Dynamic modelling of the ALSFRS-R: leveraging population-based scores using neural networks

Robert McFarlane,[a,*] Robert Ross,[b] Éanna Mac Domhnaill,[a] Adriano Chiò,[c] Philippe Corcia,[d] Caroline Ingre,[e] Christopher J. McDermott,[f,l] Mónica Povedano Panadés,[g] Pamela J. Shaw,[f,l] Philip van Damme,[h] Leonard van den Berg,[i] Ammar Al-Chalabi,[j] Cathal Walsh,[k] and Orla Hardiman[a]

[a]Academic Unit of Neurology, Trinity Biomedical Sciences Institute, School of Medicine, Trinity College Dublin, The University of Dublin, Dublin, Ireland
[b]Department of Computer Science, Technical University Dublin, Dublin, Ireland
[c]"Rita Levi Montalcini" Department of Neuroscience, University of Turin, Turin, Italy
[d]Centre de Reference Maladies Rares SLA, CHU Tours, Tours, France
[e]Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden
[f]Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, United Kingdom
[g]Hospital Universitari de Bellvitge-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain
[h]Neurology Department, University Hospitals Leuven, University of Leuven, and Neuroscience Department, Leuven Brain Institute, University of Leuven (KU Leuven), Leuven, Belgium
[i]Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands
[j]Department of Basic and Clinical Neuroscience, King's College London, Maurice Wohl Clinical Neuroscience Institute, London, United Kingdom
[k]Department of Biostatistics, Trinity College Dublin, Dublin, Ireland
[l]NIHR Sheffield Biomedical Research Centre, Sheffield Teaching Hospitals NHS Foundation Trust, United Kingdom

eBioMedicine
2025;122: 106029
Published Online 15 November 2025
https://doi.org/10.1016/j.ebiom.2025.106029

Summary

Background Amyotrophic lateral sclerosis (ALS) is a rapidly progressive neurodegenerative disorder with highly heterogeneous trajectories. The Revised ALS Functional Rating Scale (ALSFRS-R) is challenging to model due to irregularly spaced data and patient-level variability. Here we sought to develop and validate a short-horizon prediction tool leveraging a fully connected neural network (FCNN) to forecast individual ALSFRS-R trajectories, providing a natural history benchmark for trials and clinical practice.

Methods We retrospectively analysed 29,992 ALSFRS-R measurements from 5319 people living with ALS (plwALS) in the population-based PRECISION-ALS dataset. plwALS were randomised (80:20) into a training and test cohort using group-based splitting. A three-layer FCNN was built in TensorFlow to predict a third ALSFRS-R score given two historical scores and their respective time intervals. Performance was evaluated on the PRECISION-ALS test set and externally on the PROACT database. Linear extrapolation served as a baseline comparator.

Findings On the PRECISION-ALS test set, the FCNN achieved a mean absolute error (MAE) of 0.0552 (95% CI 0.0547–0.0576) on a normalised 0–1 scale, corresponding to 2.65 (2.63, 2.76) points on the 48-point ALSFRS-R. This remained consistent across all post-diagnostic periods. The model generalised well to the PROACT dataset, with an improved MAE of 0.0485 (95% CI 0.0481, 0.0489). Linear extrapolation performed significantly worse across all metrics. Error remained consistent across all clinical groups investigated, such as sex, genotype, site of onset, age at diagnosis, age at onset and diagnostic delay.

Interpretation A short-horizon FCNN can provide clinically interpretable, individualised ALSFRS-R forecasts from sparse, irregularly spaced data. By supporting rapid identification of those who step outside of the model, this approach holds promise for optimising patient counselling, clinical trial monitoring, and early intervention strategies. This approach allows us to better utilise our growing bank of ALS patient data to support decision making.

Funding R McFarlane is supported by a grant from Target ALS, Precision ALS is funded by Taighde Éireann (Research Ireland, formerly Science Foundation Ireland).

*Corresponding author.
E-mail address: mcfarlro@tcd.ie (R. McFarlane).

Keywords: ALS; Amyotrophic lateral sclerosis; Machine learning; Neural networks; ALSFRS-R

Research in context

Evidence before this study

The revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) is the primary clinical outcome measure used in ALS clinical trials, capturing patient symptoms and function over time. Previous studies have highlighted significant heterogeneity in ALSFRS-R trajectories, making accurate prediction challenging. Sparse and irregularly-timed measurements further complicate efforts to reliably model disease progression using traditional linear methods or simple population averages. This presents a problem in ALS research and in the evaluation of clinical trial effects.

Added value of this study

This study leverages an extensive international ALS dataset (PRECISION ALS) comprising 29,992 ALSFRS-R measurements from 5319 patients within a fully connected neural network. The neural network demonstrates robust predictive performance in modelling short-term ALSFRS-R trajectories, effectively capturing patient-level heterogeneity from sparse longitudinal data. This method significantly outperforms linear extrapolation and generalises well to unseen patient data (PROACT).

Implications of all the available evidence

Using neural networks to predict ALSFRS-R trajectories provides a powerful tool for enhancing patient recruitment strategies, monitoring efficacy, and individualising patient management in the clinic and in clinical trials. Furthermore, this approach enables the development of telemedicine-based early warning systems, offering real-time clinical insights and proactive interventions based directly on patient data.

Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder characterised by the degeneration of motor neurons, leading to muscle weakness, paralysis, cognitive change and death.[1] The Revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) is the most widely used clinical tool to assess functional status in people living with ALS (plwALS) and monitor disease progression over time both in a research and clinical setting.[2] The ALSFRS-R evaluates plwALS across four domains: bulbar function, fine motor function, gross motor function, and respiratory function. Twelve questions are asked, each scored between 0 and 4, totalling scores ranging from 0 to 48—the higher the score, the better the plwALS's functional status.[3]

Despite its widespread use, the ALSFRS-R has well-documented limitations, including floor and ceiling effects, non-linear progression—particularly in the early stages of the disease—and a baseline degree of heterogenicity which causes difficult to predict longitudinal trajectories.[4,5] These heterogeneous longitudinal trajectories are hallmarks of the background heterogenicity of the disease and a degree of subjectivity in grading which is unavoidable. This complicates statistical analysis in a research setting and during patient counselling in the clinic. Nevertheless, the ALSFRS-R remains the primary outcome measure in almost every current and past ALS clinical trial.[2]

Advancements in machine learning and the recent availability of large-scale longitudinal population-based datasets[6,7] present an opportunity to improve predictive modelling of ALS progression and better understand the differing temporal dependencies of the scores which make standard statistical analysis difficult. Other studies have focussed on post-diagnostic slope prediction[8] or on score cluster analysis[5] in the context of survival by using a combination of statistical techniques.[8,9] This work tends to rely on the well-known PROACT clinical trial dataset due to its size, before adjusting for sparse population-based data. Whilst some success has been demonstrated with neural networks and the ALSFRS-R using the PROACT dataset as training data,[10] including within classification tasks,[11] this has often required sophisticated preprocessing, domain adaptation, or the use of a large set of features which may not be present at every clinic visit. Thus far an individualised patient-level prediction has not yet been explored or represented with a single statistical technique. The use of population-based data here as training data aims to capture the wider heterogeneity seen in clinical practice before applying this information to a clinical trial database, as validation.

