted Trees (GBT)	0.78	0.71	0.61	0.67
Extreme Gradient Boosted Trees (XGBoost)	0.78	0.72	0.62	0.70
Explainable Boosting Machine (EBM)	**0.81**	**0.74**	**0.64**	**0.73**
Neural Network Classifier (Deep Learning)	0.76	0.7	0.62	0.68

TABLE 4. Model Performance Comparison: Performance metrics comparison for all applied machine learning algorithms.

The ROC curve for the EBM, displayed in Figure 6, visually represents the trade-off between sensitivity (recall) and specificity across various thresholds, providing a comprehensive measure of the model's performance.

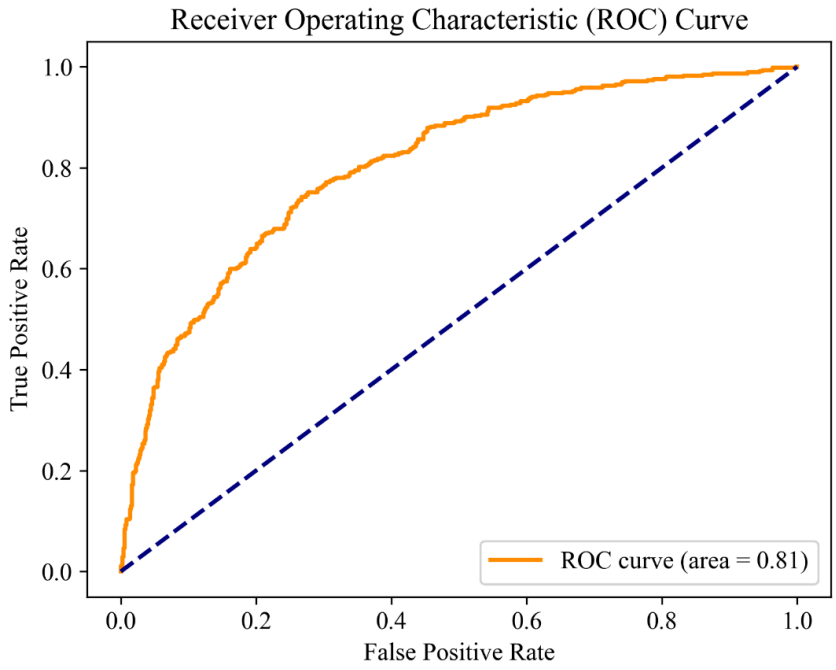

FIGURE 6. ROC curve for the leading model which is Explainable Boosted Machine

The recall metric is particularly crucial due to the nature of the health problem addressed. Recall, or the sensitivity of the model, measures the ability to correctly identify true positives—in this study, patients who are fast progressors of ALS. Given the severe implications of missing the identification of a fast-progressing patient, it is essential to minimize false negatives. To address this issue, the classification threshold was set at 40%, deliberately lower than typical defaults. This adjustment ensures the model is more inclusive in predicting fast progressors, even if it occasionally misclassifies some slow progressors as high-risk. By prioritizing patient safety, this approach minimizes the chance of overlooking high-risk patients and enhances the overall accuracy of the prognosis. The higher recall of 0.73 achieved by the Explainable Boosting Machine (EBM) marks a significant advantage, underscoring the model's utility in practical healthcare applications.

Figure 7 displays the feature importance from the Explainable Boosting Machine (EBM) model, highlighting Bicarbonate, Creatinine, Phosphorus, Chloride, White Blood Cells, and Calcium as key biomarkers in the propensity model. Moreover, 'Days since onset,' 'Illness Onset in the last 12 months,' and the 'Starting ALSFRS score' are identified as the most significant predictors in the model. Furthermore, the first ALS-related physical symptom observed, being bulbar in nature, and the presence of gastrointestinal issues significantly accelerates the progression of the disease. Riluzole, the most prescribed medication for ALS, has a very low feature importance at 0.03, which supports literature findings on the limited efficacy of the treatment.

