

Machine Learning-Based Scoring System to Predict the Risk and Severity of Ataxic Speech Using Different Speech Tasks

Bipasha Kashyap¹, Student Member, IEEE, Pubudu N. Pathirana², Senior Member, IEEE, Malcolm Horne³, Laura Power, and David J. Szmulewicz⁴

Abstract—The assessment of speech in Cerebellar Ataxia (CA) is time-consuming and requires clinical interpretation. In this study, we introduce a fully automated objective algorithm that uses significant acoustic features from time, spectral, cepstral, and non-linear dynamics present in microphone data obtained from different repeated Consonant-Vowel (C-V) syllable paradigms. The algorithm builds machine-learning models to support a 3-tier diagnostic categorisation for distinguishing Ataxic Speech from healthy speech, rating the severity of Ataxic Speech, and nomogram-based supporting scoring charts for Ataxic Speech diagnosis and severity prediction. The selection of features was accomplished using a combination of mass univariate analysis and elastic net regularization for the binary outcome, while for the ordinal outcome, Spearman's rank-order correlation criterion was employed. The algorithm was developed and evaluated using recordings from 126 participants: 65 individuals with CA and 61 controls (i.e., individuals without ataxia or neurotypical). For Ataxic Speech diagnosis, the reduced feature set yielded an area under the curve (AUC) of 0.97 (95% CI 0.90-1), the sensitivity of 97.43%, specificity of 85.29%, and balanced accuracy of 91.2% in the test dataset. The mean AUC for severity estimation was 0.74 for the test set. The high C-indexes of the prediction nomograms for identifying

the presence of Ataxic Speech (0.96) and estimating its severity (0.81) in the test set indicates the efficacy of this algorithm. Decision curve analysis demonstrated the value of incorporating acoustic features from two repeated C-V syllable paradigms. The strong classification ability of the specified speech features supports the framework's usefulness for identifying and monitoring Ataxic Speech.

Index Terms—Cerebellar ataxia, speech processing, cerebellar dysarthria.

I. INTRODUCTION

THE cerebellum integrates information from a range of sensory inputs with the aim of aiding the production of coordinated movement. Cerebellar Ataxia (CA) refers to the uncoordinated movement resulting from dysfunction of the cerebellum; it is caused by many processes, including neurodegeneration, multiple sclerosis, stroke and trauma. As the cerebellum regulates many aspects of movements, CA results in uncoordinated movements of the limbs, trunk, gait and eyes. Speech is also regulated by the cerebellum and dysfunction in relevant cerebellar regions [1] can result in cerebellar ataxia of speech, sometimes referred to as Ataxic Dysarthria but referred to here as 'Ataxic Speech'. Clinically Ataxic Speech is recognised as increased variability of impaired timing, highly variable syllables (large variation in the duration of syllables and interval between syllables and loudness of individual syllables), articulatory imprecision [2], variations in pause durations [3], [4] and peak amplitude [5].

An important tool for emphasising these features of Ataxic Speech is to ask the subject to repeat a pair of syllables consisting of a Consonant followed by a Vowel (C-V) [6]. Interestingly, at the bedside, the distinction between the speech of individuals with ataxia and controls (i.e., individuals without ataxia or neurotypical) is more readily achieved using C-V repetition than sentence utterances [7], [8]. Most of the studies in the literature used time-based measurements (perturbation measures, jitter, shimmer), frequency features (fundamental frequency, low-to-high frequency components ratio, harmonics-to-noise ratio (HNR) and pitch) and level of sound pressure to distinguish Ataxic Speech from healthy speech [9], [10]. A few studies considered cepstral-spectral measurements [11], [12], which included signal measurements such as the prominence of the cepstral peak (CPP) and smoothed CPP (CPPs). These methods were examined in relatively small cohorts and scored using perceptual assessments:

Manuscript received 28 August 2022; revised 27 February 2023, 16 July 2023, and 7 November 2023; accepted 8 November 2023. Date of publication 20 November 2023; date of current version 13 December 2023. This work was supported by the National Health and Medical Research Council (NHMRC) under Grant GNT1101304. (Corresponding author: Bipasha Kashyap.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Human Research and Ethics Committee, Royal Victorian Eye and Ear Hospital, Australia, under HREC Reference No. 11/994H/16.

Bipasha Kashyap and Pubudu N. Pathirana are with the Networked and Sensing Control (NSC) Laboratory, School of Engineering, Deakin University, Waurin Ponds, VIC 3216, Australia (e-mail: bkashy@deakin.edu.au; pubudu.pathirana@deakin.edu.au).

Malcolm Horne is with the Florey Institute of Neuroscience and Mental Health, Parkville, VIC 3052, Australia (e-mail: malcolm.horne@florey.edu.au).

Laura Power is with the Balance Disorders & Ataxia Service, Royal Victorian Eye and Ear Hospital (RVEEH), East Melbourne, VIC 3002, Australia (e-mail: laura_power@live.com.au).

