Title	Prediction of quality of life in people with ALS: on the road towards explainable clinical decision support
Authors(s)	Antoniadi, Anna Markella, Galvin, Miriam, Heverin, Mark, Hardiman, Orla, Mooney, Catherine
Publication date	2021-06
Publication information	Antoniadi, Anna Markella, Miriam Galvin, Mark Heverin, Orla Hardiman, and Catherine Mooney. "Prediction of Quality of Life in People with ALS: On the Road towards Explainable Clinical Decision Support" 21, no. 2 (June, 2021).
Publisher	ACM
Item record/more information	http://hdl.handle.net/10197/13123
Publisher's statement	© the Authors, 2021. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in ACM SIGAPP Applied Computing Review, {21, 2, (2021)} http://doi.acm.org/10.1145/3477127.3477128
Publisher's version (DOI)	10.1145/3477127.3477128

Prediction of Quality of Life in People with ALS: On the Road Towards Explainable Clinical Decision Support

Anna Markella Antoniadi
University College Dublin
Belfield, Dublin 4, Ireland
anna.antoniadi@ucdconnect.ie

Miriam Galvin
Trinity College Dublin
Dublin 2, Ireland
galvinmi@tcd.ie

Mark Heverin
Trinity College Dublin
Dublin 2, Ireland
mark.heverin@tcd.ie

Orla Hardiman
Trinity College Dublin
Dublin 2, Ireland
hardimao@tcd.ie

Catherine Mooney
University College Dublin
Belfield, Dublin 4, Ireland
catherine.mooney@ucd.ie

ABSTRACT

Amyotrophic Lateral Sclerosis (ALS) is a rare neurodegenerative disease that causes a rapid decline in motor functions and has a fatal trajectory. ALS is currently incurable, so the aim of the treatment is mostly to alleviate symptoms and improve quality of life (QoL) for the patients. The goal of this study is to develop a Clinical Decision Support System (CDSS) to alert clinicians when a patient is at risk of experiencing low QoL. The source of data was the Irish ALS Registry and interviews with the 90 patients and their primary informal caregiver at three time-points. In this dataset, there were two different scores to measure a person's overall QoL, based on the McGill QoL (MQoL) Questionnaire and we worked towards the prediction of both. We used Extreme Gradient Boosting (XGBoost) for the development of the predictive models, which was compared to a logistic regression baseline model. Additionally, we used Synthetic Minority Over-sampling Technique (SMOTE) to examine if that would increase model performance and SHAP (SHapley Additive exPlanations) as a technique to provide local and global explanations to the outputs as well as to select the most important features. The total calculated MQoL score was predicted accurately using three features – age at disease onset, ALSFRS-R score for orthopnoea and the caregiver's status pre-caregiving – with a F1-score on the test set equal to 0.81, recall of 0.78, and precision of 0.84. The addition of two extra features (caregiver's age and the ALSFRS-R score for speech) produced similar outcomes (F1-score 0.79, recall 0.70 and precision 0.90).

CCS Concepts

•Computing methodologies → Supervised learning by classification; •Applied computing → Health informatics;

Keywords

Amyotrophic Lateral Sclerosis, Quality of Life, Clinical Decision Support System, Machine Learning, Explainable Artificial Intelligence

1. INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS), also known as Motor Neuron Disease (MND) is a rare neurodegenerative disease. It is characterised by the progressive death of upper and lower motor neurons, causing muscle atrophy and paralysis, and it has a fatal trajectory usually within 3-4 years from symptom onset [46]. As ALS is currently incurable, treatment is mostly palliative and aims to alleviate symptoms and improve quality of life (QoL) [34, 51].

Knowledge of the determinants of a patient's QoL is of interest to clinicians as they can better guide their treatment and support. However, QoL may depend on several characteristics in different combinations according to each individual, including but not limited to the patient's physical disability. These combinations and interactions between the factors relating to QoL may be complicated for a clinician to identify. Machine Learning (ML) allows for the identification of such interactions and the prediction of an outcome of interest, which can be "learned" from the analysis of large volumes of data. For this reason it is used in different aspects of healthcare, such as in the development of Clinical Decision Support Systems (CDSS).

1.1 Clinical Decision Support Systems & Explainable Artificial Intelligence

Clinical Decision Support Systems are computer systems that "provide clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care" [38]. CDSS have been found to improve clinical practice by enhancing the quality of treatment and patient safety, improving adherence to care standards and reducing medical errors [27, 10]. Low resource settings can benefit from CDSS as they can reduce costs of care

[10]. Generally, if used appropriately, such systems can save health care providers' time by a significant amount (estimated around 65%) [50]. CDSS can also reduce cognitive burden for clinicians, but since a system that is 100% accurate does not exist, clinicians are still the main decision makers; ML-based systems have been called "a second set of eyes to inspect a clinical case". Thus, clinicians and CDSS can work as a team towards improved healthcare outcomes for patients and their families, but also for themselves and the healthcare system.

Artificial Intelligence (AI) and ML in particular can be deployed in CDSS due to the variety of techniques available for the processing of that data to create accurate predictions of important clinical outcomes. Accurate predictions may not always come from models that can be explained and understood, especially by non-ML experts. It is argued that, especially in the field of medicine and CDSS, there is great need for explanations of the decisions that ML models suggest [8, 2], and that regardless of their usefulness, the adoption of these models in clinical practice depends on their explainability in most cases [48, 23]. Explainability provides insights into the decision-making reasoning and the term may also be associated with the level of understanding of this reasoning [11]. Explainability is important not only due to the need for clinicians to explain the decision to the patients or other stakeholders, but also for the developers of the system who should be able to validate and improve the model or CDSS under development. Regulations such as the GDPR have also reinforced the need for explainability. According to the GDPR, data subjects have "the right not to be subject solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her" (Article 22). In case the data subject consents to being subject to automated processing, they have "the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision" (Article 22(3), Recital 70).

Explainable AI (XAI) is a popular field of AI that focuses on the development of techniques that can lead to explainable AI models. Techniques can be classified based on their scope into those providing global explanations of the entire system's reasoning and those providing local explanations of single predictions, for a specific instance/individual [2]. Techniques can also be grouped by whether they are model agnostic or model specific (i.e. they can be applied to any or a specific ML algorithm respectively) [2], and by whether they are "ante-hoc" or "post-hoc" explainability methods. Ante-hoc methods are explainable by design (e.g. linear and logistic regression, decision trees, k-nearest neighbours, general additive models and Bayesian models) [24, 8]. Post-hoc methods may provide explanations for models that are developed using a "black-box" or non-explainable technique (e.g. random forests, support vector machines, neural networks) [30, 24, 8]. Although ante-hoc methods are considered "simpler" and are explainable by design, there may be issues with their use that many times lead to a preference towards "black-box" models with post-hoc explanations. In high-dimensional datasets with complex interactions between features, or deep decision trees, these explainable-by-design methods can still be difficult to explain [35], while in many cases "black-box" techniques lead to more accurate predictions [8]. Although the need and benefits of explainability have been expressed in literature, there is a limited number of XAI-enabled CDSS [4].