Neural networks are interconnected mathematical relationships which aim to better represent a high-dimensional loss space.[12] Here a fully connected neural network (FCNN) is used, which can be conceptualised as a collection of interconnected regression models which are tuned according to data characteristics before making an estimation to unseen data. Neural networks have been used to model time-series data in other domains such as stock price fluctuations[13] or to predict global weather.[14] However, the application of these networks to ALSFRS-R data has not been fully

explored using population-based data, due to the amount of patient data required.[15]

Here, we leverage the largest ALSFRS-R population-based longitudinal database to date in order to provide individualised predictions using a FCNN. The PRECISION-ALS project data is validated internally with unseen data and then externally on the PRO-ACT database, aiming to show the applicability of this approach to clinical trial datasets, which are known to be notably different from the clinical heterogenicity present in clinical practice.[16–18]

By reframing ALSFRS-R modelling as a machine learning time-series prediction problem, this paper aims to provide a virtual benchmark within a statistical model. Clinicians can then use this model in clinical trials and routine care to compare their patient's trajectory against broader natural history data.

Methods

Data sources

All ALSFRS-R values in this study were collected by a trained investigator during a clinical interview.

Training: PRECISION-ALS

Under General Data Protection Regulation (GDPR)-compliant data-sharing agreements, we accessed de-identified longitudinal data from the PRECISION-ALS project.[19] This dataset was manually harmonised containing information collected from nine specialised European ALS centres with active national or sub-national prospectively collected population-based registries. Data are collected at numerous points from diagnosis to death according to either clinician follow-up in clinic, community outreach, or both. All plwALS diagnosed with "possible," "probable" (with or without laboratory support), or "definite" ALS according to the revised El Escorial criteria were eligible for inclusion.

Validation: PRO-ACT

For external validation, we utilised the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database, which is publicly available to researchers and contains data from multiple interventional ALS clinical trials where the investigational medicinal product was proven to not have any statistically significant effect. There is no overlap between the PRECISION-ALS and PRO-ACT datasets, ensuring an unbiased assessment of model generalisability.

ALSFRS-R segments

50% of people living with ALS (plwALS) within the PRECISION database had 3 ALSFRS-R measurements or greater, three points were then taken as the shortest collection of points that can capture temporal information and variation. Using ALSFRS-R total score and months from diagnosis at which this score was taken, the data were split into all possible 3 sequential points. These 'segments' were then fed into the model; the model aims to predicts a third total score, given the two previously recorded patient scores and the time post-diagnosis they were taken, at the third timepoint.

Fully connected neural network (FCNN) construction

The model was developed in Python using TensorFlow,[20] an open-source machine learning framework that facilitates the construction and training of neural networks. Parameters such as the number of units, dropout rates, and regularisation strength, were selected based on iterative experimentation and evaluation of model performance on the test and validation dataset. Confidence intervals were calculated by bootstrapping model errors via Scikit-learn.[21]

Model training factors

The neural network was implemented as a feedforward architecture with three hidden layers, each comprising ReLU-activated dense units. The first hidden layer contained 64 nodes, the second hidden layer 32 nodes, and the third hidden layer 16 nodes. Dropout was applied after each hidden layer to mitigate overfitting. The final output layer consisted of a single linear neuron that provided continuous ALSFRS-R predictions. The network was optimised via the Adam optimiser, with a learning rate of 0.001 and a mean squared error (MSE) loss function.

Training was conducted for a maximum of 100 epochs, with early stopping applied based on validation loss. The early stopping criterion was set with a patience of 5 epochs, meaning training would halt if no improvement in validation loss was observed for five consecutive epochs. Model convergence was achieved within 10 epochs, beyond which no substantial improvement in validation performance was noted. The initial training loss was 0.0610 (MSE), which progressively decreased, reaching a final loss of 0.0081 (MSE) after 10 epochs. The mean absolute error (MAE) also declined, starting at 0.1782 and stabilising around 0.0665. The training and validation loss are presented in Fig. 1.

Long Short-Term Memory (LSTM) modelling

To assess whether neural network memory mechanisms improved prediction accuracy, a Long Short-Term Memory (LSTM) model was constructed for the same task.[22] Inputs remained the same and two stacked LSTM layers with 25 units each were used with dropout layers in between (0.3). Return sequences was set to true in the first layer to pass the full temporal encoding to the second layer, producing a summary of the input. This was then combined with the future time input and processed through a dense ReLU layer and L2

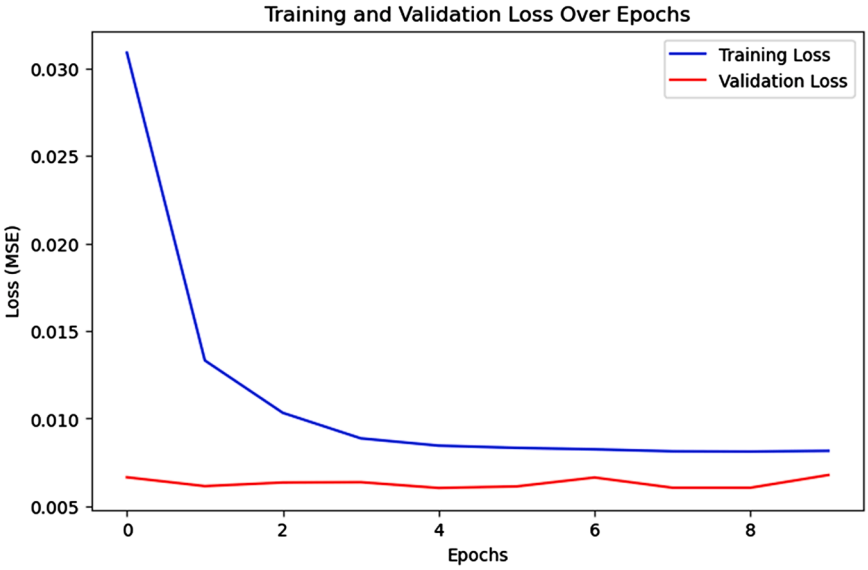

***Fig. 1:* Model training loss diagram**. A graph demonstrating the training of the FCNN model. The network was optimised via the Adam optimiser, with a learning rate of 0.001 and a mean squared error (MSE) loss function. A maximum of 100 epochs was set, with early stopping applied based on validation loss to prevent overfitting. The early stopping criterion was set with a patience of 5 epochs. Model convergence was achieved within 10 epochs. The initial training loss was 0.0610 (MSE), which progressively decreased, reaching a final loss of 0.0081 (MSE) after 10 epochs. The validation loss did not show a large increase, a promising sign that there was minimal overfitting.

regularisation. After another dropout layer, the final output was produced by a single linear activation neuron, as with the FCNN. This model was compiled with the same optimiser and loss function as previously.