David J. Szmulewicz is with the Florey Institute of Neuroscience and Mental Health, Parkville, VIC 3052, Australia, also with the Balance Disorders & Ataxia Service, Royal Victorian Eye and Ear Hospital (RVEEH), East Melbourne, VIC 3002, Australia, and also with Cerebellar Ataxia Clinic, Alfred Hospital, Prahran, VIC 3004, Australia (e-mail: dsz@me.com).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNSRE.2023.3334718>, provided by the authors. Digital Object Identifier 10.1109/TNSRE.2023.3334718

illuminate unique features of Ataxic Speech that may not be as readily discernible through a single speech task. By combining the outcomes of these two tasks, the identification of Ataxic Speech and its severity assessment was expected to be enhanced. The features were then used in nomograms to generate interpretive visual scores, that would indicate the likelihood of Ataxic Speech being present or absent in a particular individual and their severity.

To summarise, in this study, four speech domains (time, spectral and cepstral domains along with non-linear dynamics) were quantitatively assessed by objective acoustic analyses to determine specific dysarthric features and estimate their reliability in separating Ataxic Speech from control subjects' speech and estimating the severity of Ataxic Speech, expressing both in terms of probability. The aims of this study can be summarised as follows:

- 1) Examine speech features in the temporal, spectral, cepstral and non-linear dynamics domains to extract a complementary dataset that effectively captures different signalling attributes of Ataxic Speech.
- 2) Assess whether the combined use of linear and non-linear speech measures improves the distinction of Ataxic Speech from control subjects' speech and its severity.
- 3) Investigate the potential effectiveness of two separate speech tasks in the assessment of Ataxic Speech, with the objective of describing variations in Ataxic Speech features across tasks.
- 4) Design a fully automated quantitative system based on machine learning to separate Ataxic Speech from control speech and develop interpretive visual scoring charts (using nomograms) for estimating the probability of Ataxic Speech risk and Ataxic Speech severity. The nomograms are designed based on the binary and ordinal logistic regression models.

II. MATERIALS AND METHODS

A. Participants and Speech Corpus

Speech data from 126 Australian native English speakers were collected. The patients were 65 individuals previously diagnosed with ataxia. Our study specifically investigated Pure (central) CA [22], CA with Bilateral Vestibulopathy (CABV) [23], [24], and CABV with Somatosensory impairment [25]. Additionally, all patients in our study exhibited pure CA without any co-occurring dysarthria. The controls were 61 volunteers with no history of speech difficulties or neurological disease. The demographics of the participants are summarised in Table I. None of the participants (with ataxia or without ataxia/neurotypical) had undergone speech therapy prior to the study.

B. Ethics Approval and Informed Consent

This study was approved by the Human Research and Ethics Committee, Royal Victorian Eye and Ear Hospital, Australia (HREC Reference Number: 11/994H/16) and funded by the National Health and Medical Research Council (NHMRC) Grant: GNT1101304. The study was conducted according to

the NHMRC's "Australian Code for the Responsible Conduct of Research, 2018" [28] and written consent was sought from all the participants prior to their enrolment. The subject shown in Figure 1 provided informed consent to publish their image.

C. Comprehensive Objective Speech Assessment System

The Ataxic Speech assessment was performed using the following steps:

- 1) *Speech Inputs* Participants performed the following two speech tasks:
 - a) **Speech Task 1:** Participants were instructed to recite the phrase *British Constitution* (BC) three times. This phrase is a well-established benchmark for assessing the characteristics of Ataxic Speech. The acoustic measurements were obtained by taking the average of the three recorded instances for each individual.
 - b) **Speech Task 2:** Participants repeated the syllable */ta/* continuously for five seconds, producing a syllabic sequence of */ta/-ta/-ta/*. This task will be referred to as the *Repeated Ta* (RT) task.

Speech intelligibility in the individuals with ataxia was perceptually scored by an experienced clinician in accordance with the Scale for Assessment and Rating of Ataxia (SARA) [29] from the subject's performance of the above two tasks. The SARA speech item ranges from 0 to 6 indicating the severity of Ataxic Speech where "0" is normal speech, "1" is disturbed speech, "2" is distorted speech, but simple to comprehend, "3" is where it is sometimes difficult to understand words, "4" is where it is difficult to understand several words, "5" represents only single words being comprehensible and "6" represents unintelligible speech. The evaluation of the SARA speech item was performed by the same clinician for all patients. The evaluator was blinded to other patients' conditions and was not informed about the specific diagnosis or treatment status of the patients to reduce the risk of bias and ensure that the evaluations were as objective as possible.

- 2) The speech recordings were *captured* by a condenser microphone clipped at an average distance of 10 cm from the subject's lips, in a quiet room with low ambient noise. The recording was conducted using the BioKinMobiTM [30] application on an Android phone, under the supervision of a trained investigator. These speech tasks resulted in 252 speech recordings, with 126 recordings corresponding to each speech task.
- 3) Wireless transmission occurred to a blockchain based distributed *cloud* network [31] where the proposed *machine learning scoring framework* (Section II-D) was applied.
- 4) Data analysis results are transformed into a *clinically relevant format*.

A pictorial representation of the assessment platform is illustrated in Figure 1(A).