1.2 Relevant Literature and Aims of Current Study

Previous work in the investigation of QoL in ALS has identified some predictors of QoL using linear models and simple statistics, while there are no ML-based predictive models or CDSS to support the faster prediction of QoL. Some scientists have found a correlation between physical functioning and QoL [52], while some contradict that [43, 40, 17, 31], or have only found relevance with the Physical Functioning QoL sub-scale [29]. Regarding symptoms, dysphagia is an issue that relates to QoL [39, 45], as well as breathing difficulties and fatigue [52]. Palliative care provided by the multidisciplinary team has been shown to improve QoL in two studies [26, 9]. In this work we have a rich set of information available and the support of ML techniques to assist us in gaining insight. In our previous work with the same cohort we predicted caregiver burden and QoL in carers of people with ALS (PALS), and the results were promising [6, 3, 7].

The focus of this work is the development of a CDSS which will provide information on a patient's global QoL (low or high) along with explanations on the predicted outcome. The features used in the analysis contained patient and primary caregiver demographic information, along with use of health services, financial support, social status, aids and appliances received, expenses for treatment, clinical attributes of the patient and caregiving duties of the caregiver. The primary caregiver is the adult who took main responsibility for the patient by offering unpaid support. We believe that the caregiver plays an important role in the care of a patient with ALS, so some information about them might be related to the patient's QoL. Our goal is to create a system that will require a small number of easy to collect or readily-available inputs that will alert a clinician on the risk of a PALS experiencing low levels of QoL, as awareness of this information is clinically useful [42]. We did not include other questionnaires in the study, for example, one that measures depression or anxiety which has been shown to affect QoL. Although this information is relevant their inclusion would have decreased the ease of information collection and clinical usability. An additional aim of our CDSS is to provide explanations for the reasons behind the alerts so that the clinicians can make an informed decision. Our proposed system will provide both global as well as local/individual explanations. This way clinicians will be aware and devise a personalised plan for the provision of the necessary support to the PALS and their caregiver.

We are working towards the prediction of the overall QoL that a patient experienced within the past two days as it is measured by the McGill QoL questionnaire (MQoL) [19]. There are two scores that quantify this, one is a single-item score (MQoL-SIS), the self perceived QoL in the past two days, while the other (MQoL) is the mean value of the scores on five sub-scales that measure physical well-being, physical symptoms, existential well being, psychological symptoms

and support. These two measures of QoL both describe global QoL and have been found to be highly correlated [43], but they are calculated in a different manner so it is useful for clinicians to investigate both.

2. DATA

This study included information on 90 PALS and their primary caregiver, who were interviewed at their residence at three time-points (at four-month to six-month intervals). The patients attended the National ALS/MND multidisciplinary Clinic (MDC) at Beaumont Hospital, Dublin. The pairs of participants were identified through the MDC and were asked to consider participating in the study. They provided informed consent for participation and for access to the patients' clinical information through the National ALS/MND Register after the follow-up interviews. All information was pseudonymised after collection and before conducting any analysis for this work. This study was granted ethical approval from Beaumont Hospital Medical Research Ethics Committee and all participants provided informed written consent.

The McGill Quality of Life (McGill) questionnaire was used to measure the quality of life on a scale of 0 ("very bad") to 10 ("excellent") for each question [19]. The stage of the disease for the patients was measured using both the Milano-Torino (MiToS) functional staging [18] and King's clinical staging [41] systems. The El Escorial [13] criteria were used to determine diagnosis. The Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R) [14] was used as a measure of the progression of ALS. A high ALSFRS-R score shows a better functionality of the patient. The total score ranges between 0 and 48 and the individual questions (ranging between 0 and 4) quantify symptoms of the regions that are used in the El Escorial criteria. Cognitive and behavioural impairment were assessed by the Edinburgh Cognitive and Behavioural ALS Screen (ECAS) [1, 37] and Beaumont Behavioural Inventory (BBI) [20].

To calculate the total QoL in McGill, namely "MQoL", we used the method described in [19], while the "MQoL-SIS" score is a single-item global measure of quality of life. We present the distribution of the two QoL scores in the form of histograms in Figure 1. The vertical red line is used to show where the split in Low and High QoL is made. The scores were split in two classes according to their median value: 6 for MQoL-SIS and 6.85 for MQoL. A higher than median value is equivalent to high quality of life and is represented by 0 in our data, while a lower than median value is equivalent to a low quality of life and is represented by 1 (shown in Eq. 1). For the prediction of the two outcomes we used two datasets with the same features (N=137 predictors), but a different number of entries due to missingness. Entries with a missing value in the target feature were removed leaving 167 entries in the MQoL dataset and 176 entries in the MQoL-SIS dataset.

$$Binary\ QoL = \begin{cases} 1\ or\ low & QoL\ score\ \leq median\ value, \\ 0\ or\ high & QoL\ score\ > median\ value. \end{cases}$$
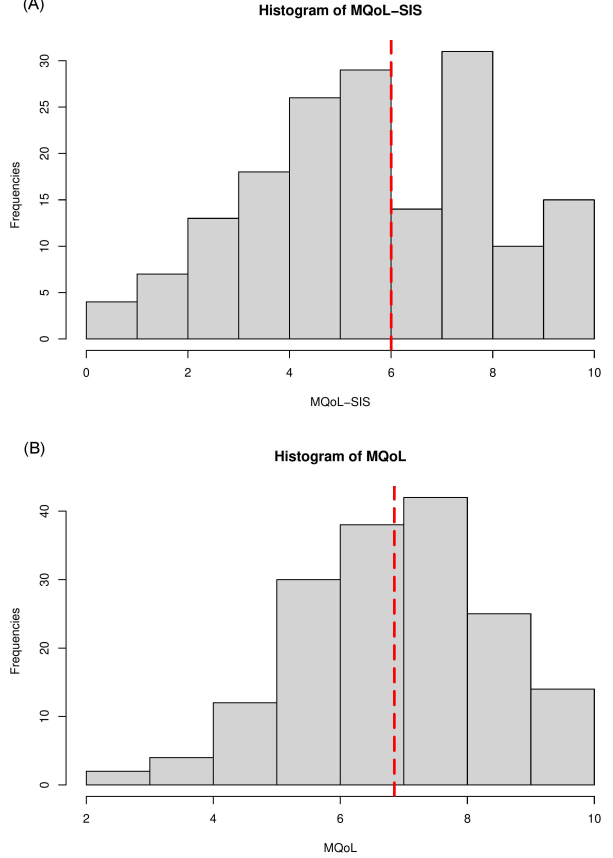$$(1)$$

Figure 1: Histograms of (A) MQoL-SIS scores and (B) MQoL scores.

3. METHODS

A summary of the methodology can be seen in the workflow diagram in Figure 2. This process was repeated for the two different outcomes in parallel. We then demonstrate the global and local explanations for the best models that predict MQoL, along with two case studies.
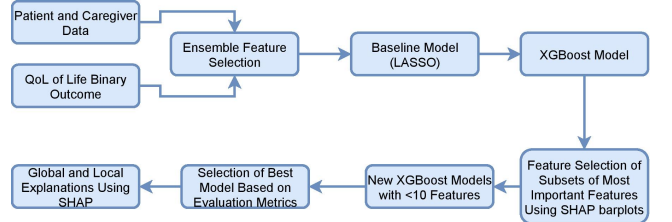

Figure 2: Workflow diagram that shows a summary of the methodology that was used in this paper. These steps were repeated for each of the two outcomes, MQoL and MQoL-SIS.