Ethics and funding

All plwALS who contributed to this study have provided their infomed consent, which is stored locally at each of the participating sites. Performance of this study was approved by each of local instuitional ethics review boards and collected according to the TRICALS consortium agreement. PRECISION ALS is funded by Taighde Éireann. Robert McFarlane is supported by a research grant from Target ALS, a charitable organisation supporting ALS research.

Role of funders

No funders had any role in study design, data collection, data analysis, interpretation or writing of this manuscript.

Results

A total of 48,375 ALSFRS-R measurements from 10,633 plwALS were initially obtained from the PRECISION-ALS dataset. After quality control, 47,341 valid measurements from 10,633 individuals remained. Exclusions occurred primarily for impossible ALSFRS-R values (>48 or <0) and large score increases (>4 points). Those with three scores or greater totalled 5319 plwALS across 29,992 measurements. The background clinical characteristics of the included groups are summarised in Table 1.

These 5319 patient trajectories were further divided into an 80% train set of 4272 plwALS across 29,992 measurements and a 20% test set from 1047 plwALS with 5702 measurements. A group-based split was performed based on individual ID to ensure that a plwALS only appeared in either the test or train sets (using GroupShuffleSplit). The split features of the training and test sets (time from diagnosis and total score) were then separately passed through a MinMax Scaler. The train set was then fed through the model and performance evaluated on the test set and the PROACT database as external validation. The same preprocessing steps were performed on the PROACT data set, which was only used post model construction as validation. All those with valid ALSFRS-R measurements were included totalling 4610 plwALS across 28,743 measurements.

Model evaluation

The FCNN was evaluated on the unseen test dataset and the external proact dataset. On the unseen PRECISION dataset, the model achieved a MSE of 0.0060 (95% CI 0.0058, 0.0066), a RMSE of 0.0777 (95% CI 0.0759, 0.0815) and a MAE of 0.0552 (95% CI 0.0547, 0.0576). The coefficient of determination (R2) was 0.8694 (95% CI 0.8557, 0.8752) when the predicted and actual values were compared, which is represented in Fig. 2. This MAE translates to 2.65 (95% CI 2.63, 2.76)

points on the original 48-point scale. A lack of systematic bias in prediction is demonstrated as a Bland-Altman plot in Supplementary Figure S1. Four example individual patient-level trajectories are included to illustrate how the model can perform on individual trajectories and the modelling constraints considered in Fig. 3.

When the model was then externally validated on the PROACT dataset all metrics of precision improved. MSE reduced to 0.0038 (95% CI 0.0037, 0.0039), RMSE to 0.0624 (95% CI 0.0612, 0.0628) and MAE to 0.0485 (95% CI 0.0481, 0.0489). R2 improved to 0.8961 (95% CI 0.8931, 0.8992). This MAE translates to 2.33 (95% CI 2.31, 2.35) points on the original 48-point scale. This confirms the model's ability to generalise and actual versus predicted scores are presented on a normalised scale in Fig. 4. This is further supported by Bland-Altman plots showing no systemic bias in prediction as Supplementary Figure S2.

Linear approximation

As two points are being used to approximate the third, a linear extrapolation of the two original points was taken to ensure that the model is learning the temporal complexities of the progression profiles of plwALS involved rather than simply regressing to a third point. The model performed significantly better than the linear approximation across all metrics. The linear model achieved a MSE of 0.0814 (95% CI 0.0457, 0.1294), a RMSE of 0.2852 (95% CI 0.2138, 0.3598), a MAE of 0.0961 (95% CI 0.0896, 0.1039) and a R square value of –0.7649 (95% CI –1.8008, 0.0428). This MAE translates to 4.61 (95% CI 4.30, 4.99) points on the original 48 point scale. All error metrics in this section are reported in Table 2 for comparison.

Clinical variables

As a post-hoc analysis, the degree of error was tested within clinical groups that are known to have differing effects on disease progression. A Mann–Whitney U test revealed no significant difference in error between males and females ($p = 0.1385$). Similarly, Kruskal–Wallis tests demonstrated no effect of *C9orf72* status ($p = 0.3032$), *SOD1* status ($p = 0.5398$), *FUS* status ($p = 0.5403$), *TARDBP* status ($p = 0.4338$), or site of onset ($p = 0.6397$) on model error.

Spearman correlation analyses found no significant relationship between error and age at onset ($r = 0.006$, $p = 0.6763$), age at diagnosis ($r = 0.008$, $p = 0.5552$), or diagnostic delay ($r = -0.002$, $p = 0.8659$). Collectively, these findings indicate that the predictive accuracy of the model is not meaningfully affected by the examined demographic or genetic variables.

Time

As indicated by other published works, the trajectory can change over time, the FCNN model error was then tested across time intervals to ensure that it can predict with consistent error across post-diagnostic periods. Short term (0–12 months), medium term (12–24 months), long term (24–48 months) and very long term (>48 months) were chosen.

MAE remained constant with scores of 0.0545 (95% CI 0.0523, 0.0567), 0.0571 (95% Cis 0.0543, 0.0600), 0.0514 (95% CI 0.0486, 0.0543) and 0.0415 (95% CI 0.0382, 0.0451) respectively. RMSE showed the same trend with 0.0752 (95% CI 0.0711, 0.0791), 0.0819 (95% CI 0.0760, 0.0880), 0.0757 (95% CI 0.0698, 0.0816) and 0.0601 (95% CI 0.0539, 0.0670) respectively. R squared improved sequentially with scores of 0.815 (95% CI 0.7942, 0.8346), 0.837 (95% CI 0.8126, 0.8595), 0.886 (95% CI 0.8671, 0.9038) and 0.9332 (95% CI 0.9166, 0.9472) respectively. These error metrics are reported in Table 3 for comparison. This slight improvement in longer time periods points to a reduced variability in those who survive for a longer time period, which correlates to clinical experience.