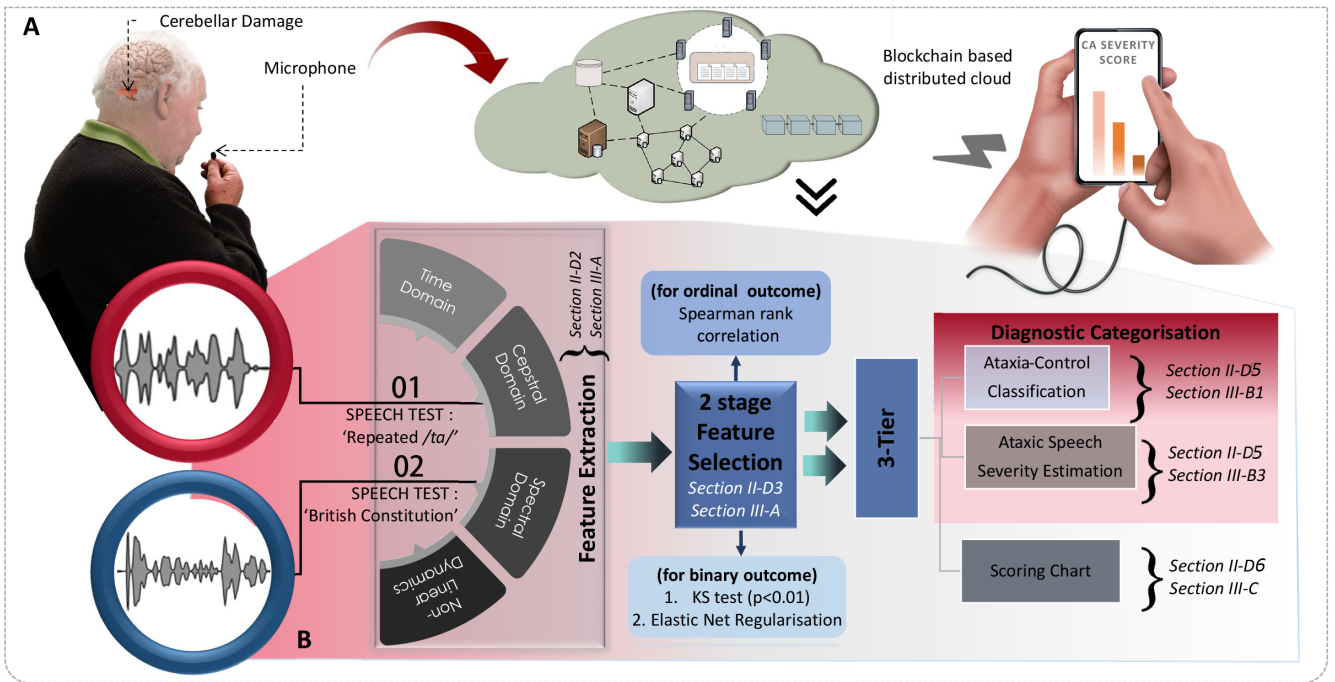


Fig. 1. A. Comprehensive Objective Speech Assessment System, B. Workflow of the machine learning scoring algorithm.

D. Machine Learning Scoring Framework

In this study, a fully automated framework (Figure 1(B)) for the scoring system was developed through the following stages.

1) **Data Representation**: Let M ($M = 126$) be the total number of individuals enrolled in this study. The k^{th} individual, where $1 \leq k \leq M$, is instructed to perform T ($T = 2$) speech tasks in a day. Distinctive acoustic features are extracted from each speech task. Further, the j^{th} speech task, where $1 \leq j \leq T$, consists of N_j acoustic features, where N_j varies between 24(RT) and 77(BC). Let $\sum_{j=1}^T N_j = F$ such that $F = 101$, that is the total number of features extracted in this study. For each individual k , the feature set consisting of all the features extracted from j speech tasks can be represented through the row matrix,

$$X_k = [x_{k11} x_{k12} \dots x_{k1N_1} x_{k21} x_{k22} \dots x_{k2N_2}] \quad (1)$$

where $X_k \in R^{1 \times (N_1 + N_2 + \dots + N_T)}$. If the i^{th} feature of the j^{th} speech task is measured for the k^{th} individual, then his every feature can be statistically represented as x_{kji} , where $1 \leq k \leq M$, $1 \leq j \leq T$ and $1 \leq i \leq N_T$. Therefore, the total feature vector X for M individuals in this study has a size of $M \times F$. A model consisting of M individuals, the input matrix feature set X , and the output response level set Y can be denoted as column matrices,

$$X = [X_1 X_2 \dots X_k \dots X_M]^T, \quad Y = [Y_1 Y_2 \dots Y_k \dots Y_M]^T. \quad (2)$$

The dimensions of X are $M \times (N_1 + N_2 + \dots + N_T)$, the dimensions of responses Y are $M \times 1$ and the dimensions of the obtained dataset combining features and responses are $M \times ((N_1 + N_2 + \dots + N_T) + 1)$.

The algorithm's output, Y_k , can be either binary or dichotomous, used for Ataxia-Control classification, or ordinal for determining the severity of Ataxic Speech, based on SARA speech scores. In the upcoming section, a machine-learning algorithm using the framework illustrated in Figure 1 (B) is presented.

2) **Feature Set Construction**: In this study, the time-based features were extracted from speech task 1 [10]; the spectral and cepstral [13] features were extracted from speech task 2; and the traditional and non-linear features were extracted from both speech tasks (1 & 2). Supporting Document Table I presents a brief description of all 101 acoustic features extracted in this study.

a) **Time-based features**: Let S_k be of the set of (RT) syllables for an individual k , where $S_k = [S_{k1}, S_{k2}, \dots, S_{kn}]$. Similarly, their corresponding measures of full prominence, widths/ time-duration at half prominence, and position of peaks are denoted by P_k, Wh_k, P_{kk} respectively, where $P_k = [P_{k1}, P_{k2}, \dots, P_{kn}]$, $Wh = [Wh_{k1}, Wh_{k2}, \dots, Wh_{kn}]$, $P_{kk} = [(P_{kk1}, Tp_{k1}), (P_{kk2}, Tp_{k2}), \dots, (P_{kkn}, Tp_{kn})]$, P_{kk} represents the elevation and Tp_k represents the time-point. The variable n varies from one RT recording of an individual to another.