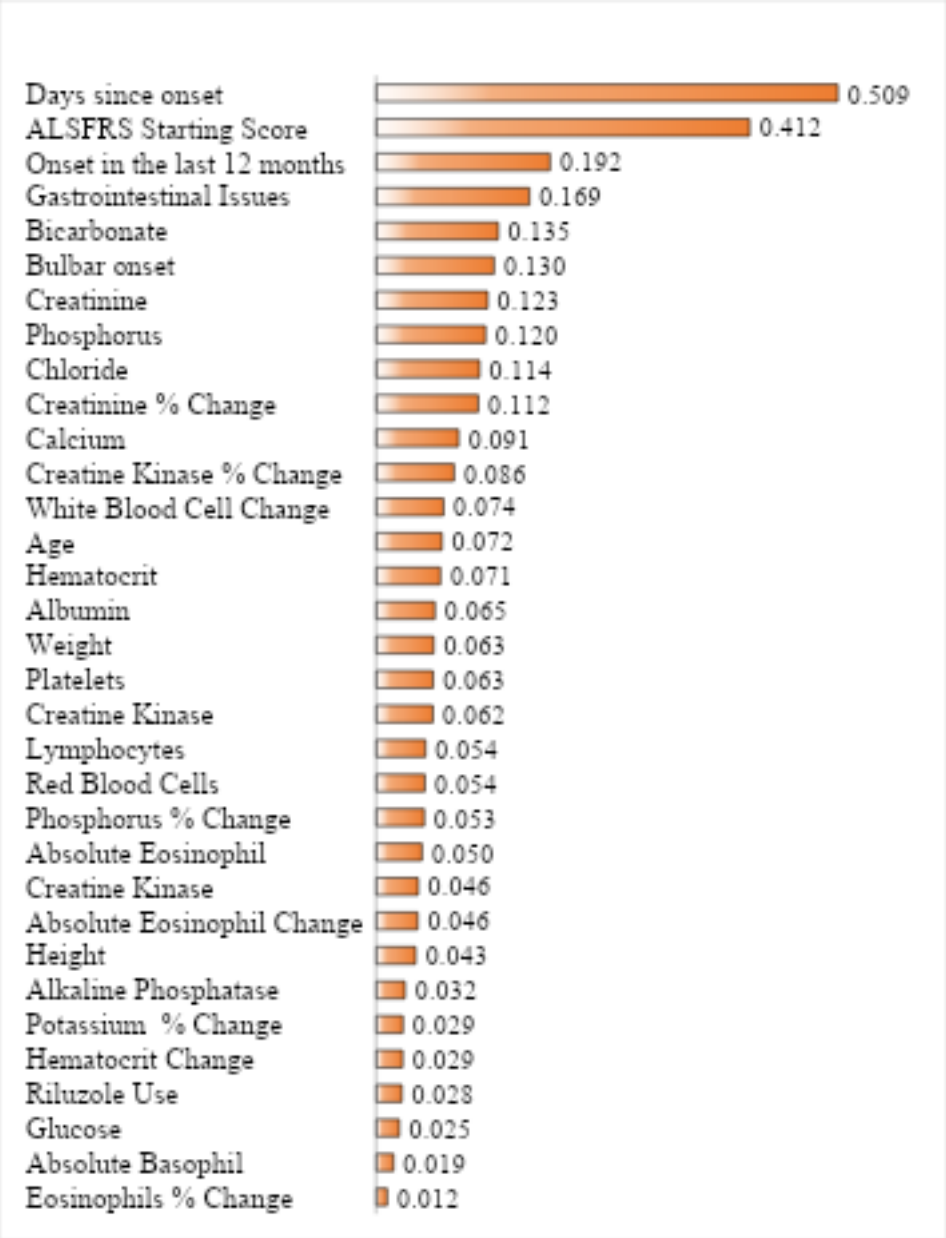

FIGURE 7. Feature importance comparison for leading features from Explainable Boosting Machine.

For the majority of the features, the Mann-Whitney U test revealed p-values below 0.001, highlighting their strong statistical significance, also in the bivariate context. These features, along with their Fold Change values, are detailed in Table 5.

Mann-Whitney U test & Fold Change		
Feature	**p-value**	**Fold Change**
Days since first symptom	1.3E-102	0.721
ALSFRS Starting Score	2.9E-58	0.938
Onset in the last 12 months	1.1E-46	2.002
Phosphorus	3.5E-28	1.026
Bulbar onset	5.3E-22	1.585
Bicarbonate	5.9E-18	1.022
Creatine Kinase	4.8E-17	0.849
Chloride	2.6E-16	0.994
Creatinine	2.9E-15	0.948
Creatine Kinase	8.7E-14	0.878
Weight	4.6E-13	0.965
Age	4.7E-11	1.037
Calcium	2.2E-08	1.005
Lymphocytes	6.2E-08	0.969
Creatinine % Change	9.5E-06	-8.287
Alkaline Phosphatase	3.5E-05	1.021
Albumin	3.8E-	0.992
Creatine Kinase % Change	3.8E-04	0.830
Absolute Eosinophil	1.5E-03	0.838
Platelets	1.5E-03	1.025
Hematocrit	5.4E-03	1.022
Glucose	1.4E-02	1.014
Hematocrit Change	2.1E-02	-6.319
Riluzole Use	3.4E-02	0.974

TABLE 5. Mann-Whitney U test p-values and fold change for significant features from EBM model

The Explainable Boosting Machine (EBM) is a 'glass-box' model, distinct from 'black-box' models such as XGBoost, Neural Networks, and Random Forests, due to its additive nature. As an additive model, EBM systematically combines simple models in a transparent way, allowing each feature's impact on predictions to be clearly understood and quantified.

Figure 6 shows how different features are attributed to the propensity score of two patients. In the cases of Patient 1 and Patient 2, both patients have a very high propensity score for fast progression. The feature comparison highlights that higher levels of chloride and

bicarbonate—indicated by positive values in the model—contribute to accelerating the progression of the disease for both patients.Figure 8. Comparison of model feature attribution for Patient 1 and Patient 2.