ALSFRS-R sub scores

ALSFRS-R sub scores were also available in this data set, and a FCNN following the same architecture was constructed and trained according to the same data protocol. Model metrics were adjusted to prevent overfitting as previously and the same early stopping methodology was used. All ALSFRS-R sub scores (0–4) for the 12 items were included as a replacement for the total value model feature. The number of nodes at the level of each dense layer was adjusted exponentially to ensure that the FCNN was able to have enough opportunity to learn given the increase in dimensionality

	Train (PRECISION ALS)	Test (PRECISION ALS)	External validation (PROACT)
Male (n, %)	2851 (61.31%)	692 (65.85%)	3265 (63.26%)
Female (n, %)	1799 (38.69%)	359 (34.15%)	1896 (36.74%)
Bulbar Onset (n, %)	845 (18.17%)	239 (22.74%)	312 (17.97%)
Spinal Onset (n, %)	3032 (65.19%)	663 (63.18%)	1424 (82.03%)
Time between measurements per patient (median, months, IQR)	3.40 (3.02)	3.61 (6.41)	2.47 (1.50)
Total duration of data capture per patient (median, months, IQR)	6.80 (6.05)	7.22 (6.98)	4.19 (2.95)
Age of Onset (years median [IQR])	60.87 (17.06)	63.55 (15.66)	54.36 (15.06)
Age of Diagnosis (years median [IQR])	62.17 (16.95)	64.25 (16.56)	55.22 (14.96)
Diagnostic Delay (years median [IQR])	0.95 (0.92)	0.85 (0.89)	0.76 (0.64)
C9orf72 expansion (n, %)	206 (4.43%)	57 (5.43%)	N/A
Pathogenic *FUS* variant (n, %)	22 (0.47%)	6 (0.57%)	N/A
Pathogenic *TARDBP* variant (n, %)	71 (1.53%)	19 (1.81%)	N/A
Pathogenic *SOD1* variant (n, %)	255 (5.48%)	22 (2.09%)	N/A

Table 1: **Patient characteristics.**

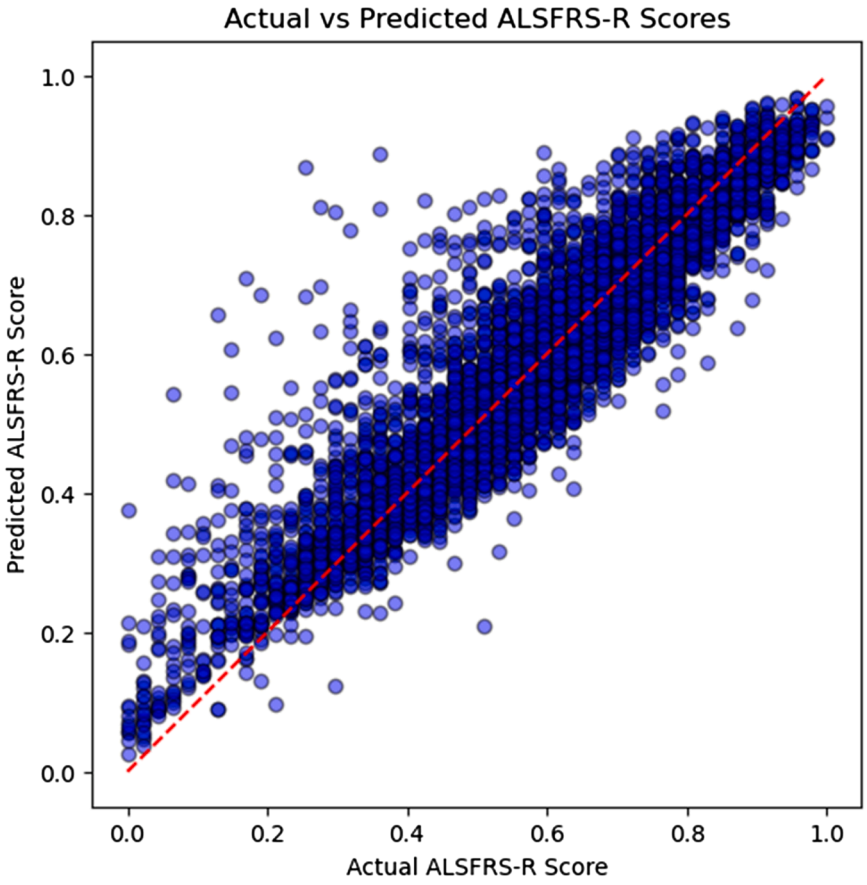

***Fig. 2:* Actual versus predicted ALSFRS-R scores**. A figure showing the predicted versus the actual ALSFRS-R score of the FCNN model on the PRECISION test dataset (1047 plwALS with 5702 measurements), after having been trained on the train dataset (4272 plwALS across 29,992 measurements). This is presented on a normalised 0–1 scale for interpretability; the coefficient of determination (R^2) was 0.8694 (95% CI 0.8557, 0.8752).

of the data inputs. Error metrics however did not improve substantially despite this. The error metrics achieved were as follows: MAE was calculated as 0.0701 (95% CI 0.0685, 0.0717), RMSE 0.0914 (95% CI 0.089, 0.0943) and R squared 0.819 (95% CI 0.810, 0.831).

LSTM modelling

When assessed on the PRECISION dataset the LSTM achieved a mean absolute error (MAE) of 0.0584 (0.0570, 0.0598), a root mean squared error (RMSE) of 0.0796 (0.0769, 0.0823), and a mean squared error (MSE) of 0.0063 (0.0059, 0.0068) on the normalised 0–1 scale. This translated to an MAE of 2.74 (2.68, 2.81) and an RMSE of 3.74 (3.62, 3.87) on the original ALSFRS-R 0–48 scale. This model achieved an R squared of 0.8629 (0.8527, 0.8725). These values are represented in Table 1.

Discussion

This study applies a deep neural network to a harmonised, population-based, pan-European ALS patient database. This is with the aim of creating a model that can be used as a tool for researchers and clinicians in the prediction of a heterogenous, non-linear degradation pattern that consistently presents a statistical challenge at analysis stage. From a computer science perspective, the prediction of the third point in a series is a time-series prediction problem—an area which has considerable literature.[23] There are however data volume and structure constraints that mean other methods of time series prediction are either impossible or extremely difficult in the context of ALS. Data here are sparse, only 50% of participants actually had three or more measurements of their ALSFRS-R. It is also irregularly spaced—reflecting the nature of actual patient follow-up—rather than consistently spaced, as in stock market calculations or in inventory management. And finally, patient-level trajectories can be markedly different, not following seasonal or cyclical trends, for which other modelling techniques are specifically designed.

