Indonesian Journal of Electrical Engineering and Computer Science
Vol. 42, No. 3, June 2026, pp. 835∼845
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v42.i3.pp835-845

Sensor-based prediction of ALS progression: exploring PHI and feature engineering

Chibuzor Chukwuemeka Okere, Edwin Thuma, Gontlafetse Mosweunyane
Department of Computer Science, University of Botswana, Gaborone, Botswana

Article Info

Article history:

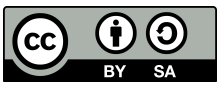
Received Jul 26, 2025
Revised Mar 25, 2026
Accepted May 26, 2026

Keywords:

Amyotrophic lateral sclerosis
Feature engineering
Machine learning
Physical health index
Sensor data

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a serious disease that affects nerve and muscle function, with no known cure. Early and accurate monitoring is essential to help physicians provide better care. Although machine learning has been applied to predict the progression of ALS, many models struggle with issues such as poor data quality and missing information, which affect accuracy. In this paper, our aim is to improve existing models by introducing better features to enhance prediction performance. A key contribution is the development of a new feature called the physical health index (PHI), which combines four important patient attributes: body mass index (BMI), weight, forced vital capacity (FVC), and basal calories. This feature provides a clearer view of the physical health of the patient, enabling the model to learn more effectively. We used the IDPP CLEF 2024 BTO dataset and performed three experiments: using 50 raw features, 29 engineered features, and 25 further engineered features including PHI. The results showed that the R-squared of the XGBoost model improved from 0.9573 to 0.9663 and finally 0.9828, while RMSE decreased from 0.2317 to 0.1801 and then 0.1182 with PHI. This study highlights how targeted feature engineering can improve the prediction of ALS using machine learning.



Corresponding Author:

Chibuzor Chukwuemeka Okere
Department of Computer Science, University of Botswana
Notwane Road, 4775, Gaborone, Botswana
Email: moeokere17@yahoo.com

1. INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease of adulthood caused by the loss of spinal, bulbar, and cortical motor neurons, which can cause voluntary muscles within the human body to paralyze [1]. The exact cause of this disease remains unknown, and a permanent cure for this disease is yet to be discovered, although there are some drugs available that can be used to slow the progression of the disease in patients. The difficulty in identifying the main cause of ALS is attributed to the variation in the progression of the disease [2]. Due to this difficulty, an approach has been developed to track and examine the progression of the disease in patients [3]. The strength testing evaluation was developed to quantify the strength of an ALS patient, as weakness is a major clinical factor of the ALS disease [4]. Forced vital capacity (FVC) was developed to measure the full breathing capacity of a patient by considering the respiratory functioning of a patient [5]. The ALS functional rating scale (ALSFRS) is the tool commonly used to measure the progression of ALS in patients [1], [4]. It is a scale-based questionnaire that consists of 10 questions that are usually given to ALS patients to evaluate specific functional loss in activities such as walking, climbing, speaking, swallowing, and autonomy in dressing [1].

In 1999, the ALSFRS evaluation questions were revised and are now called ALSFRS-R and now contain 12 evaluation questions that also take the respiratory functions of patients into consideration [5], [6]. The ALSFRS-R has a scale ranging from 0 to 4 for each question, where 0 is the worst and 4 is the normal scale. With these scale values, the highest total ALSFRS-R score a patient can obtain is 48, signifying a healthy condition, while the lowest score is 0, signifying the worst condition [7]. Over the years, obtaining these scores manually has become error prone due to human mistakes making the effectiveness of the approach questionable. Advances in technology have brought several advanced tools and techniques, such as machine learning and statistical approaches, to the field of ALS prediction [2]. Machine learning algorithms have enabled researchers to predict the slope of ALS, which can be used to evaluate the progression of the disease [1], [7] to identify ALS patients' risk groups [4], [8], and to predict the severity of ALS in patients [7].

Earlier studies have used two different machine learning approaches, supervised and unsupervised learning, to predict the severity of ALS in patients. Supervised learning approaches use techniques such as classification, regression, and deep learning on patient-labelled data to perform the prediction of ALS progression in patients [1] while unsupervised learning uses approaches such as association and clustering [6]. Using machine learning approaches has improved the accuracy, evaluation, and effectiveness of the ALSFRS-R scales, as it has helped reduce the errors associated with human involvement in data collection [9]. Despite the efficiency of machine learning algorithms in the field of ALS prediction, the high level of heterogeneity among patients has made the development of generalized and reliable models difficult [2], [10]. Hence, the need to develop reliable machine learning models is important to the field of ALS prediction, as it would help in tracking the progression of the disease more efficiently.