Six acoustic measures, namely, RT Duration Regularity (RT_Dr50), RT Gap Regularity (RT_Gr), Average RT Peak Prominence (RT_PPa), Average RT Compensation (RT_Ca), RT Damping Ratio (RT_DR75) and RT Resonant Frequency (RT_RF50) were extracted from the Repeated Ta (RT) data for an individual k according to the Topographic Prominence based algorithm of our previous study [10].

b) **Cepstral features**: The cepstral features included in this study were those that proved pivotal in the diagnosis

of ataxic speech in a related study [13]; they included 12 magnitude cepstrum features from mel-frequency cepstral coefficients (MFCCs) [32] (here referred as $BC_MFCC1, BC_MFCC2, \dots, BC_MFCC12$) and 11 phase cepstrum features from modified group delay cepstral coefficients (MGDCCs) (here referred to as $BC_MGDCC1, BC_MGDCC2, \dots, BC_MFCC12$) [13].

c) Spectral features: Spectral descriptors [14], [15] have been widely used for the perceptual analysis of audio signals in machine learning applications. The spectral features used in this study to extract features for an individual k are summarised in Supporting Document Table I along with their brief mathematical description. For each of these features extracted from the RT and BC data, we further calculated the descriptive statistics (mean (M), standard deviation (SD) and interquartile range (IQR)).

d) Classical features: Fundamental frequency and parameters describing the variability of frequency in time (jitter) [33], [34], amplitude disruptions (shimmer) [35] and pitch perturbations [36] are the most common features describing disturbances in the vibratory characteristic of the vocal folds. These classical characteristics are good candidates, for speech tremor quantification [37]. We included harmonic ratio, shimmer, jitter and their variants [9], [38] in this study.

e) Non-linear dynamic features: Abnormal speech patterns can lead to significant variations in vocal fold tension, resulting in irregular, aperiodic, and noisy-like voices. This deviation may involve sub-harmonics and chaos, which can obstruct the analysis of conventional speech signal processing techniques. Nevertheless, recent studies have indicated that nonlinear dynamical analysis can adequately explain these signal patterns [26], [27].

- Largest Lyapunov Exponent (LLE): In accordance with the Rosenstein equation, LLE is calculated as the average divergence rate of neighbouring trajectories in the attractor. For this algorithm, it is appropriate to estimate the closest neighbours to each point in the trajectories; a neighbour must complete a temporal separation greater than the “period” of the time series to be known as the nearest neighbour [27]. It is possible to say that the point separation in a trajectory is in accordance with the following expression.

$$d(t)_k = Ce^{\lambda_1 t}, \quad (3)$$

where λ_1 is the maximum Lyapunov exponent, $d(t)$ is the average divergence taken at the time t , and C is a normalization constant. The following expression can be obtained by assuming that the j^{th} pair of nearest neighbours roughly diverge at a rate of λ_1

$$\ln(d_j(i))_k = \ln(C_j) + \lambda_1(i \Delta t), \quad (4)$$

where λ_1 is the average line slope that occurs on the logarithmic plane when such an expression is drawn. Based on the Rosenstein equation adopted in our study, we also incorporated Liu et al.’s [39] first-order correction method of the exponential divergence of trajectories in state space of the original noisy speech signal to improve the signal-noise ratio while estimating LLE. The largest

Lyapunov exponent could be estimated using nonlinear least squares fitting based on the correction.

- Detrended Fluctuation Analysis (DFA): For the estimation of the scaling exponent α in non-stationary time series, the stochastic portion of the voice signals can be analysed using DFA. Any number on the real line may be inferred by the scaling exponent; however, the representation of this scaling exponent on a finite sliding scale from zero to one will be more convenient; thus we need $g : R \rightarrow (0, 1)$, a mapping function. The logistic function $g(x) = (1 + \exp(-x))^{-1}$ is one such function that finds common use in statistical and pattern recognition applications, so the normalised scaling exponent is,

$$(\alpha_{norm})_k = \frac{1}{1 + \exp(-\alpha_k)}. \quad (5)$$

On this scale, each sound will lie somewhere between the extremes of zero and one, according to the self-similarity properties of the stochastic part of the dynamics. When speech sounds have α_{norm} closer to one, this characterises speech disorders.

3) Feature Selection: In order to select features for large-scale data in a study on Ataxic Speech classification, we employed a two-stage feature selection process (Figure 1 (B)).

a) Stage 1 (for binary outcome): We used a mass univariate approach to eliminate features that were not significantly related to the binary outcome of ataxia or control, as determined in the development dataset. To do this, we conducted a feature-wise KS test to examine whether each feature varied significantly across the two groups. The resulting features were then further pruned using regularisation, with the elastic net method [40], [41] being selected due to its balance between interpretability and parsimony.

b) Stage 2 (for ordinal outcome): To achieve an ordinal outcome in the study, we assessed the correlation between each selected feature and the SARA speech score using Spearman’s rank correlation test. We found that the selected features were significantly correlated with the outcome, but not necessarily with each other. This suggests that the heterogeneous groups of features from the two speech tests may each represent a proportion of the outcome’s variability.