The use of a neural network here has captured irregular temporal dependencies and dynamically modelled something which is highly heterogenous. Compared to other published work, this model used a

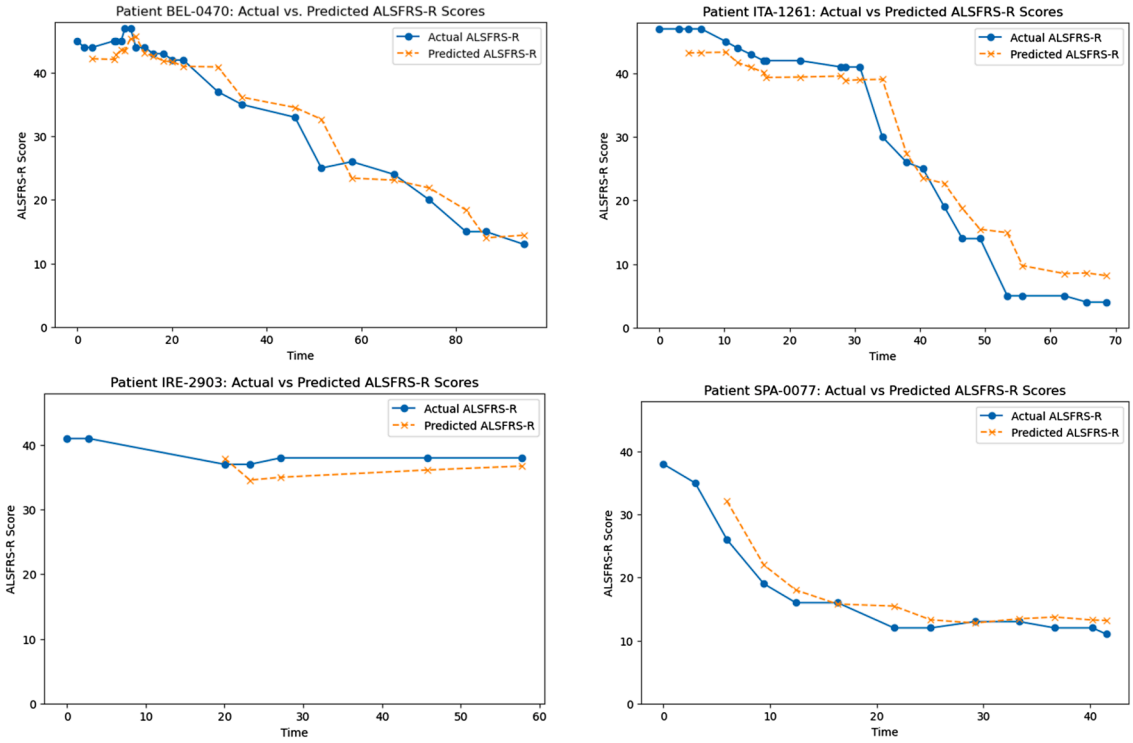

Fig. 3: **Individualised ALSFRS-R trajectories**. Four ALSFRS-R patient trajectories were taken at random from four different included countries (Belgium, Italy, Ireland and Spain) to illustrate the heterogeneity of the dataset and therefore the difficulty in statistical analysis and the reason for deciding upon a Fully Connected Neural Network (FCNN). Patients can progress both linearly and non-linearly (BEL-0470 and SPA-0077). Some show minimal or slow change (IRE-2903) while others can show a plateau before reduction (ITA-1261). Please also note the fact that the follow-up period for each patient is very variable, reflecting the reality of patient follow up in a population-based context. Included is also the prediction of the FCNN model; this allows the error of the model to be visualised at an individual patient level, to improve interpretability. Importantly, where the prediction error is largest for example at approximately 50 months post-diagnosis for patient BEL-0470 or at 35 months post-prediction for patient ITA-1261, these are precisely the types of deviations from expected progression that are of greatest clinical interest.

simpler architecture and still achieved lower RMSEs than domain-adapted LSTM and XGB across comparable time intervals.[10] This is likely due to the population-based nature of the training data. Data constraints meant that a short time horizon had to be adopted and more complex neural network structures such as Long Short-Term Memory (LSTM) models were not superior to a FCNN in testing for this specific task. Indeed, when a LSTM was applied to this task, it offered no tangible advantage over the FCNN with the simpler architecture marginally outperforming in terms of a lower MAE (2.65 versus 2.75), RMSE (3.73 versus 3.74) and a higher R squared value (0.8694 versus 0.8629). This is likely because the sequence length is determinant to learning in these types of networks. The current most comprehensive population-based data capture methods are insufficient due to the nature of real-world data collection of the ALSFRS-R and the inherent short disease duration. An idea supported by other works in this area which show that when similar sequence lengths are compared, FCNNs and similar feed-forward networks outperform more complex architectures due to underutilisation of their gating mechanisms.[24,25]

The model's short time horizon however is still useful, its prospective use in a clinical trial context could help with faster identification of 'slow' or 'fast' progressing plwALS; those which significantly step outside of the model's predictions. For example, in a data safety and monitoring board meeting (DSMB) undifferentiated ALSFRS-R scores can be compared to this population-level benchmark when only a few points are available. How far a certain group steps outside of the model could support decision making in this context with sparse, incoming data. In terms of overall fit, the model demonstrates a strong coefficient of determination (R^2 = 0.8694, 95% CI 0.8557, 0.8752), as shown in Fig. 2. However, the utility of the model is more clearly appreciated in Fig. 3, where the individual trajectories highlight the FCNN's capacity to capture patterns. Linear, non-linear, plateau and static phases of

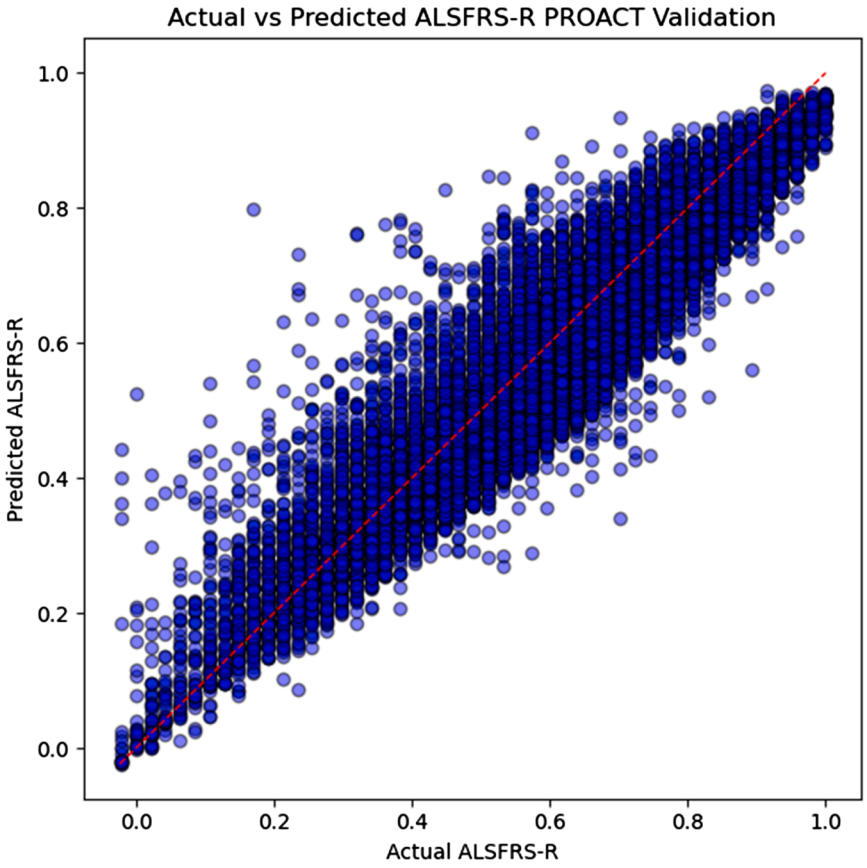

***Fig. 4:* Actual versus predicted ALSFRS-R scores**. A figure showing the actual versus the predicted ALSFRS-R scores when the FCNN model is used on the external validation cohort of the PROACT dataset. This is presented on a normalised 0–1 scale for interpretability. All error metrics improved from that of the test dataset presented in Fig. 2. The coefficient of determination (R^2) was 0.8961 (0.8931, 0.8992), an improvement from 0.8694 (95% CI 0.8557, 0.8752). This confirms the model has the ability to generalise to unseen, out of cohort data.