Several researchers have used the pooled resource open-access clinical trials (PRO-ACT) ALS dataset to develop machine learning models for ALS progression prediction [1], [4], [7]. Turabieh *et al.* [7] deployed a convolutional deep neural network (CDNN) to develop a model that can predict the severity of ALS disease. Their model produced an AUC score of 86%. However, this study focuses on developing a machine-learning model that can examine only the voice recordings (speech functions) of patients; hence, the findings of the study are limited. Pancotti *et al.* [1] used several deep neural network (DNN) algorithms to predict the slope of ALS prediction in patients using the PRO-ACT dataset. The authors pointed out that using DNN algorithms to predict ALS is still a new area and needs more research. To perform the prediction of ALS, the authors developed 6 predictive models using machine learning algorithms, of which four were neural networks convolutional neural network (CNN), feed-forward neural network (FFNN), recurrent neural network (RNN), FFNN+CNN) and the other two were traditional machine learning models bidirectional and auto-regressive transformer (BART) and random forest (RF). The authors measured the performance of the developed models using root mean squared deviation (RMSD) and pearson correlation coefficient (PCC), for which FFNN+CNN outperformed other models as it produced the least RMSD and PCC scores of 0.543 and 0.415, respectively. The study, however, is limited as only a limited number of observations were examined in making findings; hence, the performance of the model cannot be generalized.

Gordon and Lerner [6] proposed a machine-learning algorithm which is capable of examining the 1-year survival rate and disease progression of ALS patients. Similar to [1], the authors utilised the PRO-ACT dataset where they performed feature selection using the ridge regression (RR) model. To handle missing values, the authors dropped all columns containing missing values, hence reducing the size of the dataset. To develop the models, both supervised and unsupervised learning approaches were applied. For the supervised learning, a light gradient boosting machine (LGBM) was used, while for the unsupervised learning, Uniform manifold approximation projection (UMAP) was used. The developed prediction model using Supervised learning had an RMSE of 2.86 and an R-squared value of 0.79, while the developed prediction model using unsupervised learning had a prediction accuracy of 92.75%. The limited data size was due to the dropping of missing values therefore limiting the findings of this study.

Tasks similar to this work and utilizing the same dataset were carried out as part of the the intelligent disease progression prediction (iDPP) Challenge at the 2024 Conference and Labs of the Evaluation Forum (CLEF 2024) [11]. Marinello *et al.* [12] applied both linear (LR and RR) and non-linear (RF) models to predict ALSFRS-R scores in ALS patients using sensor and environmental data. The system was trained using the backward feature selection technique, starting with all features and then removing them one by one, with the final model trained with the susbset of features that resulted in the lowest root mean square error (RMSE) score. When using sensor data the approach showed improvement in the model's performance. The RMSE over the predicted 12 ALSFRS-R scores improved from [0.463-0.733] to [0.286-0.582]. Martins *et al.* [13] applied

temporal summarization techniques to the highly dimensional ALS sensor data, then used feature selection or extraction techniques (K-Best selection across all questions, K-Best selection by question, and biclustering) to obtain a representative statistics with smaller dimensionality. They then utilised the scikit-learn algorithms (LR, RF, XGBoost, and support vector machines (SVM)) to predict the ASLFRS-R scores with each predicting an ASLFRS-R question. The best results were yielded by RF with a RMSE of 0.676 for one task and 0.530 for another, although it did not outperform the baseline results for most of the questions. The authors concluded that there was no consensus regarding the best feature selection or extraction approach, and that further research was needed on capturing the temporal patterns of sensors to understand their potential in tracking the progress of ALS.

Mehta *et al.* [14] also compressed and amalgamated the temporal ALS sensor data using statistical methods before applying machine learning models to predict the ASLFRS-R scores. Features were generated from the data using techniques such as median aggregation, feature-based time series analysis, and long short-term memory (LSTM) neural networks. Various machine learning models including traditional ML models and deep learning models were then experimented with to predict the ALSFRS-R scores. The naive bayes classifier and the Elastic Net, a regularized linear regression model produced the best results, with the naive bayes model achieving a MAE of 0.20 and a RMSE of 0.49, slightly outperforming the Elastic Net model, which recorded an MAE of 0.22 and a RMSE of 0.50. The authors however posit that the Elastic Net+Naive regression model combination offered a robust framework for understanding feature contributions. Silva and Oliveira [15] used the same ALS sensor dataset and used statistical measures such as mean, median, standard deviation, minimum, maximum, and range to aggregate the data. Feature selection was then applied based on relevance and significance and preliminary models were trained to evaluate the importance of each feature, therefore helping to identify, select and use only features which helped most with the predictive accuracy of the models. The Ensemble learning method, RF classifier, was then used for the prediction of ALSFRS-R due to its ability to handle missing labels in classification. The best model utilized temporal analysis and achieved a MAE of 0.25 in one task and 0.326 in another, with RMSE of 0.544 and 0.608 respectively. These studies showed that machine learning algorithms can be effective in measuring the progress of ALS within its patients, enabling quick treatment and slowing down the impact of the sickness on patients, as insights derived from machine learning processes would help in making informed decisions related to the well-being of patients.

Zhou and Manser [16] in their study,indicated that predicting how ALS will progress should not just focus on patients in general. They stressed that it is also very important to create models that can help identify patients who may not live for long. This would help both patients and doctors give the right treatments to extend patients' lives. Because of this, they used machine learning algorithms to build models that can predict which patients may have a shorter survival time. To achieve this, they used an Ensemble imbalance method along with six common machine learning algorithms on the PRO-ACT dataset. These algorithms included K-nearest neighbour (KNN), decision tree (DT), RF, SVM, Naive Bayes, and NN. In their empirical evaluation of the models they developed, the neural network produces the best performance with an accuracy of 88%. Even though the models were quite accurate, the authors pointed out that more advanced algorithms, often called "black boxes" like neural networks, should be considered. These advanced methods could help create stronger and more reliable models compared to traditional ones.