3.1 Data Pre-processing

In order to select a subset of the most relevant features we applied an ensemble of six different feature selection methods for the elimination of biases in the process [5]. This step

was conducted in R (http://www.r-project.org/) using the "EFS" (Ensemble Feature Selection) package [36]. The ensemble consisted of the following methods: median, Pearson- and Spearman-correlation, logistic regression, and two variable importance measures embedded in the "randomForest" [12] implementation in R. From this methodology we identified the 50 and 51 important features for each of the two outcomes (MQoL and MQoL-SIS respectively).

3.2 Data Analysis

For the creation of the predictive models, the data were split at random into training and test sets (75% and 25% respectively). The distribution of patients into QoL classes is shown in Table 1. Due to the lack of existing models for this research problem to compare ours against, we developed a baseline model using logistic regression; a technique that is explainable by design. A baseline model was trained for each of the two outcomes using Least Absolute Shrinkage and Selection Operator(LASSO) [47]. LASSO was used instead of a standard logistic regression because of the large number of features in the data. The lambda parameter that was used was identified using five-fold cross validation with the method cv.glmnet from the "glmnet" [22] package in R version 3.0-2. To create the model we used the function "glmnet" from the same package. For the baseline models (LASSO) we imputed the missing data using the missForest [44] imputation method in R.

Table 1: Distribution of patients into classes. MQoL-SIS is the overall quality of life single-item Score while MQoL is the calculated total score based on the McGill QoL questionnaire.

Data	High QoL(0)	Low QoL(1)
$MQoL - SIS$		
Full	105	71
Training	81	51
Test	24	20
$MQoL$		
Full	85	82
Training	70	55
Test	15	27

We created the remaining models using the Extreme Gradient Boosting algorithm (XGBoost) [16]. XGBoost is a popular implementation of the gradient tree boosting method [21] that is easy to use, accurate and used by the winning teams of many Machine Learning competitions in different kinds of problems [16]. It creates an ensemble of Decision Trees gradually, by evaluating their predictions and introducing higher weights to the samples that were not successfully predicted in the previous tree, in order to prioritise them in the next prediction "attempt".

After creating the first models with the features identified by the EFS method, we selected their most important features as they were "revealed" by the XAI methodology we describe in Section 3.3. This was done in order to create lower-dimensional datasets. Our aim is to use the least amount of information that can be informative for the prediction of the outcome. Fewer predictive features are easier to collect, so this aspect can lead to a more usable system. Moreover, according to the European General Data Protection Regulations' (GDPR) data minimisation principle, we should collect no more personal information than "is necessary in relation to the purposes for which they are processed"(https://gdpr-info.eu/art-5-gdpr/).

To tune each model's hyperparameters we used an exhaustive search over different values on a five-fold cross-validation, with the "GridSearchCV" from sklearn using the F1 score (scoring="f1") to evaluate the different parameter combinations. The hyperparameters that were tuned were the following: the step shrinkage to prevent overfitting ("eta"), the minimum loss reduction required to make a new split in a tree ("gamma"), the minimum sum of weights required in a child ("min_child_weight"), the subsample of samples to be selected for the creation of each tree ("subsample"), the subsample ratio of columns to create each tree ("colsample_bytree"), and the L2 Regularization term on weights ("lambda"). We set default values equal to 1000 for the number of estimators and 4 for the maximum depth of each Decision Tree in the ensemble. The set of values that were used in the grid search are shown in Table 2 and the final values of all hyperparameters in all models are presented in Table 3. Regarding missing values, XGBoost automatically learns the appropriate branch directions during training, so no imputation method was used in this part of the study.

Table 2: Hyperparameter values used in the grid search for the tuning of the XGBoost models.

Hyperarameter	Set of values
eta	[0.001, 0.01, 0.1, 0.2, 0.3]
gamma	[0.05, 0.5, 1, 1.5]
min_child_weight	[5, 7, 9, 10]
subsample	[0.5, 0.8, 1]
colsample_bytree	[0.6, 0.8, 1]
lambda	[0.1, 0.5, 1]

To evaluate the models we used the F1 score and recall on the training and test sets to see how they compare, and the precision and AUC-ROC (Area Under the Receiver Operator Characteristic Curve) on the test set. Recall is significant in this work as it is more important to accurately predict the people who need support – in this case, the low QoL class. We also present the best models' Precision-Recall curves for the test sets. The formulas of all metrics are presented below:

$$Recall = \frac{TP}{P}$$

$$Precision = \frac{P}{TP + FP}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall},$$

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

where P stands for Positives, TP for True Positives, FP for False Positives, FN for False Negatives, and TN for True Negatives.

3.2.1 Oversampling the MQoL Dataset

We repeated the data analysis methodology that was previously described in the section and in Figure 2 after applying an oversampling technique to increase the training dataset's size. The methodology for model development was again XGBoost, with the use of SHAP barplots to select subsets of features for simpler models. We followed this process only for MQoL, on the dataset that was already reduced in terms of features (after the EFS) to test whether that would lead to more powerful models. The method we selected was SMOTE - Synthetic Minority Over-sampling Technique [15]. This is a very popular technique since its development in 2002, which selects examples that are close in the feature space, and creates a new synthetic example between them. We used the SMOTE method from imblearn in Python, with random_state=0, and we created a total of 175 new random picked points, to end up with 150 PALS in each class. We also tried creating 120 and 200 samples in each class for the full model instead, but we found that 150 yielded better results in the test set. We started with the full model, as before, and then selected smaller samples based on feature importance, oversampling again with 150 PALS in each class. For the the needs of the oversampling implementation, we performed an imputation of the missing data using k-Nearest Neighbors. We used the method KNNImputer from sklearn.impute in Python, with the default parameters.

3.3 Explainable Artificial Intelligence (XAI)

As discussed in the Introduction, our aim is to create a system that can assist clinicians not only by predicting QoL, but by also explaining its predictions. Our baseline model uses a technique that is explainable by design (ante-hoc explainability) and the remaining models we developed require post-hoc explanations. For this reason we used a state-of-the-art XAI technique, namely SHAP (SHapley Additive exPlanations) [33]. SHAP values are suggested as a unified method of six other XAI methods to measure feature importance. This methodology offers the opportunity of both local explanations, for a specific prediction, as well as global explanations that summarise many local ones to give an overview of the model's structure. We used the Tree Explainer [32] as it provides optimal local explanations and interactions, as well as global insights in the structure of tree-based models, such as the XGBoost that we applied in this study. With regard to global explanations, we used local model summarization to present an overview of the model and feature importance.

4. RESULTS

In the first part of this section we present the evaluation of all the fitted models using F1 score, Precision, Recall and AUC ROC score. We show how well the two outcomes are predicted by the Baseline model that used LASSO Regression, the full model using XGBoost and the models that used subsets of features using XGBoost (Table 4). The Baseline model for MQoL-SIS overfits the training data and the results in the training set are not good. The Full model is also overfitted on the training dataset with no good results in the test set. When a subset of ten or five most important

features was used, we saw some improvement that would not suffice for the development of a CDSS.