decline are demonstrated. This performance suggests the model has captured some of the clinical heterogeneity characteristic of ALSFRS-R scores.

Importantly, where the prediction error is largest for example at approximately 50 months post-diagnosis for patient BEL-0470 or at 35 months post-prediction for patient ITA-1261 (Fig. 3), these are precisely the types of deviations from expected progression that are of greatest clinical interest. Using this as a form of error detection, an FCNN could be used to prevent and identify mismatched arms of trials, helping to decide if a trial should be continued or stopped early. Additionally, certain centres have moved towards patient-directed ALSFRS-R measurements, for example through an app,[26] smartphone[27] or wearable devices.[28] This model can be applied as an initial assessment or early warning system for clinicians, providing a level of data-driven clinical governance for scores which are performed at distance.

Model metrics	MSE (95% CI)	RMSE (95% CI)	MAE (95% CI)	R Squared (95% CI)	MAE (ALSFRS-R Scale) (95% CI)	RMSE (ALSFRS-R Scale (95% CI)
FCNN PRECISION (test)	0.0060 (0.0058, 0.0066)	0.0777 (0.0759, 0.0815)	0.0552 (0.0547, 0.0576)	0.8694 (0.8557, 0.8752)	2.65 (2.63, 2.76)	3.73 (3.64, 3.91)
FCNN PROACT	0.0038 (0.0037, 0.0039)	0.0624 (0.0612, 0.0628)	0.0485 (0.0481, 0.0489)	0.8961 (0.8931, 0.8992)	2.33 (2.31, 2.35)	2.99 (2.94, 3.01)
FCNN ALSFRS-R Sub domains	0.0084 (0.0079, 0.0089)	0.0914 (0.089, 0.0943)	0.0701 (0.0685, 0.0717)	0.819 (0.810, 0.831)	3.36 (3.29, 3.44)	4.39 (4.27, 4.53)
LSTM	0.0063 (0.0059, 0.0068)	0.0796 (0.0769, 0.0823)	0.0584 (0.0570, 0.0598)	0.8629 (0.8527, 0.8725)	2.74 (2.68, 2.81)	3.74 (3.62, 3,87)
Linear	0.0814 (0.0457, 0.1294)	0.2852 (0.2138, 0.3598)	0.0961 (0.0896, 0.1039)	−0.7649 (−1.8008, 0.0428)	4.61 (4.30, 4.99)	13.69 (10.26, 17.27)

***Table 2:* Model error metrics.**

Time intervals	MAE (95% CI)	RMSE (95% CI)	R squared (95% CI)	MAE (ALSFRS-R scale) (95% CI)	RMSE (ALSFRS-R scale) (95% CI)
0–12 months	0.0545 (0.0523, 0.0567)	0.0752 (0.0711, 0.0791)	0.8150 (0.7942, 0.8346)	2.62 (2.51, 2,72)	3.61 (3.41, 3,80)
12–24 months	0.0571 (0.0543, 0.0600)	0.0819 (0.0760, 0.0880)	0.8370 (0.8126, 0.8595)	2.74 (2.61, 2.88)	3.93 (3.65, 4.22)
24–48 months	0.0514 (0.0486, 0.0543)	0.0757 (0.0698, 0.0816)	0.8860 (0.8671, 0.9038)	2.47 (2.33, 2.61)	3.63 (3.35, 3.92)
>48 months	0.0415 (0.0382, 0.0451)	0.0601 (0.0539, 0.0670)	0.9332 (0.9166, 0.9472)	1.99 (1.83, 2.16)	2.88 (2.59, 3.22)

Table 3: **Model error metrics across time intervals.**

One way of determining the clinical utility of this model's predictions is by examining them through the established metrics of 'minimal detectable change' (MDC)[29] and 'minimal important difference' (MID).[30] MDC is a statistical representation of the limitations of the scale itself in detecting change which exceeds measurement error whilst MID is the minimal amount of change in an outcome measure that is perceived by plwALS or their clinicians. For the ALSFRS-R, recent work has idenitifed 1.59 points as the MDC and 3.24 points as the MID.[31] Our MAE of 2.65 (95% CI 2.63, 2.76) sits between these numbers, suggesting that the model's average prediction error is both plausible within the measurement structure whilst still smaller than the level of change which would be regarded as clinically meaningful by plwALS or their clinicans.

Extensive epidemiological work has gone into identifying the clinical determinants of disease progression in ALS.[18,32–35] Sex, age at diagnosis, genetic status, diagnostic delay and site of onset are all associated with differing clinical progression and their interaction is often important. This model seems to perform with a consistent level of error across these clinical variables, and across the overall disease duration.

The model also seems to generalise to unseen data. When applied to the PROACT dataset, a clinical trial dataset, the model performance improved. This likely represents the fact that those chosen for clinical trials tend to have different, and often more favourable, survival characteristics than those collected in a prospective, whole-population manner as in the case of PRECISION-ALS. Previous work has shown that in ALS clinical trial cohorts tend to be younger, have more spinal-onset disease and progress more slowly.[16–18] These cohort differences are demonstrated in our datasets in table one with a lower median age in PROACT (54.36 years versus PRECISION 63.55 years) and a difference in site of onset (82.03% spinal PROACT versus 63.18% spinal PRECISION). These factors suggest reduced heterogeneity of the PROACT dataset which was more easily predicted and as such all performance metrics showed improvement on out of cohort validation.

When the specific sub scores of the ALSFRS-R were incorporated into a FCNN, rather than just the total and time as initially presented, all performance metrics declined but still performed better than the linear approximation of the next point. This demonstrates the difficulty in balancing the constraints of higher dimensional inputs (subdomains), limited longitudinal follow-up, and a relatively small database in the wider context of machine learning research.