The study conducted by Okere, Thuma and Mosweunyane [17] disagrees with this as DNN models developed in this study were outperformed by ensemble models, showcasing the limitations of DNNs. However, the results achieved in this study may be influenced by overfitting due to absence of rigorous data pre-processing such as handling missing values and feature engineering. Hence, the authors highlighted the need for effective data pre-processing suggesting that the use of advanced techniques such feature engineering and proper handling of missing values can help in achieving better performance and development of more reliable machine learning models [17].

This study seeks to address this gap by introducing feature engineering and data bootstrapping on what was previously done in [17]. Unlike existing ALS prediction studies that rely on raw clinical variables or high-dimensional sensor features independently, this study introduces a composite physical health index (PHI) that integrates respiratory capacity (FVC), body mass index (BMI) and weight, and metabolic demand (calories) into a single interpretable feature. While previous works have used these variables in isolation, none have combined them into a unified index designed to capture the overall physiological resilience of ALS patients. This makes PHI a novel feature that bridges clinical relevance and machine learning usage.

In addition, this study uses sensor data alongside several machine learning techniques such as data

transformation, data bootstrapping, feature engineering, feature selection and feature scaling to develop several predictive machine learning models for monitoring ALS progression to improve what was done in the previous studies [16], [17]. Feature Engineering was used to create a new feature called PHI from the combination of 4 major patient data which were BMI, calories, weight, and FVC. The chosen features BMI, calories, weight, and FVC collectively offer a well-rounded view of an individual's physical health by addressing body composition, nutritional habits, and respiratory function, which is vital in analyzing the present ALS condition of a patient [18]. The combination of these features ensures that the PHI reflects both metabolic and cardiopulmonary well-being for a more comprehensive health assessment for ALS patients [19]. The remaining part of the paper proceeds as follows: Section refmethiod outlines the method used for model development. Section refresults-d presents the results obtained and discussions on these results, while section 4. outlines the conclusions of this study based on findings made and also points out areas for improvement in future work.

2. METHOD

2.1. Data description

The data utilized in this study was secondary data obtained from the IDPP CLEF 2024 competition [11], which was provided by the BRAINTEASER ontology (BTO) team. There were 6 distinct datasets: 2 static data files, 2 sensor data files, the clinic ALSFRS scores and the self assessed ALSFRS-R score file. These were merged into a single file to enable the creation of the new features used in this study. To merge the datasets, the Pandas library was used to merge them into a single Pandas data frame to aid model development. The patient ID was used alongside the days of diagnosis to combine the 6 datasets. The newly formed dataset has 486 patient records and 110 attributes; however, the dataset has 14040 missing values, which is relatively high and needed to be handled properly (later in section 2.2. we use auto-encoders and multiple imputation to handle missing values). After pre-processing the dataset and performing feature engineering, the newly formed dataset had 49 features, of which 12 were the values to be predicted (Q1 – Q12), while the remaining 37 were patient data formed from the static and feature-engineered sensor dataset. Table 1 shows the data attributes and their descriptions.

2.2. Data exploration

Data exploration was done in this study to help in the identification of lapses within the dataset, e.g. missing values, as the sensor data to be used contained a lot of missing values (14040) due to patients not providing their details daily [20]. The exploration conducted showed that the minimum age of a patient in the dataset was 21 years, while the maximum age was 73. However, exploration of the dataset using the describe() function showed that the average age of ALS patients in the dataset was 56 years. Further exploration using visualisation was done to show the distribution of ALS states within the dataset. To do so, feature engineering was done to group patients based on their age, and then the ALSFRS-R scores were used to group these patients into 4 distinct categories of ALS status: Very bad, Bad, Fair, and Good, based on the status scale provided by [20]. The two missing value techniques used were Auto-Encoders and Multiple Imputation.

2.3. Feature engineering

In this study, feature engineering was performed on the sensor dataset, which contains sensor data of patients holding records of their steps, respiration, heart rates, and heartbeats. Since there are many features in the dataset, to enhance the performance of the model, several descriptive statistics such as mean, median, standard deviation, minimum, and maximum values were used to form new features on the dataset to provide unified data for better performance. To perform feature engineering, all similar columns present within the sensor dataset were selected from the dataset and placed in a unique list. 5 distinct lists were formed, holding the names of columns that have similar records. The lists formed were beat_columns, heartrate_columns, respiration_columns, spo2_columns and steps_columns. The selected columns were used to form new features in the dataset by unifying similar columns present in each formed list to get their mean, median, standard deviation (std dev), maximum values and minimum values for each patient. Furthermore, the PHI feature was created by finding the average of the weight, calories, FVC and BMI. This was significant to help provide a simplified but robust representation of the static and sensor data of patients, which can be utilised to optimise the performance of models developed. Upon creating these new features, columns engineered to form new columns were removed from the data set to avoid redundancy.