4) Handling Multicollinearity: We also incorporated the Correlation matrix/ Correlation plot and Variation Inflation Factor (VIF) to detect the multicollinearity. An optimal value of $VIF < 10$ is selected in our experiment. The application of the elastic net regularization method in the selection of features for the binary outcome (as discussed in Section II-D.3.a) effectively addressed the issue of multicollinearity in our highly dimensional dataset. This was accomplished by removing highly correlated predictors from the data [42].

5) Model Building and Ataxic Speech Assessment: This section involved designing a 3-Tier Ataxic Speech automated assessment architecture as depicted in Figure 1 (B): (i) classify speech as ataxic or control, (ii) estimate the severity of Ataxic Speech, and then (iii) use a nomogram based scoring chart to predict Ataxic Speech risk and probability of Ataxic Speech severity.

To construct models for (i) and (ii), we employed binary and ordinal logistic regression, respectively. Subsequently, nomograms were developed for (iii) utilizing the binary and ordinal logistic regression models. Nomograms have been widely used in different clinical settings [43], [44], to indicate the probability of an event, such as death or the presence of a disease, primarily by reducing statistical predictive models to a single numerical estimate tailored to the individual patient profile.

6) Regression Models and Nomogram Construction (Scoring Chart): Let Y_k denote the response of an individual k belonging to an outcome category J .

a) Construction of nomogram (binary responses): When estimating the probability of Ataxic Speech over control speech, the outcome categories, $J(J = 2)$, have the responses 1 (ataxic) or 0 (control). The odds of Y_k being equal to either 1 or 0 are denoted, respectively, by,

$$\begin{aligned} P(Y_k = 1) &= \frac{\exp(\alpha + \beta X_k)}{1 + \exp(\alpha + \beta X_k)}, \text{ and} \\ P(Y_k = 0) &= \frac{1}{1 + \exp(\alpha + \beta X_k)} \end{aligned} \quad (6)$$

b) Construction of nomogram (ordinal responses): In our study, the outcome categories, $J(J = 4)$ are assumed to have a natural order and takes four levels, namely, *SeverityLevel* (SL) 0 (nil), 1 (mild), 2 (moderate), 3 (severe).

Therefore, the ordinal logistic regression models take the forms,

$$\begin{aligned} \ln \frac{P(Y_k \leq 1)}{P(Y_k > 1)} &= \ln \left(\frac{\pi_1}{\pi_2 + \pi_3} \right) = \alpha_1 + \beta_1 X_{k1} + \beta_2 X_{k2}, \dots, \beta_p X_{kp} \\ &= \alpha_1 + \beta X_k, \end{aligned} \quad (7)$$

$$\begin{aligned} \ln \frac{P(Y_k \leq 2)}{P(Y_k > 2)} &= \ln \left(\frac{\pi_1 + \pi_2}{\pi_3} \right) = \alpha_2 + \beta_1 X_{k1} + \beta_2 X_{k2}, \dots, \beta_p X_{kp} \\ &= \alpha_2 + \beta X_k, \end{aligned} \quad (8)$$

where $\pi_j, j = 1, 2, 3$ are the category probabilities, he covariate $X_k = [X_{k1}, \dots, X_{kp}]$ consists of the selected p features from the speech recordings of the individual k , the slope, $\beta = [\beta_1, \dots, \beta_p]$ is a vector of regression coefficients and α_j is an intercept, depending on j .

Equations (7)-(8) share the same coefficient β for covariate X_k . However, the intercept α varies and is denoted by different annotations (α_1 to α_3). The intercepts can be used to calculate the predicted probability of patients with a given set of characteristics being in a specific category.

The probability of a patient to be in a category j , where $j = 1, 2, 3$ can be computed as,

$$\begin{aligned} P(Y_k \leq j) &= P(Y_k < j) - P(Y_k < (j - 1)), \\ P(Y_k = 1) &= P(Y_k \leq 1), \end{aligned} \quad (9)$$

where $P(Y_k \leq j)$ is the cumulative probability of Y_k less than or equal to a particular category, $j = 1, \dots, J - 1$.

E. Data Distribution, Statistics and Model Performance Metrics

To evaluate the generalization performance of the proposed framework on previously unseen data, the study utilized a randomly selected 20% (26 participants) holdout subset of the dataset for testing the trained models, while the remaining 80% (100 participants) of the observations were used for development and validation of the models (Table I). All statistical tests were two-tailed, and a p-value < 0.01 was considered statistically significant. To tune the hyperparameters (α and λ) during the elastic net regularization (Section II-D.3.a), a nested cross-validation scheme was applied in the development-validation dataset and the number of folds (k) varied from 2 to 100. The process was iterated 100 times for each fold where the development-validation dataset was randomly permuted in each iteration. All data management and statistical analyses were carried out using the software, MATLAB and R-version 4.1 (R Foundation for Statistical Computing, Vienna, Austria).