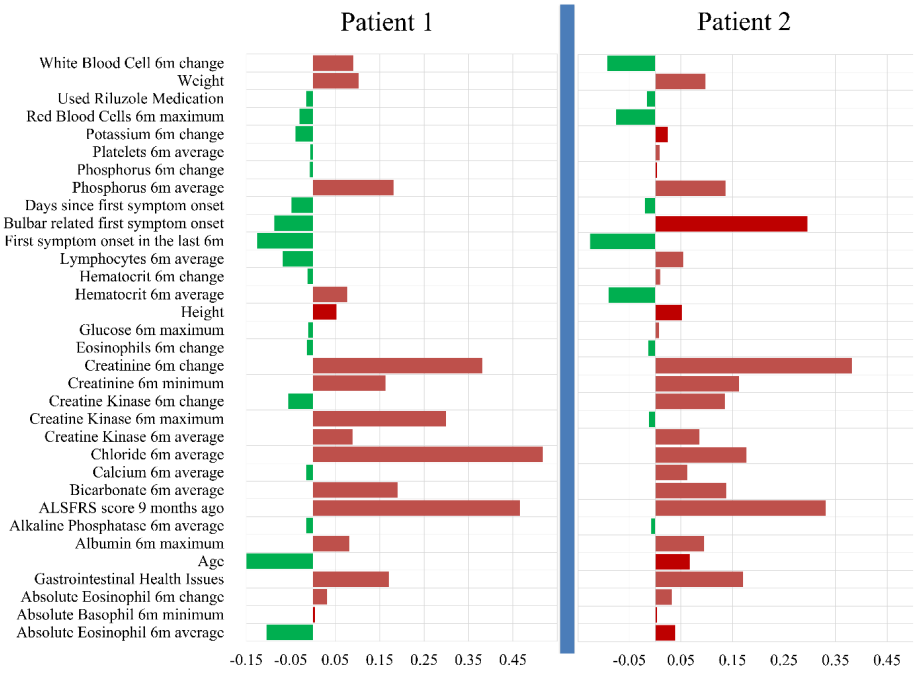

FIGURE 8. Comparison of model feature attribution for Patient 1 and Patient 2.

For Patient 1, higher 6-month averages of absolute eosinophil, calcium, and creatine kinase increase ALS progression speed, while hematocrit and white blood cell changes slow it down. On the other hand, for Patient 2, higher 6-month averages of absolute eosinophil and calcium slow ALS progression, but creatine kinase change speeds it up. Hematocrit and white blood cell changes indicate faster progression.

This transparency provides outputs including patient-level summaries of significant features contributing to the propensity score, which leads to a more nuanced understanding of model decisions. Consequently, EBM serves as an effective tool for creating personalized treatment plans based on detailed patient data. By identifying which factors are most influential in accelerating ALS progression, clinicians can better target interventions to potentially mitigate these negative influences, thereby adapting treatment plans more effectively to individual patient needs by addressing the most impactful predictors for each patient.

7. Discussion

7.1 Summary of Significant Biomarkers Identified

This research aims to enhance our understanding of ALS progression through advanced analytical models. It identifies key biomarkers such as bicarbonate, creatinine, phosphorus, chloride, white blood cells, and calcium as most significant in determining the progression of ALS. Additionally, eosinophil, weight, alkaline phosphatase, albumin, and creatine kinase are recognized as significant, though to a lesser extent, which contrasts with previous studies that identified these as the most significant biomarkers (Chiò et al., 2014; Ong et al., 2017; Yang et al., 2023).

7.2 Comparison with Previous Studies and Deepening Insights Further

White Blood Cells:

The white blood cell count includes neutrophils, lymphocytes, monocytes, eosinophils, and basophils, all of which are vital for the immune response. This study finds that changes in the aggregate white blood cell count, along with specific components such as lymphocytes, absolute eosinophils, changes in eosinophil levels, and absolute basophils, are significant in ALS progression.

Total White Blood Count: This study aligns with Murdock et al. (2017), who found a significant correlation between changes in total white blood cell counts in a study involving 119 participants. However, they noted that it remains unclear whether white blood cell count had a positive or negative effect on ALS progression. This study deepens the understanding of how changes in white blood cell counts impact ALS progression, revealing that significant changes (2 × 10^9/L cells or more) correlate with faster disease progression, whereas more stable counts are associated with slower progression. On the other hand, Cui et al. (2022) observed increases in total white blood cell count over time in ALS patients without an association with disease progression, which contrasts with both this study and Murdock et al. (2017).

Neutrophils: Cui et al. (2022), observed increases in neutrophils over time in ALS patients without an association with disease progression. In contrast, Murdock et al. (2017) and Murdock et al. (2021) findings support the correlation between neutrophils and ALS. Specifically, Murdock et al. (2021) identified a correlation between higher neutrophil counts and shorter survival time in a study of 269 participants. However, they did not

make a claim on whether neutrophil levels had a positive or negative correlation with ALS progression. Although neutrophils does not emerge as a significant biomarker in this study, changes in the overall white blood cell count, which would track changes in neutrophils, are identified as a significant feature.

Eosinophil and Lymphocyte Levels: In the findings of Yang et al. (2023), based on a small cohort study of 59 ALS patients, a statistically significant difference in eosinophil count and lymphocyte percentage was observed. However, their study found only eosinophil count to be correlated with ALS progression, whereas this study identifies both eosinophil count and lymphocyte percentage as significant. Additionally, Murdock et al. (2017) found a significant correlation between changes in CD4 T cells, a type of lymphocyte, and ALS progression. Partial dependence plots from this study reveal a slight positive correlation with faster ALS progression at lower eosinophil levels (0.05-0.184 × 10^9/L cells) and a slight negative correlation at higher levels (2.46-2.73 × 10^9/L cells), although the effect size remains small. Moreover, lymphocyte percentage levels lower than 25% are correlated with increased risk of ALS progression, with the normal range for lymphocytes being 16-33%.

Other Blood Cells Indicators:

Hematocrit: Hematocrit is a measure of the proportion of red blood cells in the blood, helping to assess overall health, particularly the blood's ability to carry oxygen. This study shows that higher hematocrit levels and increases in hematocrit are correlated with a higher risk of fast ALS progression. Mandrioli et al. (2017)'s findings align with this study, noting that hematocrit is directly associated with the odds of survival or tracheostomy. Additionally, red blood cell count is identified as a medium feature importance predictor for ALS progression in this study.

Platelets: Platelets are small blood cells that play a crucial role in blood clotting and wound healing. In this study, platelets rank as medium significance in ALS progression prediction. Although there is limited research directly linking platelet levels in the blood with ALS, few studies have explored the correlation between platelet serotonin levels, platelet malfunction, and ALS (Dupuis et al., 2010; Leiter and Walker, 2020).