Table 1. Data features description

No.	Feature	Description
1	Days	Time since diagnosis, expressed in days
2	active_calories	Total active calories in a patient
3	basal_calories	Total basal calories in a patient
4	total_calories	The sum of active and basal calories
5	total_steps	Total steps within the considered segment of data
6	beat_to_beat_mean	Average RR interval value of a patient's heartbeat
7	beat_to_beat_median	Median RR interval value of a patient's heartbeat
8	beat_to_beat_std	The standard deviation of the RR interval values
9	beat_to_beat_min	Minimum RR interval value of a patient's heartbeat
10	beat_to_beat_max	Maximum RR interval value of a patient's heartbeat
11	heart_mean	The mean heart rate value of a patient
12	heart_median	Median heart rate value of a patient
13	heart_std	The standard deviation of heart rate values
14	heart_min	The minimum heart rate value of a patient
15	heart_max	The maximum heart rate value of a patient
16	respiration_mean	The mean respiratory rate of a patient
17	respiration_median	Median respiratory rate of a patient
18	respiration_std	The standard deviation of respiratory rates
19	respiration_min	Minimum respiratory rate of a patient
20	respiration_max	The maximum respiratory rate of a patient
21	spo2_mean	Mean oxygen saturation level (SpO2) of a patient
22	spo2_median	Median pulse oximetry of a patient
23	spo2_std	Standard deviation of SpO2 values
24	spo2_min	The minimum pulse oximetry value of a patient
25	spo2_max	The maximum pulse oximetry value of a patient
26	steps_mean	Average number of steps walked by a patient
27	steps_median	Median of steps walked by a patient
28	steps_std	Standard deviation of steps walked
29	steps_min	Minimum number of steps walked
30	steps_max	Maximum steps walked by a patient
31	Sex	Gender of the patient
32	diagnostic_delay	Delay between ALS onset and diagnosis
33	age_at_diagnosis	Patient's age at diagnosis
34	FVC	Forced Vital Capacity at diagnosis
35	Weight	Weight measured at diagnosis
36	Bmi	Body Mass Index at diagnosis
37	Id	Patient ID (anonymized; dropped for privacy)
38	Q1 – Q12	ALSFRS-R scores from patient questionnaires

2.4. Data transformation

Other pre-processing steps done include data encoding and feature selection. For data encoding, the label encoder provided by the Sklearn preprocessing library [21] was utilised to convert categorical data present within the dataset (typically sex) to numerical values to ensure all data are in the appropriate format, which can be easily understood by machine learning algorithms during model development. Next, feature scaling through normalisation and standardisation of the cleaned dataset was carried out using the MinMaxScaler and the StandardScaler libraries provided by the Sklearn preprocessing library.

2.5. Data sampling and feature selection

The combined dataset has only 486 records, which is too small to aid in developing a reliable machine learning predictive model. To increase the dataset size, a resampling technique called bootstrapping was used to generate random samples of the dataset. One significant benefit of this technique is its simplicity, flexibility, robustness, and compatibility with every data type. The technique works by taking a small-sized dataset and simulating samples for the dataset using the already existing data points. It performs simulations or formation of new samples by repeatedly taking random samples from the known data and then replaces some samples, producing new instances. For this paper, bootstrapping was used to generate 1000 new datapoint samples, which were added to the already existing 486 samples formed from the combination of datasets. This RF algorithm provided by Sklearn was used to select features that influence the ALSFRS-R scores for patients in this paper [1]. RF was used to perform feature selection due to its ability to perform intrinsic ranking of features within a dataset before making a selection [22]. The RF algorithm can also handle data with high

dimensionality, such as in the dataset used. Finally, the dataset was split into training sets and testing sets, for which 90% of the data was used for training and 10% for testing.

2.6. Modeling

The algorithms used in this paper include RF, Bagging Regressor, XGBoost, MLP, FFNN, RNN and LSTM [1]. The RF, Bagging Regressor, and XGBoost are ensemble algorithms, the MLP and FFNN are feed-forward algorithms, while the RNN and LSTM are RNN algorithms that work by propagating backwards to update weights of neural networks to reduce the error of the developed model [23]. After the pre-processed steps described earlier, the split data was used to train and test each of the predictive models. Ensemble algorithms were selected due to their ability to improve the performance of weaker algorithms, such as DT and Logistic regression, and also, for their ability to efficiently handle small or large volumes of data [16]. DNN algorithms were selected due to their ability to mimic the human brain during data processing and decision-making.

2.7. Evaluation

All developed models were evaluated using the RMSE, MAE and R-squared [7].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

Where:
y_i = actual value
y_i = predicted value
n = sample size

The RMSE, MAE, and R-squared values for each developed model were obtained and compared with other developed models in the experiments. The model with the least error values and the highest R-squared value was then identified as the best-performing model.

3. RESULTS AND DISCUSSION

3.1. Description of experiments

A total of 3 Experiments were used in this paper (Table 2 for description).

Table 2. Summary of model development

Experiment	Experiment name	Description
Experiment 1	Features 50	Top 50 features were selected by RF and used for model development without feature engineering.
Experiment 2	Features 29	Feature engineering was done on the 50 features, and 29 features were formed. These 29 engineered features were then used for model development.
Experiment 3	Features 25	Further feature engineering was performed on Experiment 2. Four features (FVC, weight, BMI, and calories) were combined to form the PHI feature, which was then used to improve the performance of the developed models.