The prediction performance of the Ataxic Speech assessment models was evaluated using the concordance index (C-index) value on a scale from 0 to 1 with a 95% confidence interval (CI), as well as the area under the curve (AUC) for the receiver operator characteristic (ROC) plot [45]. Generally, a value of C-index > 0.70 and AUC > 0.80 indicates that the model is good for discrimination. The nomogram was validated by measuring calibration curves both internally (validation set) and externally (test set). The calibration was analysed using the Hosmer-Lemeshow goodness-of-fit test (HL test), which assesses how well the speech pattern in the data under analysis is described; non-significant p-values indicated that the fit of the model was good [46]. For the calibration curves, the results in the development dataset were validated using the bootstrap method (bootstrap = 500). Discrimination between observed and predicted outcomes was also assessed using the Mean Absolute Error (MAE), Mean Squared Error (MSE), Somers' Dxy (Dxy), and Nagelkerke R2 index (R2). A good discriminative model has a high value of R2 (> 0.7) and low values of MSE and MAE. Large values (tending towards -1 or 1) for Somers' Dxy indicate the model has good predictive capacity. Goodness-of-fit tests, such as the likelihood ratio test, show model suitability, and the Wald statistics evaluate the significance of individual independent variables. The Brier score was calculated; complete precision [47] is indicated by a Brier score of 0. Discrimination and calibration [48] does not measure clinical effectiveness, or the potential to make better decisions with a model than without; therefore, we used an alternate method [48] to perform a decision-curve analysis in our study.

III. EXPERIMENTAL RESULTS

This section describes the results of implementing our machine learning framework for Ataxic Speech assessment.

Recently published epidemiological data indicated 26/100,000 children and 2.7/100,000 adults diagnosed with a dominantly inherited cerebellar ataxia [49], [50]. These studies also report the prevalence of recessively inherited cerebellar ataxias as 3.3/100,000. Hence, for a given large effect size (AUC = 0.97 and Cohen's d of 2.7), a minimum

sample size of 50 (for each group) was calculated by power analysis, with the error probability set at 0.05 and a false negative rate set at 0 (that is a power of 1).

To retain more response information while ensuring each response category has a relatively large number of respondents, an ordinal scaled version (four groups) of the original SARA-rated version (five groups), was analysed. The four category ordinal version is grouped as: Group 1(SARA‘0’), Group 2(SARA‘1’), Group 3(SARA‘2’), Group 4(SARA‘3’ & ‘4’) with 17, 17, 16 and 15(10 & 5) ataxic speakers respectively. Our results revealed that there was no significant difference in age between control and ataxia speaker groups in both the development-validation dataset ($t(98) = 1.25$, $p = 0.215$) and the test dataset ($t(24) = 0.87$, $p = 0.393$). Therefore, we concluded that age was not a confounding variable in our study.

A. Feature Extraction and Selection

A total of 101 spectral, cepstral, time-domain, non-linear and traditional features were initially extracted from speech recordings of the two speech tasks. A Correlation heatmap for the original feature set (101 features) and reduced feature sets are provided in Supporting Document Figure 1. Further, a box plot distribution of 101 features is presented in Supporting Document Figure 2, where the blue shaded plots represent the features that were statistically significantly different for ataxic and control (KS test, $p \leq 0.01$) in elastic net regularisation ($\alpha = 0.2$).

This initial feature set when subjected to the two-stage feature selection approach (Section II-D.3) resulted in the following two reduced feature subsets (FS).

- 1) For Ataxia-Control Categorisation (8 features),
FS 1 = $RT_PPa, RT_JitterSD, BC_MGDCC1, BC_JitterRAP_SD, BC_HR_M, BC_SpecRoP_SD, BC_SpecFlux_IQR, BC_SpecKurt_IQR$,
- 2) For Ataxic Speech Severity Estimation (4 features),
FS 2 = $RT_PPa, RT_Gr, BC_SpecKurt_M, BC_SpecSkew_M$.

The feature selection process for Ataxia-Control classification using the elastic net regularisation technique is demonstrated in Figure 2. Figures 2(A) and 2(B) show the trace plots and corresponding cross-validated deviance of the elastic net fit. Figure 2(C) indicates the performance of the α (elastic-net mixing parameter) when varied from 0 to 1. When α increases, the number of selected features reduces. By the principle of parsimony [51], eight features were selected in this step at $k = 10$, $\alpha = 0.9$, λ (regularization penalty) = 0.0609 (Accuracy = 93%, Sensitivity = 94%, Specificity = 91.8%) as indicated by the red dotted line in Figure 2(C, D).

B. Diagnostic Categorisation

1) *Experiment 1 - Ataxia-Control Classification*: The binomial logistic regression ascertained the effects of the feature set FS 1 on the likelihood that participants have Ataxic Speech. The logistic regression model was statistically significant, $\chi^2(18) = 100.77$, $p < 0.001$. The model explained 73% (Nagelkerke R²) of the variance in Ataxic Speech and