Regarding the MQoL models, in the Baseline model that used LASSO Regression, the algorithm did not converge, the predictive model consists only of a constant value and predicts all outcomes as 0.476, which means they were all classified as 0. As a result, recall is 0 and the precision and F1 scores are NaN. However, when we used the XGBoost algorithm the results were good on both the training and test sets, although in this case, the model underfits the training data. The good performance of the model that used only five features lead us to remove more features and we found that three features were also quite predictive of the outcome. We are mostly interested in the prediction of the low QoL class (MQoL=1) in order to provide assistance to these patients, so higher recall is a priority. As a result we would recommend MQoL5 or MQoL3 (the models with only five or three features) as they use a very small number of features, while the recall in both the training and test sets is high. The recall in MQoL3 is slightly higher while the precision is slightly lower compared to MQoL5. Figure 3 shows the Precision-Recall curves for the test sets of these models for different thresholds. In MQoL5 we see a significant drop of precision after recall reaches a value of 0.7, but then we see small drops in precision for small increases in recall, while in MQoL3 the average precision (AP) is slightly higher than in MQoL5 and there are larger drops in precision for small increases in recall.

For MQoL we also developed predictive models after oversampling the training set. The models' performance is shown in Table 5. The model overfits the training data and the performance in the test set is reduced, compared to the results in Table 4. Nonetheless, the 3-feature models agree in terms of features.

4.1 Model Explanations

Due to their good predictive power, we are focusing on using the XAI techniques on models MQoL5 and MQoL3. Figure 4 shows a bar plot of the features in order of impact based on the mean absolute value of the SHAP values for each feature, to provide a global explanation of the model. The patient's age at disease onset is the most important predictor with an average magnitude almost equal to 1. It is followed by the caregiver's employment status prior to caregiving with almost 0.8 and then the remaining features have less magnitude although still important for the model. Figure 5 shows the summary plots of the two models. The features are displayed in order of importance, each of the dots is a different entry in the dataset and the color represents the feature value (red is high, blue is low). The reason some dots pile up vertically is to give a sense of density, and the reason they spread horizontally signifies their importance for some individual predictions, even if globally they might not be as important [32]. The plot is split by a vertical line on 0, so all feature values on its left have a negative impact on the output while on its right they have a positive impact. If the colour on one side of the vertical line is mostly constant (close to red or close to blue) then it is easy to explain the relationship with the outcome. For instance, high values (good functionality) in the ALSFRS-R 11 orthopnoea score,

Table 3: Hyperparameter tuning of predictive models. MQoL-SIS refers to the overall QoL Single-Item Score, while the MQoL refers to the the calculated total QoL. Where there is no number in the dataset name, the full dataset is implied, as it was selected by the feature selection method, while the numbers refer to the datasets with the 10, 5, 3 and 2 most important features. The number of estimators were 1000 and the maximum depth of each tree was set to 4.

Model	Hyperpameters
MQoL-SIS Baseline	alpha: 1, lambda: 0.0576
MQoL-SIS Full	colsample_bytree: 0.6, eta: 0.1, gamma: 0.05, lambda: 0.1, min_child_weight: 7, subsample: 1
MQoL-SIS10	colsample_bytree: 0.6, eta: 0.3, gamma: 1, lambda: 0.5, min_child_weight: 5, subsample: 0.8
MQoL-SIS5	colsample_bytree: 1, eta: 0.2, gamma: 0.5, lambda: 0.1, min_child_weight: 5, subsample: 0.8
MQoL Baseline	alpha: 1, lambda: 0.1186
MQoL Full	colsample_bytree: 0.8, eta: 0.2, gamma: 1, lambda: 0.5, min_child_weight: 5, subsample: 0.8
MQoL10	colsample_bytree: 0.8, eta: 0.1, gamma: 0.05, lambda: 0.1, min_child_weight: 5, subsample: 1
MQoL5	colsample_bytree: 0.8, eta: 0.3, gamma: 0.5, lambda: 0.1, min_child_weight: 5, subsample: 1
MQoL3	colsample_bytree: 0.8, eta: 0.01, gamma: 0.05, lambda: 0.1, min_child_weight: 5, subsample: 1
MQoL2	colsample_bytree: 0.6, eta: 0.3, gamma: 0.05, lambda: 0.5, min_child_weight: 5, subsample: 0.8

Table 4: Evaluation of predictive models. MQoL-SIS refers to the overall QoL Single-Item Score, while the MQoL refers to the the calculated total QoL. The Baseline and "Full" models use the whole dataset, while the numbers next to the remaining model names indicate the number of features they use that were selected based on importance. AUC is the Area Under the ROC (Receiver Operator Characteristic) Curve.

Model	f1-training	Recall-training	f1-test	Recall-test	Precision-test	AUC-test
MQoL-SIS Baseline	0.61	0.54	0.17	0.12	0.29	0.49
MQoL-SIS Full	0.54	0.90	0.41	0.35	0.50	0.53
MQoL-SIS10	0.64	0.86	0.50	0.45	0.56	0.58
MQoL-SIS5	0.61	0.82	0.53	0.45	0.64	0.62
MQoL Baseline	NaN	0	NaN	0	NaN	0.50
MQoL Full	0.54	0.91	0.79	0.70	0.90	0.79
MQoL10	0.66	0.96	0.80	0.74	0.87	0.77
MQoL5	0.67	0.82	0.79	0.70	0.90	0.79
MQoL3	0.62	0.84	0.81	0.78	0.84	0.76
MQoL2	0.65	0.62	0.78	0.67	0.95	0.80

Table 5: Evaluation of predictive models for the prediction of MQoL after oversampling. The"Full" model uses the whole dataset, while the numbers next to the remaining model names indicate the number of features they use that were selected based on importance. AUC is the Area Under the ROC (Receiver Operator Characteristic) Curve.

Model	f1-training	Recall-training	f1-test	Recall-test	Precision-test	AUC-test
Over-MQoL Full	0.99	0.99	0.67	0.56	0.83	0.68
Over-MQoL10	0.99	0.99	0.73	0.67	0.82	0.70
Over-MQoL5	0.95	0.97	0.75	0.70	0.79	0.69
Over-MQoL3	0.92	0.91	0.75	0.67	0.86	0.73

which is a measure of shortness of breath when lying down, are predictive of higher QoL.

In order to explore the features some more we used dependence plots (Figure 6) to show how the model output varies by feature value (on the five-variable model for MQoL). The feature used for coloring is automatically chosen based on a potential interaction. In a similar manner to the summary plot, we can explain Figure 7 of local explanations for two hypothetical patients, one with a predicted QoL of 0 (the outcome in probabilities was equal to 0.33) and one with a predicted QoL of 1 (the outcome in probabilities was equal to 0.7). The features on the right (in blue) have a negative impact to the predicted outcome, while the red ones are "pushing" the predicted value to the right, meaning they have a positive impact. In this graph we can see the predicted probability, as well as the specific values the impor-

tant features took in each individual case. It is important to note here that a "positive" impact is numerically positive, which means it moves the output value closer to 1, which is equivalent to low QoL. In Section 5 we present the two hypothetical patients as case studies in order to put the CDSS in context.