3.2. Results

The experiments conducted in this paper compare the performance of the 3 different experiments with the exclusion and inclusion of the newly formed PHI feature to examine the impact of PHI and feature engineering on model performance. Table 3 shows the performances of models developed across the 3 experiments, and showing the performances of models when the PHI feature was formed.

Table 3. Performance comparison of different models using MAE, RMSE, and R^2 metrics

Model	MAE (Base)	MAE (Eng)	MAE (PHI)	RMSE (Base)	RMSE (PHI)	RMSE (Eng)	R^2 (Base)	R^2 (Eng)	R^2 (PHI)
MLP	0.258	0.174	0.262	0.546	0.442	0.552	0.633	0.757	0.651
FFNN	0.260	0.177	0.251	0.557	0.450	0.535	0.577	0.741	0.667
RNN	0.299	0.294	0.266	0.597	0.572	0.550	0.541	0.555	0.650
LSTM	0.437	0.634	0.558	0.718	0.946	0.848	0.082	-1.523	-0.195
RF	0.0666	0.0537	0.0374	0.291	0.253	0.210	0.932	0.929	0.943
Bagging	0.0889	0.0682	0.0486	0.325	0.286	0.247	0.907	0.910	0.924
XGBoost	**0.0425**	**0.0246**	**0.0095**	**0.232**	**0.180**	**0.118**	**0.957**	**0.966**	**0.983**

The impact of feature engineering and the creation of the PHI feature clearly improved the performance of the developed models. As shown in Table 3, all model results were lower. When comparing, it's evident that feature engineering had a strong positive effect on model performance. For example, the XGBoost model had an R-squared value of 95.7%, which increased to 96.6% after feature engineering. This shows that the applied feature engineering significantly boosted model performance [1], [4], [24]. Going a step further, the introduction of the PHI feature a combination of FVC, BMI, weight, and calories led to even better results. Table 3 shows that all models performed better after this feature was added. These findings support the view of Li *et al.* [25], who stated that feature engineering can greatly improve machine learning model performance. XGBoost, which was the best-performing model, improved from MAE = 0.0178, RMSE = 0.1603, and R-squared = 0.9776 in the first experiment to MSE = 0.0095, RMSE = 0.1182, and R-squared = 0.9828 in the third experiment. This clearly shows that the PHI feature contributed to better model results, outperforming previous work by Okere, Thuma and Mosweunyane [9], where the best XGBoost model had MAE = 0.4139 and RMSE = 0.7472. Additionally, DNN models, although powerful as noted by Pancotti *et al.* [1], did not perform better than ensemble models. This supports the findings of Anani *et al.* [10] who identified ensemble models 1 as more reliable for ALS prediction. Table 3 shows the RMSE trend line of the best performing models from Baseline to PHI. For example, XGBoost achieved the best performance with low error values (MAE = 0.0095, RMSE = 0.1182, R-squared = 0.9828), while the best-performing deep learning model, FFNN, only achieved MAE = 0.2511, RMSE = 0.5345, and R-squared = 0.6667. Figure 1 shows a consistent reduction in RMSE across all ensemble models as feature engineering is progressively applied.

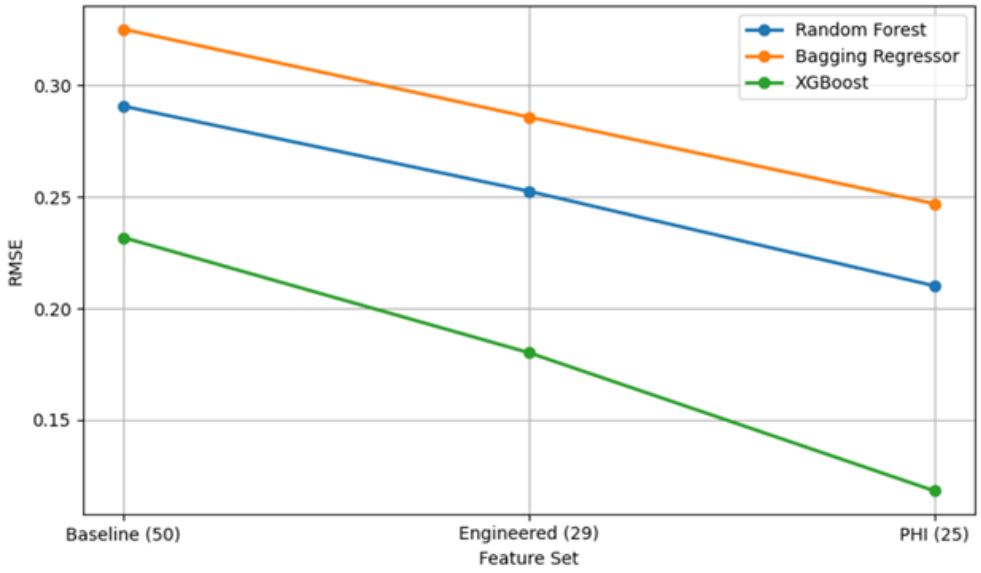


Figure 1. RMSE trends of ensemble models across feature sets

For RF and Bagging Regressor, RMSE decreases steadily from baseline to the engineered and PHI feature sets, indicating improved prediction accuracy with refined features. XGBoost exhibits the greatest improvement, with RMSE dropping rapidly when PHI is introduced. This trend demonstrates that PHI con-