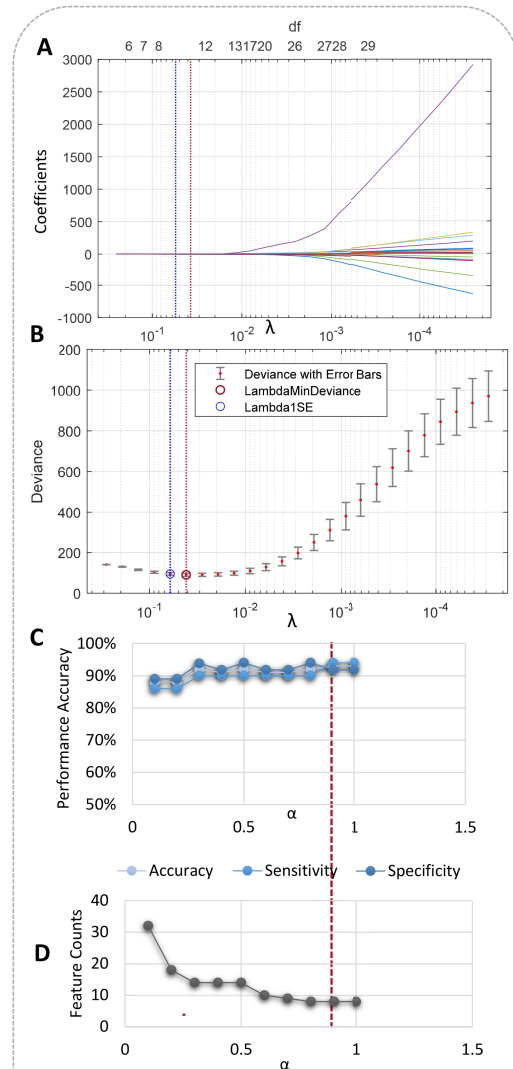


Fig. 2. Elastic net fit plots A. Trace plots of coefficient fit by elastic net B. Cross-validated Deviance of Elastic net fit C. Accuracy estimation and a number of characteristics selected when the alpha elastic-net mixing parameter takes values from 0 to 1 D. Eight features were selected in this step with $\alpha = 0.9$ (Accuracy = 93%, Sensitivity = 94%, Specificity = 91.8%) as indicated by the red dotted line.

correctly classified 90% of cases. In the validation subset, the AUC(ROC) was 0.98 (95% CI 0.89-1), sensitivity was 90%, specificity was 84%, positive predictive value was 89.9%, negative predictive value was 83.3%, and balanced accuracy was 87.1%; in the test set, the model showed similar results (0.97 (95% CI 0.90-1), 97.43%, 85.29%, 84.2%, 96%, and 91.2% respectively). Of the 8 predictor variables, all were statistically significant (as shown in Table III).

2) *Decision Curve Analysis*: The standardised net benefit for Model 1 (all features from FS 1), Model 2 (RT features from FS 1), and Model 3 (BC features from FS 1) were plotted against the threshold probability for categorising a subject's speech as Ataxic Speech (Figure 3(B)). The "all" line shows the net benefit of detecting all subjects with ataxia, and the "none" line is the net benefit of detecting controls. The plot demonstrates that Model 1 is superior to Models 2 and 3 across almost all threshold probabilities (0.1-0.85), with the highest difference at a threshold probability of around 0.6; at that

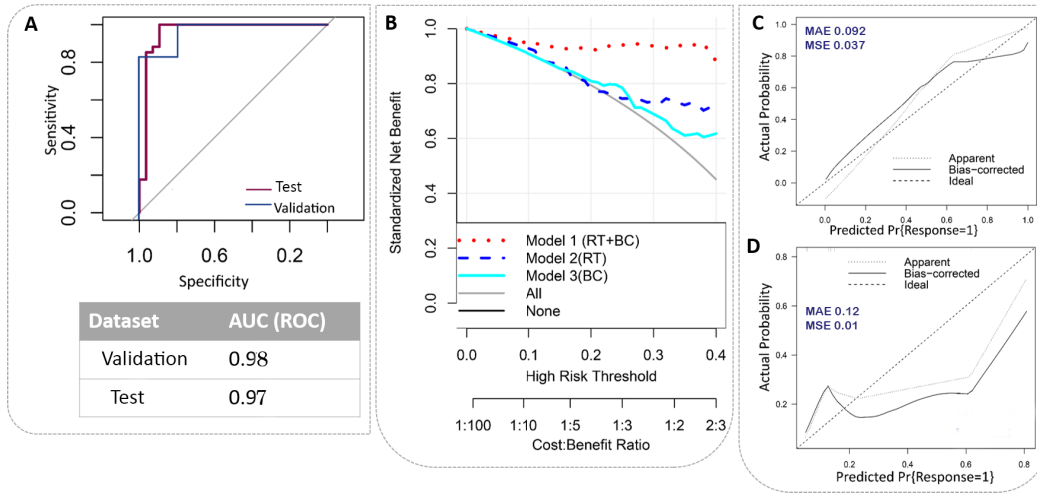


Fig. 3. **A.** Area under the curve (AUC) for the receiver operator characteristic (ROC) plot of Ataxic Speech diagnosis for the validation and test datasets are 0.98 and 0.97 respectively **B.** Decision curve analysis for Model 1 (All features from FS 1), Model 2 (RT features from FS 1), and Model 3 (BC features from FS 1) for diagnosing CA. The three curves were compared to the curves of detecting none (black) and all (grey) individuals with ataxia. The Calibration Curves for Ataxic Speech risk prediction nomogram for **C.** validation dataset **D.** and test dataset. Performance statistics, namely, Mean Absolute Error (MAE) and Mean Squared Error (MSE), are indicated for both the plots (**C.** and **D.**). The x-axis showed the predicted probability of Ataxic Speech severity. The y-axis showed the actual probability of Ataxic Speech severity. The solid line showed the efficiency of the nomogram. An almost close to the diagonal dotted line indicates a good predicted capability.