Adding additional sub domains allowed the model to memorise patient-specific trajectories, resulting in overfitting and a lack of ability to generalise to unseen data. This is a manifestation of the classical bias/variance trade off in machine learning.[36] Whilst more recent literature has called into question the U-shaped risk curve of this trade off in neural networks,[37] this work is often performed in much larger datasets with more complex neural network architectures such as convolutional neural networks in the context of imaging analysis.[38,39] This is an example of some of the real-life problems in modelling complex health data, where the modelling architecture and data structure must be simultaneously observed, understood, and adjusted to provide meaningful predictions. With growing, multimodal datasets in ALS, such as those envisioned by PRECISION ALS, these more complex architectures may become more feasible however were not able to be meaningfully applied for this task due to the interaction between model capacity and data sparsity.

The ALSFRS-R is an imperfect scale. However, it measures clinical outcomes that matter to plwALS and that continue to be applied in a clinical trial context. Factors such as climbing stairs, dressing yourself and turning in bed matter to plwALS and have been associated with survival and quality of life.[40] The model presented here better utilises decades of natural history data collected and harmonised to support real-time decision making, both in trials and in the clinic.

Limitations

The data here are sparse and irregularly spaced, the rolling window approach aims to mitigate this however missing data in real-world registries remains an inherent challenge. The plwALS included in the testing, training and validation are significantly biased towards white Europeans, and this should be considered in the interpretation of the results. We do not present decision thresholds here as further validation would be required before this is achieved. Whilst population-based data capture was used, it is also the case that those more able to participate in longitudinal testing are likely to have been overrepresented and this should be considered.

Conclusion

Decades of population-based natural history information has been captured by specialists across the world about how ALS and the ALSFRS-R progresses. Considerable heterogeneity of the trajectories presents a statistical challenge for clinicians and in clinical trials. Here we have shown how a dynamic neural network approach can capture this heterogenicity and provide individual patient-level predictions. This allows us to leverage extant population data to support clinical decision making in trials and at the bedside. Models such as the one presented here are becoming increasingly possible through the growing power of our ALS patient databases and international collaboration. With larger and more multimodal datasets, it is likely that more complex deep learning is possible.

Plain language summary

ALS progresses differently in every person, making it difficult to predict how quickly a patient's function and ALSFRS-R score will decline. Using 5319 patient's ALSFRS-R trajectories, we trained a simple type of neural network to forcast short term changes in functional ability. This model was tested both on unseen population data and unseen clinical trial databases as validation of performance. The model predicted more accurately than linear methods, even when information was limited or irregularly spaced. Importantly, prediction accuracy was not affected by genetic, demographic or clinical variables. This demonstrates that routinely collected patient data can be used to build generalisable, individualised prediction tools using neural networks, capturing natural history data. This may help clinicians and trial invesitigators identify patients whose disease course deviates from expected patterns and therefore respond earlier or perform further investigation.

Contributors

All authors have verified and have direct access to the raw data used in this study. All authors have read and approved the final version of the manuscript.

Robert McFarlane: conceptualisation, literature search, data curation, formal analysis, funding acquisition, investigation, methodology, resources, validation, visualisation, writing—original draft, and writing—review & editing.

Robert Ross: conceptualisation, investigation, methodology, project administration, resources, supervision, visualisation, writing– review & editing.

Éanna Mac Domhnaill: data curation, investigation, project administration, resources, validation, writing—original draft, and writing—review & editing.

Adriano Chiò: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Philippe Corcia: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Caroline Ingre: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Christopher J McDermott: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Mónica Povedano Panadés: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Pamela J Shaw: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Philip van Damme: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Leonard van den Berg: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Ammar Al-Chalabi: conceptualisation, data curation, funding acquisition, investigation, project administration, resources, and writing—review & editing.

Cathal Walsh: conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, resources, supervision, validation, visualisation, writing—original draft, and writing—review & editing.

Orla Hardiman: conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, writing–original draft and writing—review & editing.

Data sharing statement

The full model weights are available to any researcher after application to the PRECISION ALS scientific steering committee (www.precisionals.ie). Training data is held and managed by the PRECISION ALS consortium.

Declaration of interests

Robert McFarlane: Target ALS neurology residents grant (NR-2024-NRG-S1).

Robert Ross: none declared.

Éanna Mac Domhnaill: none declared.

Adriano Chiò: has received consulting fees from Biogen, Zambon, Mitsubishi and Roche. Has received an honorarium from Zambon. Participates in the DSMB/Advistory role for Verge Corcept and Eli Lilly.

Philippe Corcia: none declared.

Caroline Ingre: has consulted for Cytokinetics, Pfizer, BioArctic, Novartis, Tikomed, Ferrer, Amylyx, Prilenia, Mitsubishi, Abbvie and 2N Pharma. She is also a board member of Tobii Dynavox; all outside the submitted work. Oversight chair for Data Monitoring Committee (EXPERTS-ALS).

Christopher J McDermott: none declared.

Mónica Povedano Panadés: has received consulting fees for Eikonizo Therapeutics Inc. Has received honoraria from Biogen Ferrer and Zambon, paid to her instuition. Has had travel funded to attend a meeting by Italfarmaco SA. Participates in an unpaid advisory capacity to TRICALS and the ELA Law Expert Committee. Is a member of the board of the Funcación Luzón.

Pamela J Shaw: none declared.

Philip van Damme: has received a grant from CSL Behring, paid to insuition. Has received honoraria, payable to insutition, from Biogen, Amylyx and Zambon. Participates in a DSMB/advistory capacity for Biogen, CSL Behring, Alexion Pharmaceuticals, Ferrer, QurAlis, Cytokinetics, Argenx, UCB, Muna Therapeutics, Alector, Augustine Therapetics, VectorY, Sapreme Technologies, Novartis, Prilenia, Trace Neuroscience and NRG Therapetics.

Leonard van den Berg: serves on the scientific advisory board of Novarits, Nurabio, Prilenia, Ferrer, Amylyx, Corcept and Verge.

Ammar Al-Chalabi: Reports grants from the MRC, NIHR, Darby Rimmer Foundation, EU H2020 program, Motor Neurone Disease Association, MyName'5Doddie Foundation, Alan Davison foundation. Has received royalities from Iltoo Pharma. AAC reports consultancies

or advisory boards for Amylyx, Apellis, Biogen, Clene Therapeutics, Cytokinetics, GenieUs, GSK, Lilly, Mitsubishi Tanabe Pharma, Novartis, OrionPharma, Quralis, and Sanofi. Reports Patent WO2024121173A1. AAC in the co-director of the UK MND Research Instuite an is an attendee at the all party parliamentary group on MND (UK).

Cathal Walsh: none declared.

Orla Hardiman: has received consulting fees from Biogen, Novartis, Trace, Uniqure and Medicinova. Has received honoraria for a lecture from Zambon. Has participated in a DSMB/advisory capacity for Uniqure and Medicinova. Prof Hardiman is editor-in-chief of the journal Amyotrophic Lateral Sclerosis and Frontotemporal Degenerations.