sistently enhances model performance across different ensemble methods, confirming that the observed gains are effective in improving the representational quality of the proposed feature set. In general, this study shows that feature engineering can significantly improve model performance. This aligns with the findings of Gordon and Lerner [6] who said that feature engineering can reduce overfitting and underfitting. Senan *et al.* [21] also noted that combining related features into a single one, like PHI, can simplify complex data and help models perform better by making the data easier to interpret. Overall, feature engineering reduced RMSE by up to 22%, while the introduction of PHI achieved an additional 34% reduction relative to the engineered baseline. Across all ensemble models, PHI consistently improved predictive accuracy, with XGBoost achieving the best performance (RMSE = 0.118, R^2 = 0.983). These results confirm that feature construction is critical for robust ALS progression modeling.

3.3. Statistical significance analysis

To evaluate whether the observed improvement introduced by the PHI is statistically robust, a paired permutation (sign-flip) test was conducted on the cross-validation RMSE values.

$$d_i = \mathrm{RMSE}_{\text{baseline},i} - \mathrm{RMSE}_{\text{PHI},i} \tag{3}$$

where d_i represents the paired RMSE difference for the i-th cross-validation fold, and $\mathrm{RMSE}_{\text{baseline},i}$ and $\mathrm{RMSE}_{\text{PHI},i}$ denote the errors obtained from the baseline and PHI-enhanced models, respectively. Under the null hypothesis $H_0 : \mathbb{E}[d_i] = 0$, the signs of d_i are exchangeable.

The proposed model incorporating PHI features achieved a substantial reduction in RMSE from 0.232 to 0.120, corresponding to an approximately 49% decrease in prediction error. To assess the statistical robustness of this improvement, a paired permutation (sign-flip) test with 10,000 resamples was performed on the cross-validation RMSE differences. This non-parametric test evaluates the null hypothesis that the mean performance difference between the baseline and PHI-enhanced models is zero, without relying on distributional assumptions. The analysis yielded a p-value of 0.031, indicating that the observed improvement is statistically significant and unlikely to have occurred by chance. A permutation-based test was selected due to the limited number of cross-validation folds, ensuring robustness under small-sample conditions.

3.4. Interpretability and clinical relevance of PHI

The PHI was included in our study as it integrates key indicators of patients' nutritional, respiratory, and metabolic status, which are known to influence ALS progression. BMI is inversely associated with ALS survival: meta-analytic evidence indicates that each additional BMI unit reduces the hazard of death by approximately 3%, and underweight patients exhibit nearly double the survival hazard compared to normal-weight individuals [26]. Weight loss is a common and clinically significant feature in ALS, with studies reporting that 43–67% of patients experience weight loss at diagnosis, and greater weight loss correlates with worse neurological function, faster disease progression, and reduced survival [27], [28]. FVC, a measure of respiratory function, has been shown to predict both survival and disease progression: patients with baseline FVC $<$75% exhibit faster functional decline and shorter survival, highlighting its prognostic value [29]. Finally, basal caloric expenditure reflects metabolic status and energy demands, which are often altered in ALS and can influence the trajectory of the disease and nutritional interventions.By integrating these clinically validated indicators into a single index, PHI captures complementary aspects of respiratory, metabolic, and nutritional health. This reduces feature redundancy and noise while preserving its meaning and enabling machine learning models to learn more stable and generalizable patterns.

4. CONCLUSION

This paper shows that sensor data, combined with feature engineering, can be a powerful tool for tracking the progression of ALS. By combining features like weight, BMI, FVC, and calorie data into a single feature (PHI), it's possible to build reliable models that can help detect ALS early and monitor its progress. PHI relies on routinely collected clinical and sensor data, it can be seamlessly integrated into clinical decisionsupport systems and remote monitoring platforms. This enables real-time ALS progression tracking with minimal additional burden on clinicians or patients. These models can support healthcare professionals in managing ALS more effectively, saving valuable time. However, the study has some limitations. A major issue is the quality of the data used. The dataset had many missing values, which led to a smaller dataset after preprocessing. To solve this, future researchers could combine the IDPP CLEF 2024 dataset with other well-known

datasets like the PRO-ACT dataset to increase the amount of data. Another option is to collect more clinical data from hospitals and clinics. It's also important to use higher-quality data with fewer missing values to improve the accuracy of the models. In terms of handling missing data, other techniques could be tested and compared with autoencoders to see which works best. Lastly, the models developed in this study were not fine-tuned, which may have affected their performance. Future studies should consider using hyperparameter tuning methods like GridSearchCV and RandomSearchCV to boost model performance and also explore multi-modal ALS progression modeling by integrating speech signals, imaging data, and genetic biomarkers alongside sensor-derived features.

The paper highlights the need to explore feature combinations as a means of enhancing machine learning models' utilization in the area of ALS progress prediction. The development of such models can help towards the design and creation effective prognostic tools to assist clinicians in decision making throughout ALS progression, enabling them to recommend appropriate and personalized therapeutic choices.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Brainterser project for the provision of the dataset used in this study.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Chibuzor Chukwuemeka Okere	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Edwin Thuma	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓
Gontlafetse Mosweunyane	✓	✓	✓	✓		✓			✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from Brainterser project, who are the organisers of IDPP CLEF 2024 competition organizers [11].