Electrolytes:

Electrolytes are minerals in the body that carry an electric charge and are essential for various physiological functions, including fluid balance, nerve signaling, muscle contractions, and maintaining acid-base balance. Out-of-normal values of electrolytes can indicate various illnesses or health conditions, making their monitoring important for overall health. In

this study, chloride, bicarbonate, and potassium are identified as significant electrolytes impacting ALS progression.

Chloride and Bicarbonate: Chloride and bicarbonate levels are associated also with acid-base (pH) imbalance. This study corroborates previous findings from Stambler et al. (1998), Qureshi et al. (2008), and Manera et al. (2023), which indicate that lower chloride levels are associated with faster disease progression and shorter survival time in ALS patients. Ong et al. (2017) further supports this by showing that decreasing chloride levels correlate with a higher risk of death. However, in this study, the change in chloride levels was not statistically significant in the machine learning model, even though the chloride level itself is a significant factor.

Qureshi et al. (2008) also demonstrated that higher bicarbonate levels correlate with shorter survival times in ALS patients. Additionally, Hadjikoutis and Wiles (2001) indicated that high serum bicarbonate and low chloride levels are metabolic indicators of chronic respiratory acidosis, which is associated with respiratory muscle weakness in patients with Motor Neuron Diseases (MND). Ong et al. (2017) expanded on this by suggesting that patients with higher bicarbonate and lower chloride levels are more likely to experience severe respiratory issues, with respiratory failure being a more likely cause of death.

This study enhances the current understanding of ALS progression by identifying a critical bicarbonate level of 26 mmol/L, above which the risk of fast ALS progression increases, compared to the normal bicarbonate range of 18-23 mmol/L. Additionally, chloride levels below 101 mmol/L are associated with faster ALS progression, with the normal range for chloride being 98-106 mmol/L. The study further shows that bicarbonate levels are monotonically positively correlated, and chloride levels are monotonically negatively correlated, with faster ALS progression. These findings underscore the importance of closely monitoring these electrolyte levels in ALS patients to better understand and potentially manage disease progression.

Potassium: This study indicates that a decrease in potassium levels correlates with better outcomes, which aligns with the findings of Gentile et al. (2023). Their study of 836 patients demonstrated a negative correlation between potassium levels and changes in ALSFRS-R scores. However, Sun et al. (2020) contrast with this study, as they did not find a clear association between potassium levels and ALS mortality risk.

Sodium: While Manera et al. (2023) found a significant correlation between serum sodium levels and ALSFRS-R, this research did not

identify sodium as a significant factor in ALS progression. This suggests that although sodium levels may be correlated with the ALSFRS-R score, they are not predictive of score decline or disease progression. Similarly, Sun et al. (2020) did not find a clear association between sodium levels and ALS mortality risk, aligning with this study.

Mineral Balance:

Mineral balance, particularly involving calcium and phosphorus, is critical for various physiological functions, including those that support neurological health. This study shows that calcium and phosphorus levels are related to ALS progression.

Calcium: Calcium is essential for neurons, and research has shown a connection between calcium imbalance and neurodegenerative processes in amyotrophic lateral sclerosis (ALS) (Katzeff et al., 2020). Most research in ALS literature has focused on intracellular calcium, with limited studies on serum calcium. This study finds that higher calcium levels, specifically above 2.3 mmol/L, where normal range is 2.2-2.5 mmol/L, are associated with faster ALS progression. However, Sun et al. (2020) contrast with this study, as they did not find a clear association between serum calcium levels and ALS progression.

Phosphorus: Although not widely studied in ALS literature, higher phosphorus levels are found to accelerate ALS progression in this research. Gordon and Lerner (2019) showed that phosphorus levels correlate with ALSFRS scores, and this study extends their findings by demonstrating that phosphorus levels play a significant role in ALS progression.

Metabolites and Other Biochemical Markers:

Albumin: The relationship between albumin and ALS progression has been explored by Sun et al. (2020), Hertel et al. (2022), and Gentile et al. (2023). While Hertel et al. (2022) observed a correlation between albumin levels and ALSFRS scores, they did not find a significant impact on disease progression. In contrast, both Sun et al. (2020) and Gentile et al. (2023) observed that lower albumin levels were associated with shorter survival. Similarly, this research confirms a similar association between lower albumin levels and faster ALS progression in a larger patient cohort, analyzed within the context of a much broader set of features. This study enriches the understanding of albumin's role, demonstrating a more nuanced relationship between albumin levels and ALS progression. The analysis reveals that the risk of disease progression increases until albumin levels reach 40 g/L, beyond which the risk decreases, suggesting that

while albumin is beneficial, its impact varies across different levels. The normal range of albumin is 35-50 g/L.

Creatinine: This study corroborates the findings of Sun et al. (2020), who studied 399 patients, and Gentile et al. (2023), who analyzed 836 patients. Both studies identified a strong association between lower creatinine levels and increased ALS progression, noting a link to higher mortality risk. Specifically, this study demonstrates that creatinine levels below 60 mmol/L are associated with accelerated ALS progression, indicating a worse prognosis. The normal range for creatinine levels is 53-106 mmol/L. Additionally, the study reveals a significant correlation between changes in creatinine levels and ALS progression: a more than 12% increase in creatinine levels within the cohort is linked to a higher risk of progression and a sharper decline in ALSFRS scores over the six-month observation period.

Creatine Kinase (CK): This study finds that significant fluctuations in CK levels—specifically, a decline of 50% or more or an increase of 70% or more—are strongly associated with ALS progression. This contrasts with the findings of Gentile et al. (2023), who observed a positive correlation between CK levels and ALSFRS-R scores at the time of diagnosis but did not find any correlation between CK levels and the rate of disease progression.