5. CASE STUDIES

This section introduces two hypothetical patients along with their predicted probability of low QoL and the explanations of the prediction, in order to put the CDSS in context and to illustrate the factors that were predictive of their QoL. In Figure 7 we show the output of the 3-variable model for better visualisation using SHAP. This model uses the patient's age at disease onset and ALSFRS-R score for orthopnoea, as well as the caregiver's employment status before caregiving
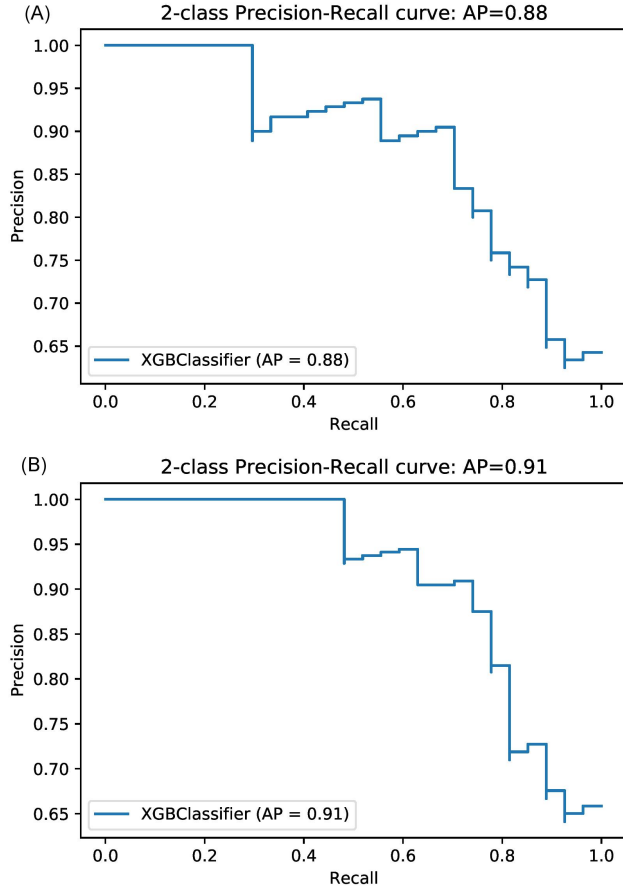
(A) 2-class Precision-Recall curve: AP=0.88

(B) 2-class Precision-Recall curve: AP=0.91

Figure 3: Precision-Recall curves for the test sets of models (A) MQoL5 and (B) MQoL3, along with their average precision (AP).
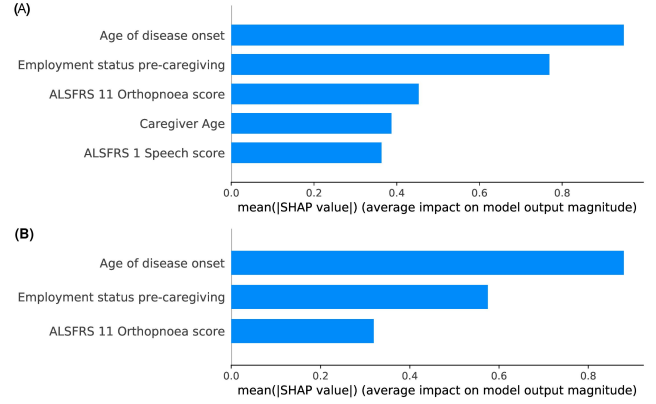

Figure 4: Bar plots that show features in order of importance based on the mean absolute value of the SHAP values for each feature in (A) MQoL5 and (B) MQoL3.
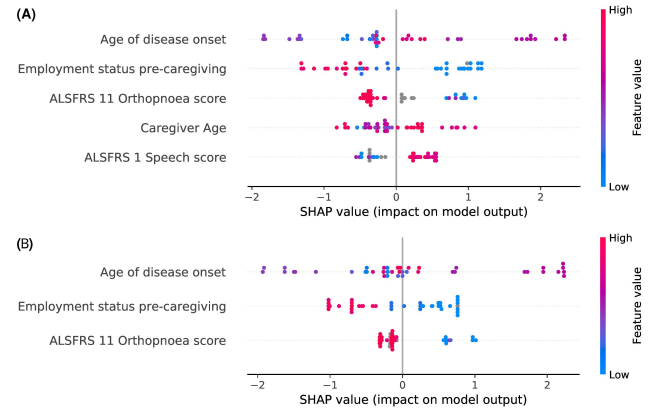

Figure 5: Summary plots of Models MQoL5(A) and MQoL3(B). Features appear in order of their sum of SHAP value magnitudes, and SHAP values show the impact each feature has on the model output. The color represents the feature value (red high, blue low). This reveals for example that a high ALSFRS–R 11 score for orthopnoea lowers the predicted risk of low QoL (or we could say that it is associated with higher QoL).

to make a prediction. A high orthopnoea score (4) corresponds to no breathing problems when lying flat, and a low score (0) represents inability to sleep without mechanical assistance. The caregiver's employment status is a categorical variable where: 1 represents "working for payment or profit"; 2 "looking for first regular job"; 3 "unemployed"; 4 "student/pupil"; 5 "looking after home/family"; 6 "retired from employment"; 7 "unable to work due to permanent sickness or disability"; and 8 "other".

5.1 Case 1

A 61-year-old Caucasian man, with no family history of ALS, presented at the multidisciplinary clinic 18 months after symptom onset, for his scheduled visit. He used to be a bank manager, but is now retired (not due to his ALS diagnosis). He was accompanied by his 63-year-old wife who, due to the fact that they had four children, used to look after the home/family; this is something she also does now, along with her caregiving duties. The patient's onset was spinal, and affected his movement, starting with weakness of his hips and legs. At this visit he was assessed and was found to need some assistance to perform his dressing and hygiene routine, and he also walks with assistance. He has

no breathing, speaking or eating difficulties.