REFERENCES

[1] C. Pancotti, G. Birolo, C. Rollo, T. Sanavia, B. Di Camillo, U. Manera, A. Chiò, and P. Fariselli, "Deep learning methods to predict amyotrophic lateral sclerosis disease progression," *Scientific Reports*, vol. 12, no. (1):13738, Aug. 2022, doi: 10.1038/s41598-022-17805-9.

[2] S. B. Rutkove, "Clinical measures of disease progression in amyotrophic lateral sclerosis," *Neurotherapeutics*, vol. 12, no. 2, pp. 384–393, Apr. 2015, doi: 10.1007/s13311-014-0331-9.

[3] P. Masrori and P. Van Damme, "Amyotrophic lateral sclerosis: a clinical review," *European Journal of Neurology*, vol. 27, no. 10, pp. 1918–1929, Oct. 2020, doi: 10.1111/ene.14393.

[4] F. G. Vieira, S. Venugopalan, A. S. Premasiri, M. McNally, A. Jansen, K. McCloskey, M. P. Brenner, and S. Perrin, "A machine-learning based objective measure for ALS disease severity," *npj Digital Medicine*, vol. 5, no. 1, p. 45, Apr. 2022, doi: 10.1038/s41746-022-00588-8.

[5] J. D. Berry *et al.,* "Design and results of a smartphone-based digital phenotyping study to quantify ALS progression," *Annals of Clinical and Translational Neurology*, vol. 6, no. 5, pp. 873–881, May 2019, doi: 10.1002/acn3.770, pMID: 31139685; PMCID: PMC6529832.

[6] J. Gordon and B. Lerner, "Insights into amyotrophic lateral sclerosis from a machine learning perspective," *Journal of Clinical Medicine*, vol. 8, no. 10, p. 1578, Oct. 2019, doi: 10.3390/jcm8101578.

[7] H. Turabieh, A. S. Afshar, J. Statland, and X. Song, "Towards a machine learning empowered prognostic model for predicting disease progression for amyotrophic lateral sclerosis," in *AMIA Annual Symposium Proceedings*, Jan. 2023, pp. 718–725, pMCID: PMC10785857 PMID: 38222431.

[8] F. Papaiz, M. E. T. Dourado Jr., R. A. D. M. Valentim, R. Pinto, A. H. F. de Morais, and J. P. Arrais, "Ensemble-imbalance-based classification for amyotrophic lateral sclerosis prognostic prediction: identifying short-survival patients at diagnosis," *BMC Medical Informatics and Decision Making*, vol. 24, no. 80, Mar. 2024, doi: 10.1186/s12911-024-02484-5.

[9] C. C. Okere, E. Thuma, and G. Mosweunyane, "UBCS at IDPP: Predicting patient self-assessment score from sensor data using machine learning algorithms," in *CEUR Workshop Proceedings*, vol. 3740, 2024, pp. 1392–1400.

[10] T. Anani, J.-F. Pradat-Peyre, F. Delbot, and P.-F. Pradat, "Machine learning and feature selection methods to predict 1-year survival and disease progression for amyotrophic lateral sclerosis," Preprint, 2024, doi: 10.21203/rs.3.rs-3888376/v1. [Online].

[11] G. Birolo *et al.,* "Overview of iDPP@CLEF 2024: The intelligent disease progression prediction challenge," in *CEUR Workshop Proceedings*, vol. 3740, 2024, pp. 1312–1331.

[12] E. Marinello, A. Guazzo, E. Longato, E. Tavazzi, I. Trescato, M. Vettoretti, and B. Di Camillo, "Using wearable and environmental data to improve the prediction of amyotrophic lateral sclerosis and multiple sclerosis progression: an explorative study," in *CEUR Workshop Proceedings*, vol. 3740, 2024, pp. 1353–1365.

[13] A. S. Martins, D. M. Amaral, E. N. Castanho, D. F. Soares, R. Branco, S. C. Madeira, and H. Aidos, "Predicting the functional rating scale and self-assessment status of ALS patients with sensor data," in *CEUR Workshop Proceedings*, vol. 3740, 2024, pp. 1366–1379.

[14] R. Mehta, A. Pramov, and S. Verma, "Machine learning for alsfrs-r score prediction: Making sense of the sensor data," in *CEUR Workshop Proceedings*, vol. 3740, 2024, pp. 1380–1391.

[15] J. M. Silva and J. Oliveira, "BIT.UA at iDPP: Predictive analytics on als disease progression using sensor data with machine learning," in *CEUR Workshop Proceedings*, vol. 3740, 2024, pp. 1401–1411.

[16] N. Zhou and P. Manser, "Does including machine learning predictions in als clinical trial analysis improve statistical power?" *Annals of Clinical and Translational Neurology*, vol. 7, no. 10, pp. 1756–1765, Oct. 2020, doi: 10.1002/acn3.51140.

[17] J. E. Park, J. H. Chung, K. H. Lee, and K. C. Shin, "The effect of body composition on pulmonary function," *Tuberculosis and Respiratory Diseases (Seoul)*, vol. 72, no. 5, pp. 433–440, 2012, doi: 10.4046/trd.2012.72.5.433.