Glucose: Sun et al. (2020) identified both higher blood glucose levels and increasing glucose levels as being related to a higher mortality risk in ALS patients. In a related vein, Gray et al. (2015) and Wuolikainen et al. (2016) observed elevated glucose levels in the cerebrospinal fluid of ALS patients. This study also finds that higher blood glucose levels correlate with faster disease progression, although the feature importance of glucose is much lower compared to other biomarkers.

HDL Levels: Mixed findings are reported regarding the impact of a patient's lipid profile on ALS prognosis. For example, Rafiq et al. (2015) found no relationship between lipid profile and ALS prognosis, whereas Hertel et al. (2022) identified HDL levels as significant. However, this study does not support the significance of HDL. It's also important to note that Hertel et al. (2022) reported borderline significance for HDL, based on a cohort of 1,084 ALS patients, with a p-value of 0.044. This borderline significance might explain why HDL levels are not showing up as significant in this study, as more significant features may have greater feature importance in explaining ALS progression.

Non-Blood Marker Factors:

Bulbar Onset: Christensen et al. (1990), Testa et al. (2004), Elamin et al. (2015), Daghlas et al. (2017), and Pudasaini et al. (2022), all demonstrated that bulbar onset is correlated with a worse ALS prognosis. This study also confirms this finding.

Age: This study corroborates the findings of Preux et al. (1996), Louwerse et al. (1997), Christensen et al. (1990), Stambler et al. (1998), and Testa et al. (2004), all of which emphasize the significance of age on ALS prognosis. Specifically, older age is correlated with faster disease progression and shorter survival time. Based on the cohort studied in this research, patients over 50 are more likely to have faster ALS progression.

Days since disease on set: Patients who experience disease onset within the last 12 months exhibit faster disease progression, as indicated by this study. These patients also have shorter times from onset to diagnosis, suggesting that their symptoms were prominent enough to prompt quicker medical attention compared to those with longer diagnostic timelines. Stambler et al. (1998) and Testa et al. (2004) similarly demonstrated that a shorter time from onset to diagnosis is a predictor of survival, aligning with this study's findings. Moreover, Turabieh et al. (2023) emphasized the importance of factors such as days since disease onset in predicting changes in the ALS Functional Rating Scale-Revised (ALSFRS-R) slope, further supporting this study's conclusions.

7.3 Evaluating Blood Marker Correlations with ALSFRS Scores: A Focus Beyond Decline

Gordon and Lerner (2019) studied the Pro-ACT data with a sample of 3,772 patients, examining the correlation between blood markers and ALSFRS scores. Their findings indicated that chloride, alkaline phosphatase, phosphorus, CK, and creatinine have higher feature importance in determining the ALSFRS score. This study extends Gordon et al.'s work by analyzing the change in ALSFRS scores, providing insights into indicators of faster decline in ALS health outcomes. The findings of this study show that Gordon et al.'s identified blood markers are also applicable to score changes. Additionally, this study highlights other highly significant factors, such as bicarbonate, red blood cell count, white blood cell count, and changes in certain blood markers over the observation period, as important indicators of changes in ALSFRS scores which are not part of significant features of the ALSFRS score correlation in Gordon and Lerner (2019) study.

7.4 Advancing ALS Biomarker Research: Extensive Data and Methodological Strengths

This study stands out by employing the Explainable Boosting Machine (EBM) model, a cutting-edge algorithm that provides transparent insights into how each feature influences ALS progression. EBM identifies non-linear relationships and interactions, making it superior for understanding the complex dynamics at play and facilitating the development of tailored treatment options at the patient level, all while maintaining transparency.

This research distinguishes itself from previous studies by leveraging one of the most comprehensive datasets available in the field, allowing for an analysis that incorporates a significantly larger number of features compared to prior efforts. Many previous studies have produced conflicting results due to their reliance on different patient datasets with limited scope, often identifying only a few significant factors in isolation, constrained by smaller sample sizes and a narrow range of features. This limitation in scope frequently prevents a nuanced understanding of the complex interactions among various biological markers and their impact on disease progression.

In contrast, the extensive dataset of 6,000 patients utilized in this study captures both the heterogeneity of ALS patients and the complexity of relevant features, offering a more nuanced understanding of the disease. By incorporating a wide array of features (400+) achieved through intensive data cleaning techniques and the creation of new longitudinal features, and analyzing them collectively, this study uncovers additional significant biomarkers that may have been overlooked in earlier research. The findings are also more comprehensive, as the analysis identified many more significant features within the same model, accounting for all interactions and dependencies between them, providing a deeper understanding of disease progression. The use of advanced statistical techniques and machine learning models allows for the capture of intricate dynamics within the data, further enhancing the study's predictive accuracy.

7.5 Limitations

These findings should be interpreted within the context of several limitations. A major challenge was the high percentage of missing values in the PRO-ACT dataset, necessitating the exclusion of numerous biomarkers and thereby limiting the depth of our analysis. Employing a subset of patients with more complete data may reveal additional

significant features. Furthermore, the intermittent data collection and the dataset's cutoff at 2022 limited our capacity to observe recent trends in ALS progression or treatment responses. The models will need retraining when new data becomes available to stay current. Lastly, traditional machine learning models do not consider confounding variables. To confirm causal relationships between significant features and ALS progression, future studies should employ causality-based models using advanced techniques such as Double ML, quasi-experimental methods, or validate the findings through clinical trials.