During the visit the patient's information is updated in the health records and based on the patient's age at disease onset, ALSFRS-R score for orthopnoea, and the caregiver's employment status before caregiving, the predicted probability for having a low QoL (a MQoL score below 6.85) is calculated. The patient has a predicted probability of low QoL equal to 0.33, which means we would classify him as a low risk patient (or high QoL) at this time. The prediction and its explanation are shown in Figure 7(A). His age of onset was 60.4 and leads to an increase in the output value (as shown by the arrow and the red colour). The caregiver's employment status before caregiving was 5 (looking after home/family) and the ALSFRS-R 11 score for orthopnoea was 4 (high functionality). Both these features have a negative impact on the predicted probability. The caregiver's
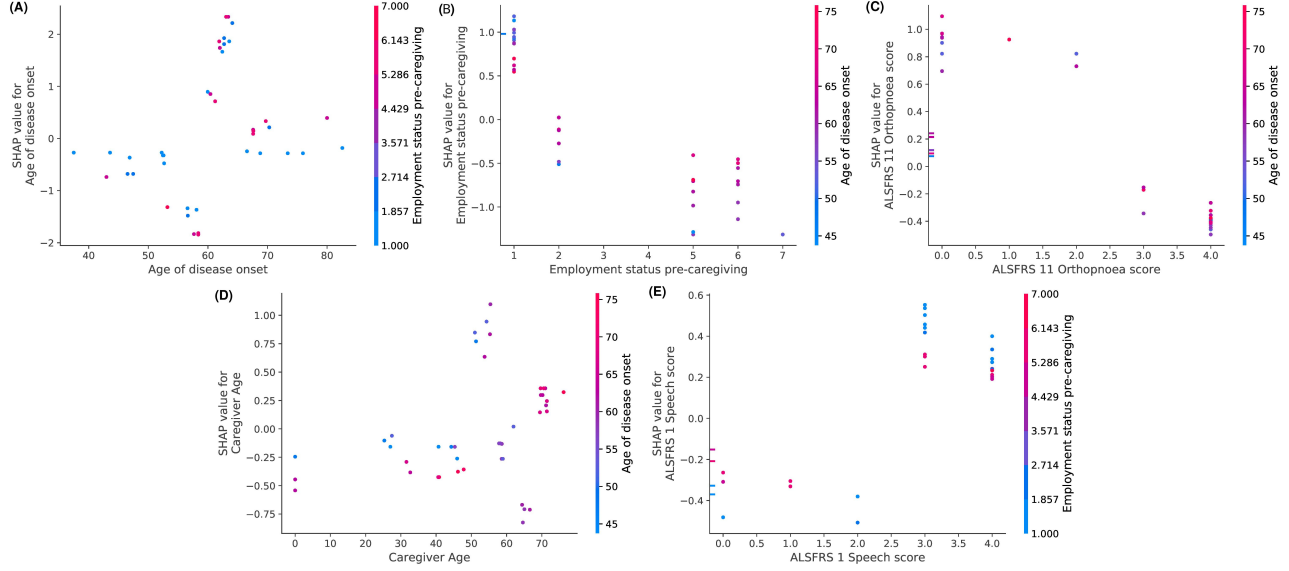
Figure 6: Dependence plots of features in MQoL5 (five-feature model for the prediction of MQoL) that show their impact on the outcome. Each dot is a different entry and the feature used for coloring is automatically chosen based on a potential interaction. (A) Age of onset, (B) Employment status pre caregiving, (C) ALSFRS-R 11 Orthopnoea score, (D) Caregiver Age, (E) ALSFRS-R 1 Speech, are all plotted against their SHAP value.

prior employment status is what, in this case, affected the prediction the most (based on the length of the blue arrow).

5.2 Case 2

A 66-year-old Caucasian woman, with no family history of ALS, presented at the multidisciplinary clinic five years after symptom onset, for her scheduled appointment. She continues having difficulty breathing and is using a BiPAP (Bilevel Positive Airway Pressure) during the night as she has respiratory insufficiency and orthopnoea. When she first attended the clinic, five years ago, she was a retired school teacher. She was accompanied to the clinic by her 68-year-old sister, who has also been retired for ten years. Her sister now lives with her because the patient needs assistance with walking, dressing and hygiene, and turning into bed, and she is her primary caregiver.

The predicted probability of low QoL for the patient was 0.70 which classifies her as an at-risk individual or a low QoL patient (Figure 7(B)). The age of onset was close to case 1 (61.2 years) and had a similar positive effect (although smaller), while the orthopnoea score was 0 (inability to sleep without mechanical assistance) and also has a positive impact on the predicted probability. The caregiver's prior employment status had a negative impact with a value of 6 (retired from employment). The feature that mostly affected the prediction in this case was the orthopnoea score.

6. DISCUSSION

The focus of this work was the development of a CDSS to alert clinicians to a PALS' risk of low QoL. We used two different scores from the McGill questionnaire that represent overall QoL. MQoL-SIS, is defined as a single item

score that describes overall self-perceived QoL in the past two days, while the MQoL score is calculated by sub-scores from the questionnaire that measure physical symptoms and well being, existential, psychological QoL, and support. We included information on the patient's clinical status, aids and appliances, use of healthcare services, expenses due to the condition, and demographic and financial status. Additionally, considering the patient and primary caregiver as a "system" that interacts, we included some demographic information on the caregiver, as well as caregiving duties, employment status and financial supports as a result of their relative's condition. We used an ensemble Machine Learning method, namely XGBoost, for the development of the models and an explainable technique, namely SHAP, to get insights for the predictions.

We found that the SIS score is not easy to predict with the available information, while the calculated total MQoL score was predicted well by only a small number of variables. We identified two models with very similar scores and slight differences in the proportions of correct positive (MQoL=1) identifications. Evaluation on a larger number of patients could clarify whether there are significant differences in performance. The two selected models were the ones with three and five predictive features (MQoL3 and MQoL5) and they were selected over the two- and ten-feature models (MQoL2 and MQoL10) due to their good recall scores and small number of variables. As an additional step to increase the models' performance, we worked towards increasing the training dataset with synthetic samples, using the SMOTE method. The models overfit the training data while the accuracy on the test set decreased compared to the previous models. The three most predictive features were the same as in the model we finally selected (MQoL3), but in the 5-feature model we had one different feature (bed receipt by
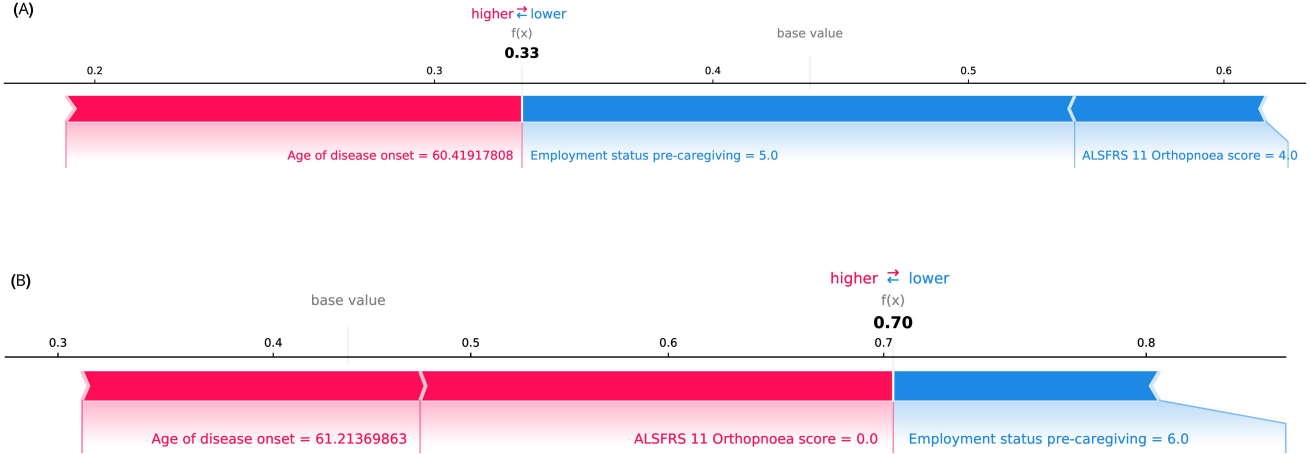
Figure 7: (A) Explanations on a patient's prediction of high MQoL. (B) Explanations on a patient's prediction of low MQoL. Both are created using model MQoL3. Blue shows a decrease in predicted outcome caused by the "blue" features, while the red show an increase. The outcome is shown in the form of predicted probability. A MQoL value of 1 means that we have a case (an "issue"), while 0 means non-case (or no "issue"). When a feature has a positive impact on the outcome it means that the predicted probability of the binary QoL score is increased and is "pushed" towards a value of 1; this value represents a low QoL score.

the patient as opposed to the caregiver's age). This feature could, nonetheless, be considered as a feature of interest in future clinical research.