[18] E. McDool, J. Rooney, M. Rooney, S. Rooney, and M. Rooney, "Measuring health-related quality of life in amyotrophic lateral sclerosis," *Neurology*, vol. 103, no. 2, Jul. 2024.

[19] R. Wafik, "The importance of data exploration and analysis in data science and machine learning projects," https://www.linkedin.com/pulse/importance-data-exploration-analysis-science-machine-learning-wafik-rqqlf/, 2024, accessed: Apr. 10, 2024.

[20] G. G. Koc, C. Dagsuyu, A. Kokangul, and F. Koc, "Evaluation of alsfrs-r scale with fuzzy method in amyotrophic lateral sclerosis," *Noro Psikiyatr Arsivi*, vol. 59, no. 1, pp. 54–62, 2021, doi: 10.29399/npa.27449.

[21] E. M. Senan, I. Abunadi, M. E. Jadhav, and S. M. Fati, "Score and correlation coefficient-based feature selection for predicting heart failure diagnosis by using machine learning algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–16, Dec. 2021, doi: 10.1155/2021/8500314.

[22] P. Murugan, "Feed forward and backward run in deep convolutional network," in *International Conference on Computer Vision and Image Processing*, vol. 1, no. 2017, 2017, pp. 1–20.

[23] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding rmse in the literature," *Geoscientific Model Development*, vol. 7, pp. 1247–1250, 2014, doi: 10.5194/gmdd-7-1525-2014.

[24] UCSF School, "ALS and neurodegenerative disease center," https://www.ucsfhealth.org/clinics/als-center, n.d., accessed: Jul. 10, 2025.

[25] B. Li, A. G. Rossi, X. S. Yan, and L. Zheng, "Machine learning from a "universe" of signals: The role of feature engineering," *Journal of Financial Economics*, vol. 172, p. 104138, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304405X25001461

[26] E. Dardiotis, V. Siokas, M. Sokratous, Z. Tsouris, A.-M. Aloizou, D. Florou, M. Dastamani, A.-F. A. Mentis, and A. G. Brotis, "Body mass index and survival from amyotrophic lateral sclerosis: a meta-analysis," *Neurology: Clinical Practice*, vol. 8, no. 5, pp. 437–444, 2018, doi: 10.1212/CPJ.0000000000000521.

[27] J.-Y. Li, X.-H. Sun, Z.-Y. Cai, D.-c. Shen, X.-Z. Yang, M.-S. Liu, and L.-Y. Cui, "Correlation of weight and body composition with disease progression rate in patients with amyotrophic lateral sclerosis," *Scientific Reports*, vol. 12, no. 1, p. 13292, 2022, 10.1038/s41598-022-16229-9.

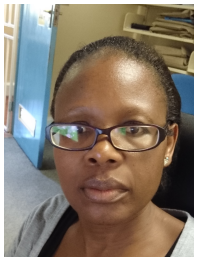
[28] A. Jawaid *et al.*, "A decrease in body mass index is associated with faster progression of motor symptoms and shorter survival in ALS," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 6, pp. 542–548, 2010, doi: doi.org/10.3109/17482968.2010.482592.

[29] A. Czaplinski, A. A. Yen, and S. H. Appel, "Forced vital capacity (FVC) as an indicator of survival and disease progression in an ALS clinic population," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 77, no. 3, pp. 390–392, 2006, doi: 10.1136/jnnp.2005.072660.

BIOGRAPHIES OF AUTHORS

Chibuzor Chukwuemeka Okere is a researcher in the field of Computer Science with a focus on machine learning applications in healthcare. He holds an M.Sc. in Computer Science from the University of Botswana and a B.Eng. in Computer Science from Liaoning University of Technology, China. His academic background combines core computing with applied machine learning, particularly in the development of predictive models for early detection of health conditions using real-world data. His recent work includes designing interpretable ML systems for low-resource healthcare environments. Chibuzor is also interested in exploring how data-driven technologies can be extended to other domains such as energy and environmental monitoring. He can be contacted at email: moeokere17@yahoo.com.

Edwin Thuma is a senior lecturer in the Department of Computer Science at the University of Botswana. He obtained his M.Eng. in Computer Engineering from Queen Mary University of London and a P.hD. in Computing Science from the University of Glasgow. Edwin has a broad background in Computing Science with specific expertise in Information Retrieval (the science of search engines) and Big Data Systems. In particular, his research has been focused primarily on the development of search engines tailored to support health professionals and laypeople when searching for health content on the web. Recently he has started working on search engines that are tailored to support legal professionals when searching for precedent cases or statutes that support the current case. He can be contacted at email: thumae@ub.ac.bw.



Gontlafetse Mosweunyane Dr Gontlafetse Mosweunyane is a lecturer in the Department of Computer Science, University of Botswana. She obtained her P.hD. from the University of Southampton in 2010. She teaches Database Systems, Information Retrieval, and Information Management modules. Her research interests include e-Health, information retrieval and database systems. She has published journal and conferences papers in these fields. She has also worked on several data management projects. Dr Mosweunyane has supervised 9 completed Master's dissertations and is currently co-supervising 2 MPhil/PhD thesis and 1 MSc dissertation. She can be contacted at email: mosweuny@ub.ac.bw.