Acknowledgements

We thank the funders Taighde Éireann (Research Ireland, formerly Science Foundation Ireland, grant no. 20SP8953) and Target ALS (Grant no. NR-2024-NRG-S1) for their support. We would also like to expressly thank the plwALS who contributed their data to this study.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2025.106029.

References

1 Hardiman O, Al-Chalabi A, Chio A, et al. Amyotrophic lateral sclerosis. *Nat Rev Dis Primers*. 2017;3:17085.
2 Genge A, Cedarbaum JM, Shefner J, et al. *The ALSFRS-R summit: a global call to action on the use of the ALSFRS-R in ALS clinical trials*. Taylor & Francis; 2024.
3 Cedarbaum JM, Stambler N, Malta E, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *J Neurol Sci*. 1999;169(1-2):13–21.
4 Franchignoni F, Mandrioli J, Giordano A, Ferro S, ERRALS Group. A further rasch study confirms that ALSFRS-R does not conform to fundamental measurement requirements. *Amyotroph Lateral Scler Frontotemporal Degener*. 2015;16(5-6):331–337.
5 Ramamoorthy D, Severson K, Ghosh S, et al. Identifying patterns in amyotrophic lateral sclerosis progression from sparse longitudinal data. *Nat Comput Sci*. 2022;2(9):605–616.
6 Bandini A, Rezaei S, Guarín DL, et al. A new dataset for facial motion analysis in individuals with neurological disorders. *IEEE J Biomed Health Inform*. 2020;25(4):1111–1119.
7 Macelaru MH, Chiuzbaian R, Pop P. Machine learning approaches in the detection of amyotrophic lateral sclerosis disease using orofacial gestures. In: *17th International Conference on Agents and Artifical Intelligence*. 3. 2025:1132–1139.
8 Pancotti C, Birolo G, Rollo C, et al. Deep learning methods to predict amyotrophic lateral sclerosis disease progression. *Sci Rep*. 2022;12(1):13738.
9 Turabieh H, Afshar AS, Statland J, Song X, Pooled Resource Open-Access ALS Clinical Trials Consortium*. Towards a machine learning empowered prognostic model for predicting disease progression for amyotrophic lateral sclerosis. *AMIA Annu Symp Proc*. 2023;2023:718–725.
10 Domain adaptation from clinical trials data to the tertiary care clinic–Application to ALS. In: Hadad B, Lerner B, eds. *2020 19th IEEE international conference on machine learning and applications (ICMLA)*. IEEE; 2020.
11 Din Abdul Jabbar MA, Guo L, Nag S, et al. Predicting amyotrophic lateral sclerosis (ALS) progression with machine learning. *Amyotroph Lateral Scler Frontotemporal Degener*. 2024;25(33-44):242–255.
12 Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev*. 1958;65(6):386–408.
13 Guresen E, Kayakutlu G, Daim TU. Using artificial neural network models in stock market index prediction. *Expert Syst Appl*. 2011;38(8):10389–10397.
14 Lam R, Sanchez-Gonzalez A, Willson M, et al. Learning skillful medium-range global weather forecasting. *Science*. 2023;382(6677):1416–1421.
15 Grollemund V, Pradat P-F, Querin G, et al. Machine learning in amyotrophic lateral sclerosis: achievements, pitfalls, and future directions. *Front Neurosci*. 2019;13:135.
16 Kiernan MC, Vucic S, Talbot K, et al. Improving clinical trial outcomes in amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2021;17(2):104–118.
17 Mitsumoto H, Brooks BR, Silani V. Clinical trials in amyotrophic lateral sclerosis: why so many negative trials and how can trials be improved? *Lancet Neurol*. 2014;13(11):1127–1138.
18 Hardiman O, Al-Chalabi A, Brayne C, et al. The changing picture of amyotrophic lateral sclerosis: lessons from European registers. *J Neurol Neurosurg Psychiatry*. 2017;88(7):557–563.
19 McFarlane R, Galvin M, Heverin M, et al. Precision ALS—an integrated pan European patient data platform for ALS. *Amyotroph Lateral Scler Frontotemporal Degener*. 2023;24(5-66):389–393.
20 Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-Scale machine learning on heterogeneous distributed systems. *arXiv*. 2016. preprint arXiv:160304467.
21 Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
22 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–1780.
23 Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H. An empirical comparison of machine learning models for time series forecasting. *Econometric Rev*. 2010;29(5-6):594–621.
24 Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*. 2018.
25 Zhou Y, De S, Ewa G, Perera C, Moessner K. Data-driven air quality characterization for urban environments: a case study. *IEEE Access*. 2018;6:77996–78006.
26 Steinfurth L, Grehl T, Weyen U, et al. Self-assessment of amyotrophic lateral sclerosis functional rating scale on the patient's smartphone proves to be non-inferior to clinic data capture. *Amyotroph Lateral Scler Frontotemporal Degener*. 2025:1–12.
27 Berry JD, Paganoni S, Carlson K, et al. Design and results of a smartphone-based digital phenotyping study to quantify ALS progression. *Ann Clin Transl Neurol*. 2019;6(5):873–881.
28 Folgado D, Rocha PS, Matias P, et al. *HomeSenseALS: a Mobile sensing and self-monitoring system to monitor patients with ALS in everyday life. ACM transactions on computing for healthcare*. 2025.
29 de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes*. 2006;4(1):54.
30 Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–415.
31 Fournier CN, James V, Glass JD. Clinically meaningful change: evaluation of the Rasch-built overall amyotrophic lateral sclerosis disability scale (ROADS) and the ALSFRS-R. *Amyotroph Lateral Scler Frontotemporal Degener*. 2023;24(33-44):311–316.
32 Longinetti E, Fang F. Epidemiology of amyotrophic lateral sclerosis: an update of recent literature. *Curr Opin Neurol*. 2019;32(5):771–776.
33 McFarlane R, Peelo C, Galvin M, Heverin M, Hardiman O. Epidemiologic trends of amyotrophic lateral sclerosis in Ireland, 1996–2021. *Neurology*. 2023;101(19):e1905–e1912.
34 Logroscino G, Traynor BJ, Hardiman O, et al. Incidence of amyotrophic lateral sclerosis in Europe. *J Neurol Neurosurg Psychiatry*. 2010;81(4):385–390.
35 Zou Z-Y, Zhou Z-R, Che C-H, Liu C-Y, He R-L, Huang H-P. Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry*. 2017;88(7):540–549.
36 Geman S, Bienenstock E, Doursat R. Neural networks and the bias/variance dilemma. *Neural Comput*. 1992;4(1):1–58.
37 Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. *arXiv*. 2016.
38 Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25.
39 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. 2014.
40 Kimura F, Fujimura C, Ishida S, et al. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. *Neurology*. 2006;66(2):265–267.