7. Conclusion

This research successfully navigates its constraints to chart promising directions for ALS research. The machine learning models deployed demonstrated performance metrics that stand out in the PRO-ACT database literature, with an expanded list of significant biomarkers identified. The EBM model exhibited superior performance compared to other algorithms employed in the study. Notably, the identification of 24 blood-marker-based features correlated with ALS progression marks a significant advancement in the field. Among the most impactful drivers identified are Bicarbonate, Creatine Kinase, Creatinine, Chloride, Calcium, and Phosphorus.

Moreover, the EBM model's transparency offers unparalleled interpretability by providing clear, additive analyses of how each feature influences disease progression. This interpretability enables more personalized treatment insights, surpassing previous studies reliant on smaller clinical samples or less interpretable black-box models.

By leveraging a comprehensive dataset that captures both the heterogeneity of ALS patients and the complexity of relevant features, this study provides a more robust foundation for understanding ALS progression. The use of advanced machine learning models ensures that the findings are reliable, offering a valuable basis for developing targeted interventions and enhancing future ALS research.

Acknowledgements

Data used in the preparation of this article was obtained from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Database. The following organizations and individuals within the PRO-ACT Consortium contributed to the design and implementation of the PRO-ACT Database and/or provided data but did not participate in the analysis of the data or the writing of this report: ALS Therapy Alliance, Cytokinetics, Inc.,

Amylyx Pharmaceuticals, Inc., Knopp Biosciences, Neuraltus Pharmaceuticals, Inc., Neurological Clinical Research Institute, Massachusetts General Hospital, Northeast ALS Consortium, Novartis, Prize4Life Israel, Regeneron Pharmaceuticals, Inc., Sanofi, Teva Pharmaceutical Industries, Ltd., and The ALS Association.

References

Abe, K., Aoki, M., Tsuji, S., Itoyama, Y., Sobue, G., Togo, M., Hamada, C., Tanaka, M., Akimoto, M., Nakamura, K., Takahashi, F., Kondo, K., Yoshino, H., Abe, K., Aoki, M., Tsuji, S., Itoyama, Y., Sobue, G., Togo, M., . . . Yoshino, H. (2017). Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *The Lancet Neurology,* 16(7), 505–512. https://doi.org/10.1016/s1474-4422(17)30115-1

Andrews, J. A., Jackson, C. E., Heiman-Patterson, T. D., Bettica, P., Brooks, B. R., & Pioro, E. P. (2020). Real-world evidence of riluzole effectiveness in treating amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 21(7–8), 509–518. https://doi.org/10.1080/21678421.2020.1771734

Arthur, K. C., Calvo, A., Price, T. R., Geiger, J. T., Chiò, A., & Traynor, B. J. (2016). Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. *Nature Communications*, 7(1). https://doi.org/10.1038/ncomms12408

Brotman, R. G., Moreno-Escobar, M. C., Joseph, J., Munakomi, S., & Pawar, G. (2024, February 12). *Amyotrophic lateral sclerosis*. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK556151/

Ceccanti, M., Pozzilli, V., Cambieri, C., Libonati, L., Onesti, E., Frasca, V., Fiorini, I., Petrucci, A., Garibaldi, M., Palma, E., Bendotti, C., Fabbrizio, P., Trolese, M. C., Nardo, G., & Inghilleri, M. (2020). Creatine Kinase and Progression Rate in Amyotrophic Lateral Sclerosis. *Cells*, 9(5), 1174. https://doi.org/10.3390/cells9051174

Chiò, A., Calvo, A., Bovio, G., Canosa, A., Bertuzzo, D., Galmozzi, F., Cugnasco, P., Clerico, M., De Mercanti, S., Bersano, E., Cammarosano, S., Ilardi, A., Manera, U., Moglia, C., Sideri, R., Marinou, K., Bottacchi, E., Pisano, F., Cantello, R., Mazzini L., Mora, G. (2014). Amyotrophic lateral sclerosis outcome measures and the

role of albumin and creatinine. *JAMA Neurology*, 71(9), 1134. https://doi.org/10.1001/jamaneurol.2014.1129

Chiò, A., Logroscino, G., Hardiman, O., Swingler, R., Mitchell, D., Beghi, E., Traynor, B. G., & Couratier, P. (2009b). Prognostic factors in ALS: A critical review. *Amyotrophic Lateral Sclerosis,* 10(5–6), 310–323. https://doi.org/10.3109/17482960802566824

Cho, H., & Shukla, S. (2020). Role of Edaravone as a Treatment Option for Patients with Amyotrophic Lateral Sclerosis. *Pharmaceuticals*, 14(1), 29. https://doi.org/10.3390/ph14010029

Christensen, P. B., Højer-Pedersen, E., & Jensen, N. B., MD. (1990). Survival of patients with amyotrophic lateral sclerosis in 2 Danish counties. *Neurology*, 40(4), 600. https://doi.org/10.1212/wnl.40.4.600

Cui, C., Ingre, C., Yin, L., Li, X., Andersson, J., Seitz, C., Ruffin, N., Pawitan, Y., Piehl, F., & Fang, F. (2022). Correlation between leukocyte phenotypes and prognosis of amyotrophic lateral sclerosis. *eLife*, 11. https://doi.org/10.7554/elife.74065

Daghlas, I., Lever, T. E., & Leary, E. (2017). A retrospective investigation of the relationship between baseline covariates and rate of ALSFRS-R decline in ALS clinical trials. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 19(3–4), 206–211. https://doi.org/10.1080/21678421.2017.1418001