Our findings suggest that QoL (as it is measured by the Total MQoL score) is related to the age of PALS at disease onset, orthopnoea score in the ALSFRS-R scale, and Speech Score in the ALSFRS-R scale. It is also related to the caregiver's age and employment status pre-caregiving. Using the dependence and summary plots we can further discuss the way each feature affects the predictions. It is important to note that the methodology does not reveal causality; rather, it is giving us insights into what affected the predicted value. Moreover, we emphasise the fact that a value of 1 means that we have a case (an "issue"), while 0 means non-case (or no "issue"). Thus, in this case, when a feature has a positive impact on the outcome it means that the predicted probability of the binary QoL score is increased and is "pushed" towards a value of 1; this value represents a low QoL score. So, a positive impact would be associated with lower QoL and vice versa.

With regards to MQoL, age of onset is the most predictive feature, but it affects the outcome in different ways. The highest ages (above 65 years) have a smaller absolute impact on QoL, while the smallest (below 55 years) have a small negative impact. Interestingly, the values in between have the biggest impact on QoL but it could be either positive or negative. This effect is not very clear and it could be associated with other factors, so we suggest further investigation. Regarding employment status pre-caregiving, we see that the primary caregiver being employed before the onset of their caregiving duties (value of 1 in the feature), has a positive impact on the outcome; thus, it is related to

lower MQoL. This outcome might be associated to the caregiver's burden after the assignment to the additional duty of caregiving or to the satisfaction of the care the patient received. The latter has been found to be associated with MQoL in a previous study [17], along with social status. Lack of orthopnoea (high score) is associated with higher QoL and previous work has also found a similar effect on QoL due to breathing problems [52]. Higher speech functionality (ALSFRS-R 1) is related to lower QoL. In this case there could be a confounding factor that we are unaware of and this would require further investigation. It can be noted that the majority of patients in our study had small or no speech disturbances. Finally, we can see that a lower caregiver's age (below 50 years) is associated with higher QoL (has a negative impact on the predicted value). For ages above 50 there is variation of the effect.

We found that two of the differentiators of the QoL classes were caregiver-related information. Additional support for caregivers could potentially impact both their and the patient's QoL. Provision of palliative care has been found beneficial for the patient's QoL and for the alleviation of caregiver burden [26, 9]. Moreover, support to the family has been recommended for the improvement of QoL [25]. Social support and satisfaction of the care that patients receive affects QoL [17], so, awareness of a patient's QoL can lead to the timely provision of this support. Finally, younger patients might be less affected, according to our findings, but early support could maintain their levels of QoL.

After this work on patient QoL we conducted a study using similar methodologies, for the prediction of caregiver QoL in ALS [7]. Our aim is to combine the models in these studies in order to create a unified CDSS for both parties in

order to simultaneously assess their levels of QoL and provide the necessary advice or support. The most predictive features for the CDSS model for caregiver QoL assessment were the weekly caregiving duties, age and health of the caregiver, and the patient's age of onset and physical functioning (ALSFRS-R scores on cutting food without having undergone a gastrostomy, speech, walking, and dressing). We see that there is an overlap in some of the features that the CDSS will use and it will require a maximum of 12 features for the prediction of both outcomes, even if we end up with the five-variable model from the current study.

6.1 Limitations and Future Work

This study is not without limitations, the main one being the small number of entries in our dataset due to the rarity of the disease. We attempted to resolve this issue by creating synthetic data to increase the data size, but there was no improvement in the performance of the models. Evaluation of the models on more patients is recommended as a further step. Previous studies have found correlations between QoL and psychological, existential and spiritual factors [43, 40, 49, 17], which were not included in this work. Further investigation of such factors combined with the findings in this study could potentially explain some of the effects that were found, and even more accurately predict the MQoL-SIS Score. Moreover, social withdrawal and patient satisfaction for the care that patients receive were not measured in this study but have been identified as predictors of MQoL-SIS [17].

In order for this work to be a step closer to a CDSS, there is a need for practical evaluation and user study. We aim to present the system to clinicians for feedback, to evaluate its quality in terms of ease of use, explainability of outcomes (understanding of the results, visualisation, terminology), and processing time. Subsequently potential users will evaluate the quality of information that the system provides in terms of accuracy and reliability, and decision support [28]. An additional feature that the system could incorporate, especially in the initial stages of its usage, is to ask clinicians to present reasons for disagreeing with the decision. This could result in future improvement of the CDSS in a dynamic manner. It may also be useful to present the results to both clinicians and patients; something that may require different presentation of them, in a manner that is understandable to the lay audience. Both these features have been found to improve system effectiveness [27].

7. CONCLUSIONS

In this work we used a Machine Learning ensemble technique (XGBoost) to create a CDSS to alert clinicians about a patient's QoL in ALS. We also applied a state-of-the art technique which can explain the outputs to clinicians for better usability and system assessment. The explanations can be both individual and also related to the overall structure of the model; the CDSS can incorporate both types of explanations but display them on the user's demand, to avoid overwhelming them with information. The most important predictors of QoL were the age at disease onset, ALSFRS-R score for orthopnoea and the caregiver's status pre-caregiving. Additionally caregiver age and ALSFRS-R

score for speech can potentially add to the predictive power but this should be further evaluated on new patients. While the predictive features would require further investigation and a bigger sample would be needed to validate the predictive models, this work successfully predicts a patient's overall QoL, as it is defined by the calculated total QoL score (MQoL) from the McGill questionnaire. This study is a proof of concept of an informatics solution to assist in the easier assessment of a patient's QoL. We aim to validate these findings against a larger European ALS cohort and to evaluate the explainability of the CDSS with the help of clinicians.

9. REFERENCES

[1] S. Abrahams, J. Newton, E. Niven, J. Foley, and T. H. Bak. Screening for cognition and behaviour changes in als. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15(1-2):9–14, 2014. PMID: 23781974.

[2] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.

[3] A. Antoniadi, M. Galvin, M. Heverin, O. Hardiman, and C. Mooney. Using patient information for the prediction of caregiver burden in amyotrophic lateral sclerosis. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–1, 2020.

[4] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney. Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: A systematic review. *Applied Sciences*, 11(11), 2021.

[5] A. M. Antoniadi, M. Galvin, M. Heverin, O. Hardiman, and C. Mooney. Identifying features that are predictive of quality of life in people with amyotrophic lateral sclerosis. In *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–2, 2020.

[6] A. M. Antoniadi, M. Galvin, M. Heverin, O. Hardiman, and C. Mooney. Prediction of caregiver burden in amyotrophic lateral sclerosis: a machine learning approach using random forests applied to a cohort study. *BMJ Open*, 10(2):e033109, 2020.

[7] A. M. Antoniadi, M. Galvin, M. Heverin, O. Hardiman, and C. Mooney. Prediction of caregiver

quality of life in amyotrophic lateral sclerosis using explainable machine learning. *Scientific Reports*, 2021.

[8] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[9] P. Bede, D. Oliver, J. Stodart, L. van den Berg, Z. Simmons, D. Ó Brannagáin, G. D. Borasio, and O. Hardiman. Palliative care in amyotrophic lateral sclerosis: a review of current international guidelines and initiatives. *Journal of Neurology, Neurosurgery & Psychiatry*, 82(4):413–418, 2011.