Dupuis, L., Spreux-Varoquaux, O., Bensimon, G., Jullien, P., Lacomblez, L., Salachas, F., Bruneteau, G., Pradat, P., Loeffler, J., & Meininger, V. (2010). Platelet serotonin level predicts survival in amyotrophic lateral sclerosis. *PLoS ONE*, 5(10), e13346. https://doi.org/10.1371/journal.pone.0013346

Elamin, M., Bede, P., Montuschi, A., Pender, N., Chio, A., & Hardiman, O. (2015). Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm. *Journal of Neurology*, 262(6), 1447–1454. https://doi.org/10.1007/s00415-015-7731-6

Gao, J., Dharmadasa, T., Malaspina, A., Shaw, P. J., Talbot, K., Turner, M. R., & Thompson, A. G. (2022). Creatine kinase and prognosis in amyotrophic lateral sclerosis: a literature review and multi-centre cohort analysis. *Journal of Neurology*, 269(10), 5395–5404. https://doi.org/10.1007/s00415-022-11195-8

Gentile, F., Maranzano, A., Verde, F., Bettoni, V., Colombo, E., Doretti, A., Olivero, M., Scheveger, F., Colombrita, C., Bulgarelli, I., Spinelli, E. G., Torresani, E., Messina, S., Maderna, L., Agosta, F., Morelli, C., Filippi, M., Silani, V., & Ticozzi, N. (2023). The value of routine blood work-up in clinical stratification and prognosis of patients with

amyotrophic lateral sclerosis. *Journal of Neurology*, 271(2), 794–803. https://doi.org/10.1007/s00415-023-12015-3

Gordon, J., & Lerner, B. (2019). Insights into Amyotrophic Lateral Sclerosis from a Machine Learning Perspective. *Journal of Clinical Medicine*, 8(10), 1578. https://doi.org/10.3390/jcm8101578

Gray, E., Larkin, J. R., Claridge, T. D. W., Talbot, K., Sibson, N. R., & Turner, M. R. (2015). The longitudinal cerebrospinal fluid metabolomic profile of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 16(7–8), 456–463. https://doi.org/10.3109/21678421.2015.1053490

Hadjikoutis, S., & Wiles, C. M. (2001). Venous serum chloride and bicarbonate measurements in the evaluation of respiratory function in motor neuron disease. *QJM*, 94(9), 491–495. https://doi.org/10.1093/qjmed/94.9.491

Hertel, N., Kuzma-Kozakiewicz, M., Gromicho, M., Grosskreutz, J., De Carvalho, M., Uysal, H., Dengler, R., Petri, S., & Körner, S. (2022). Analysis of routine blood parameters in patients with amyotrophic lateral sclerosis and evaluation of a possible correlation with disease progression—a multicenter study. *Frontiers in Neurology*, 13. https://doi.org/10.3389/fneur.2022.940375

Huang, S., Shen, Y., Peng, W., Ye, K., & Zheng, H. (2024). Edaravone for patients with amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Acta Neurologica Belgica*. https://doi.org/10.1007/s13760-024-02476-2

Iłzecka, J. (2003). Prostaglandin E2 is increased in amyotrophic lateral sclerosis patients. *Acta Neurologica Scandinavica*, 108(2), 125–129. https://doi.org/10.1034/j.1600-0404.2003.00102.x

Katzeff, J. S., Bright, F., Lo, K., Kril, J. J., Connolly, A., Crossett, B., Ittner, L. M., Kassiou, M., Loy, C. T., Hodges, J. R., Piguet, O., Kiernan, M. C., Halliday, G. M., & Kim, W. S. (2020). Altered serum protein levels in frontotemporal dementia and amyotrophic lateral sclerosis indicate calcium and immunity dysregulation. *Scientific Reports*, 10(1). https://doi.org/10.1038/s41598-020-70687-7

Leiter, O., & Walker, T. L. (2020). Platelets in Neurodegenerative Conditions—Friend or foe? *Frontiers in Immunology*, 11. https://doi.org/10.3389/fimmu.2020.00747

Liu, G., Hwang, C., Hsieh, C., Lu, C., Chang, S. L., Lee, J., Huang, C., & Chang, H. (2013). Eosinophil-Derived Neurotoxin Is Elevated in Patients with Amyotrophic Lateral Sclerosis. *Mediators of Inflammation*, 2013, 1–7. https://doi.org/10.1155/2013/421389

Louwerse, E., Visser, C., Bossuyt, P., & Weverling, G. (1997). Amyotrophic lateral sclerosis: mortality risk during the course of the disease and prognostic factors. *Journal of the Neurological Sciences*, 152, s10–s17. https://doi.org/10.1016/s0022-510x(97)00238-4

Mandrioli, J., Malerba, S. A., Beghi, E., Fini, N., Fasano, A., Zucchi, E., De Pasqua, S., Guidi, C., Terlizzi, E., Sette, E., Ravasio, A., Casmiro, M., Salvi, F., Liguori, R., Zinno, L., Handouk, Y., Rizzi, R., Borghi, A., Rinaldi, Medici D., Santangelo M., Granieri E., Mussuto V., Aiello M., Ferro S., Vinceti M.; ERRALS Group (2018). Riluzole and other prognostic factors in ALS: a population-based registry study in Italy. J*ournal of Neurology*, 265(4), 817–827. https://doi.org/10.1007/s00415-018-8778-y

Mandrioli, J., Rosi, E., Fini, N., Fasano, A., Raggi, S., Fantuzzi, A. L., & Bedogni, G. (2017). Changes in routine laboratory tests and survival in amyotrophic lateral sclerosis. *Neurological Sciences,* 38(12), 2177–2182. https://doi.org/10.1007/s10072-017-3138-8