[10] A. Belard, T. Buchman, J. Forsberg, B. K. Potter, C. J. Dente, A. Kirk, and E. Elster. Precision diagnosis: a view of the clinical decision support systems (cdss) landscape through the lens of critical care. *Journal of clinical monitoring and computing*, 31(2):261–271, 2017.

[11] U. Bhatt, M. Andrus, A. Weller, and A. Xiang. Machine learning explainability for external stakeholders. *arXiv preprint arXiv:2007.05408*, 2020.

[12] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[13] B. R. Brooks. El escorial world federation of neurology criteria for the diagnosis of amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, 124:96 – 107, 1994.

[14] J. M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, and A. Nakanishi. The alsfrs-r: a revised als functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, 169(1):13 – 21, 1999.

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[16] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.

[17] A. Chiò, A. Gauthier, A. Montuschi, A. Calvo, N. Di Vito, P. Ghiglione, and R. Mutani. A cross sectional study on determinants of quality of life in als. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(11):1597–1601, 2004.

[18] A. Chiò, E. R. Hammond, G. Mora, V. Bonito, and G. Filippini. Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 86(1):38–44, 2015.

[19] S. R. Cohen. *McGill Quality of Life Questionnaire*, pages 3883–3889. Springer Netherlands, Dordrecht, 2014.

[20] M. Elamin, M. Pinto-Grau, T. Burke, P. Bede, J. Rooney, M. O'Sullivan, K. Lonergan, E. Kirby, E. Quinlan, N. Breen, A. Vajda, M. Heverin, N. Pender, and O. Hardiman. Identifying behavioural changes in als: Validation of the beaumont behavioural inventory (bbi). *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 18(1-2):68–73, 2017. PMID: 27894191.

[21] J. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 10 2001.

[22] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, Articles*, 33(1):1–22, 2010.

[23] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.

[24] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4):e1312, 2019.

[25] B. Jakobsson Larsson, A. Ozanne, K. Nordin, and I. Nygren. A prospective study of quality of life in amyotrophic lateral sclerosis patients. *Acta Neurologica Scandinavica*, 136(6):631–638, 2017.

[26] C. Y. Karam, S. Paganoni, N. Joyce, G. T. Carter, and R. Bedlack. Palliative care issues in amyotrophic lateral sclerosis: an evidenced-based review. *American Journal of Hospice and Palliative Medicine®*, 33(1):84–92, 2016.

[27] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj*, 330(7494):765, 2005.

[28] J. Kim, Y. M. Chae, S. Kim, S. H. Ho, H. H. Kim, and C. B. Park. A study on user satisfaction regarding the clinical decision support system (cdss) for medication. *Healthcare informatics research*, 18(1):35, 2012.

[29] S. Körner, K. Kollewe, S. Abdulla, A. Zapf, R. Dengler, and S. Petri. Interaction of physical function, quality of life and depression in amyotrophic lateral sclerosis: characterization of a large patient cohort. *BMC neurology*, 15(1):1–8, 2015.

[30] Z. C. Lipton. The mythos of model interpretability. *Queue*, 16(3):31–57, 2018.

[31] D. Lulé, S. Häcker, A. Ludolph, N. Birbaumer, and A. Kübler. Depression and quality of life in patients with amyotrophic lateral sclerosis. *Deutsches Ärzteblatt International*, 105(23):397, 2008.

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

[33] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus,

S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

[34] J. E. McLeod and D. M. Clarke. A review of psychosocial aspects of motor neurone disease. *Journal of the neurological sciences*, 258(1-2):4–10, 2007.

[35] C. Molnar, G. Casalicchio, and B. Bischl. Interpretable machine learning–a brief history, state-of-the-art and challenges. *arXiv preprint arXiv:2010.09337*, 2020.

[36] U. Neumann, N. Genze, and D. Heider. Efs: an ensemble feature selection tool implemented as r-package and web-application. *BioData mining*, 10(1):1–9, 2017.

[37] E. Niven, J. Newton, J. Foley, S. Colville, R. Swingler, S. Chandran, T. H. Bak, and S. Abrahams. Validation of the edinburgh cognitive and behavioural amyotrophic lateral sclerosis screen (ecas): A cognitive tool for motor disorders. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 16(3-4):172–179, 2015.

[38] J. A. Osheroff, J. M. Teich, B. Middleton, E. B. Steen, A. Wright, and D. E. Detmer. A roadmap for national action on clinical decision support. *Journal of the American medical informatics association*, 14(2):141–145, 2007.

[39] G. Paris, O. Martinaud, A. Petit, A. Cuvelier, D. Hannequin, P. Roppeneck, and E. Verin. Oropharyngeal dysphagia in amyotrophic lateral sclerosis alters quality of life. *Journal of Oral Rehabilitation*, 40(3):199–204, 2013.

[40] R. Robbins, Z. Simmons, B. Bremer, S. Walsh, and S. Fischer. Quality of life in als is maintained as physical function declines. *Neurology*, 56(4):442–444, 2001.

[41] J. C. Roche, R. Rojas-Garcia, K. M. Scott, W. Scotton, C. E. Ellis, R. Burman, L. Wijesekera, M. R. Turner, P. N. Leigh, C. E. Shaw, and A. Al-Chalabi. A proposed staging system for amyotrophic lateral sclerosis. *Brain*, 135(3):847–852, 2012.

[42] Z. Simmons. Patient-perceived outcomes and quality of life in als. *Neurotherapeutics*, 12(2):394–402, 2015.

[43] Z. Simmons, B. A. Bremer, R. A. Robbins, S. M. Walsh, and S. Fischer. Quality of life in als depends on factors other than strength and physical function. *Neurology*, 55(3):388–392, 2000.

[44] D. J. Stekhoven and P. Bühlmann. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 2011.

[45] L. Tabor, J. Gaziano, S. Watts, R. Robison, and E. K. Plowman. Defining swallowing-related quality of life profiles in individuals with amyotrophic lateral sclerosis. *Dysphagia*, 31(3):376–382, 2016.

[46] E. O. Talbott, A. M. Malek, and D. Lacomis. The epidemiology of amyotrophic lateral sclerosis. *Handbook of clinical neurology*, 138:225–38, 2016.

[47] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[48] A. Vellido. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, pages 1–15, 2019.

[49] A. Vignola, A. Guzzo, A. Calvo, C. Moglia, A. Pessia, E. Cavallo, S. Cammarosano, S. Giacone, P. Ghiglione, and A. Chiò. Anxiety undermines quality of life in als patients and caregivers. *European Journal of Neurology*, 15(11):1231–1236, 2008.

[50] K. B. Wagholikar, R. A. Hankey, L. K. Decker, S. S. Cha, R. A. Greenes, H. Liu, and R. Chaudhry. Evaluation of the effect of decision support on the efficiency of primary care providers in the outpatient practice. *Journal of primary care & community health*, 6(1):54–60, 2015.

[51] L. C. Wijesekera and P. N. Leigh. Amyotrophic lateral sclerosis. *Orphanet journal of rare diseases*, 4(1):3, 2009.

[52] C. A. Young, J. Ealing, C. McDermott, T. Williams, A. Al-Chalabi, T. Majeed, G. Burke, A. Pinto, D. Dick, K. Talbot, et al. The relationships between symptoms, disability, perceived health and quality of life in amyotrophic lateral sclerosis/motor neuron disease. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 20(5-6):317–327, 2019.