Manera, U., Grassano, M., Matteoni, E., Bombaci, A., Vasta, R., Palumbo, F., Torrieri, M. C., Cugnasco, P., Moglia, C., Canosa, A., Chiò, A., & Calvo, A. (2023). Serum chloride as a respiratory failure marker in amyotrophic lateral sclerosis. *Frontiers in Aging Neuroscience*, 15. https://doi.org/10.3389/fnagi.2023.1188827

Murdock, B. J., Goutman, S. A., Boss, J., Kim, S., & Feldman, E. L. (2021). Amyotrophic lateral sclerosis survival associates with neutrophils in a sex-specific manner. *Neurology Neuroimmunology & Neuroinflammation*, 8(2). https://doi.org/10.1212/nxi.0000000000000953

Murdock, B. J., Zhou, T., Kashlan, S. R., Little, R. J., Goutman, S. A., & Feldman, E. L. (2017). Correlation of peripheral immunity with rapid amyotrophic lateral sclerosis progression. *JAMA Neurology*, 74(12), 1446. https://doi.org/10.1001/jamaneurol.2017.2255

Nori, H., Caruana, R., Bu, Z., Shen, J. H., & Kulkarni, J. (2021, July 1). Accuracy, interpretability, and differential privacy via explainable boosting. *PMLR*. https://proceedings.mlr.press/v139/nori21a.html

Ong, M., Tan, P. F., & Holbrook, J. D. (2017). Predicting functional decline and survival in amyotrophic lateral sclerosis. *PLoS ONE*, 12(4), e0174925. https://doi.org/10.1371/journal.pone.0174925

Park, J., Kim, S., Park, D., & Park, J. (2019). Effect of edaravone therapy in Korean amyotrophic lateral sclerosis (ALS) patients. *Neurological Sciences*, 41(1), 119–123. https://doi.org/10.1007/s10072-019-04055-3

Preux, P. M., Couratier, P., Boutros-Toni, F., Salle, J. Y., Tabaraud, F., Bernet-Bernady, P., Vallat, J. M., & Dumas, M. (1996). Survival prediction in sporadic amyotrophic lateral sclerosis. *Neuroepidemiology*, 15(3), 153–160. https://doi.org/10.1159/000109902

PRO-ACT (Project for ALS Clinical Trials). *PRO-ACT - HOME*. (n.d.). https://ncri1.partners.org/ProACT/Document/DisplayLatest/9

Pudasaini, P., Neupane, S., Dhakal, B., Rana, A., Pathak, B. D., & Dawadi, S. (2022). Bulbar onset amyotrophic lateral sclerosis: A case report. *Annals of Medicine and Surgery*, 84. https://doi.org/10.1016/j.amsu.2022.104889

Qureshi, M., Shui, A., Dibernardo, A. B., Brown, R. H., Schoenfeld, D. A., & Cudkowicz, M. E. (2008). Medications and laboratory parameters as prognostic factors in amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*, 9(6), 369–374. https://doi.org/10.1080/17482960802163614

Rafiq, M. K., Lee, E., Bradburn, M., McDermott, C. J., & Shaw, P. J. (2015). Effect of lipid profile on prognosis in the patients with amyotrophic lateral sclerosis: Insights from the olesoxime clinical trial. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 16(7–8), 478–484. https://doi.org/10.3109/21678421.2015.1062517

Stambler, N., Charatan, M., & Cedarbaum, J. M. (1998). Prognostic indicators of survival in ALS. *Neurology*, 50(1), 66–72. https://doi.org/10.1212/wnl.50.1.66

Sun, J., Carrero, J. J., Zagai, U., Evans, M., Ingre, C., Pawitan, Y., & Fang, F. (2020). Blood biomarkers and prognosis of amyotrophic lateral sclerosis. *European Journal of Neurology*, 27(11), 2125–2133. https://doi.org/10.1111/ene.14409

Testa D, Lovati R, Ferrarini M, Salmoiraghi F, Filippini G. Survival of 793 patients with amyotrophic lateral sclerosis diagnosed over a 28-year period. Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders: Official Publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases. 2004 Dec;5(4):208-212. DOI: 10.1080/14660820410021311. PMID: 15799548. https://pubmed.ncbi.nlm.nih.gov/15799548/

Thakore, N. J., Lapin, B. R., Mitsumoto, H., & Consortium, N. P. R. O. a. C. T. (2022b). Early initiation of riluzole may improve absolute survival in amyotrophic lateral sclerosis. *Muscle & Nerve*, 66(6), 702–708. https://doi.org/10.1002/mus.27724

Turabieh, H., Afshar, A. S., Statland, J., Song, X., & Consortium, P. R. O. a. C. T. (2023). *Towards a machine learning empowered prognostic model for predicting disease progression for amyotrophic lateral sclerosis*. PubMed Central (PMC). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10785857/

Wuolikainen, A., Jonsson, P., Ahnlund, M., Antti, H., Marklund, S. L., Moritz, T., Forsgren, L., Andersen, P. M., & Trupp, M. (2016). Multi-platform mass spectrometry analysis of the CSF and plasma metabolomes of rigorously matched amyotrophic lateral sclerosis, Parkinson's disease and control subjects. *Molecular BioSystems*, 12(4), 1287–1298. https://doi.org/10.1039/c5mb00711a

Yang, J., Liu, T., Zhang, L., Li, X., Du, F. P., Liu, Q., Dong, H., & Liu, Y. (2023). Eosinophils at diagnosis are elevated in amyotrophic lateral sclerosis. *Frontiers in Neurology*, 14. https://doi.org/10.3389/fneur.2023.1289467