TABLE III
MODEL STATISTICS FOR ATAXIC SPEECH DIAGNOSTIC

	Unstandardised Coefficients(B)	S.E. ^a	Wald z	Pr(> z)	Odds Ratio	Standardised Coefficients(β)	VIF ^b
Intercept	4.60	3.17	1.45	0.04			
RT_PPa	-5.73	2.67	-2.14	0.03	0.34	-4.91	9.87
RT_JitterSD	-0.0048	0.0023	-2.12	0.03	0.28	-0.0038	9.45
BC_MGDCC1	-6.65	2.30	-2.89	0.003	0.14	-5.15	1.39
BC_JitterRAP_SD	7.45	2.71	2.75	0.006	3.88	6.11	9.11
BC_HR_M	-1.88	3.51	-0.54	0.04	0.81	-0.069	10.07
BC_SpecRoP_SD	-0.006	0.001	-3.27	0.001	0.23	-0.008	8.52
BC_SpecFlux_IQR	0.06	0.02	2.60	0.009	2.61	0.08	5.22
BC_SpecKurt_IQR	0.006	0.004	1.87	0.04	3.03	0.004	2.38
Model Likelihood Ratio Test (χ^2 , d.f., Pr(> χ^2))			100.77, 8, <0.0001				
Discrimination Indexes (R2, Brier)			0.73, 0.09				
Rank Discrimination Indexes (C-index, Dxy)			0.961, 0.922				

Captions: ^a S.E. : Standard Error, ^b VIF : Variance Inflation Factor

threshold, the net benefit (for ataxia speech) is 0.70, 0.20 and 0.80 for Models 2, 3 and 1 respectively. At that threshold, according to the net benefit concept [48], one can administer about $80-70 = 10$ and $80-20 = 60$ more profitable treatments (in every 100 subjects with ataxia) when using Model 1 rather than Model 2 and Model 3 respectively; to those who would otherwise be left untreated (i.e., *net of false positive*). Hence, Model 1 had superior performance compared to Models 2 and 3 as its net benefit surpassed the net benefit of the other two models across the threshold probability range (0.1-0.85). This result indicates the effectiveness of using both speech tasks over only using one for detecting subjects with ataxia.

3) Experiment 2 -Ataxic Speech Severity Estimation: The four acoustic features in FS 2, namely, *RT_PPa*, *RT_Gr*, *BC_SpecKurt_M*, and *BC_SpecSkew_M*. were significantly correlated ($p < 0.01$) with the SARA Speech scores and were included in the severity estimation model design. An increase in the time domain feature (*RT_GR*) extracted from RT was associated with an increase in the odds of considering a high Ataxic Speech severity, with an odds ratio of 1.46, Wald $z = 2.46$, $p = 0.01$ (Table IV).

4) Experiment 3: This experiment sought to explore the effectiveness of a combined feature set, integrating features from our original Ataxic-Control Classification (Experiment 1) and Ataxic Speech Severity Estimation (Experiment 2). This experiment aimed to determine if features selected for severity estimation could also accurately classify ataxic and control speakers. Results from the validation phase showed that the combined features achieved an AUC(ROC) of 0.98 (95% CI 0.88-1), with sensitivity and specificity rates of 88% and 86%, respectively. These rates were comparable to those observed in Experiment 1, which achieved an AUC(ROC) of 0.98 (95% CI 0.89-1), alongside a sensitivity of 90% and a specificity of 84%. However, performance on the test sets revealed a slightly reduced accuracy for Experiment 3, with an AUC(ROC) of 0.95 (95% CI 0.90-1), sensitivity of 92%, and specificity of 83%, compared to an AUC(ROC) of 0.97 (95% CI 0.90-1), a sensitivity of 97.43%, and specificity of 85.29% in Experiment 1. Notably, the sensitivity was consistently higher in Experiment 1 across both sets. Considering the critical balance between sensitivity and specificity in medical diagnostics, where high sensitivity is paramount for effective

TABLE IV
MODEL STATISTICS FOR ATAXIC SPEECH SEVERITY PREDICTION

	Unstandardised Coefficients(B)	S.E. ^a	Wald z	Pr(> z)	Odds Ratio	Standardised Coefficients(β)	VIF ^b
RT_PPa	-0.81	1.19	-0.68	0.049	0.79	-0.07	9.84
RT_Gr	2.99	1.21	2.46	0.01	1.46	3.87	9.94
BC_SpecKurt_M	-3.99	5.38	-0.74	0.045	0.63	-0.99	4.82
BC_SpecSkew_M	7.70	5.51	1.40	0.002	3.17	9.67	4.58
Intercept, $y \geq 1$	-1.54	0.71	-2.16	0.03			
Intercept, $y \geq 2$	-3.03	0.79	-3.80	0.0001			
Intercept, $y = 3$	-4.94	1.01	-4.90	<0.0001			
Model Likelihood Ratio Test (χ^2 , d.f. ^a , Pr(> χ^2))			30.58, 5, <0.0001				
Discrimination Indexes (R2, Brier)			0.526, 0.157				
Rank Discrimination Indexes (C-index, Dxy)			0.812, 0.623				

Captions: ^a d.f.: degree of freedom, ^b S.E. : Standard Error, ^c y : AS (Ataxic Speech) Severity, ^d VIF : Variance Inflation Factor

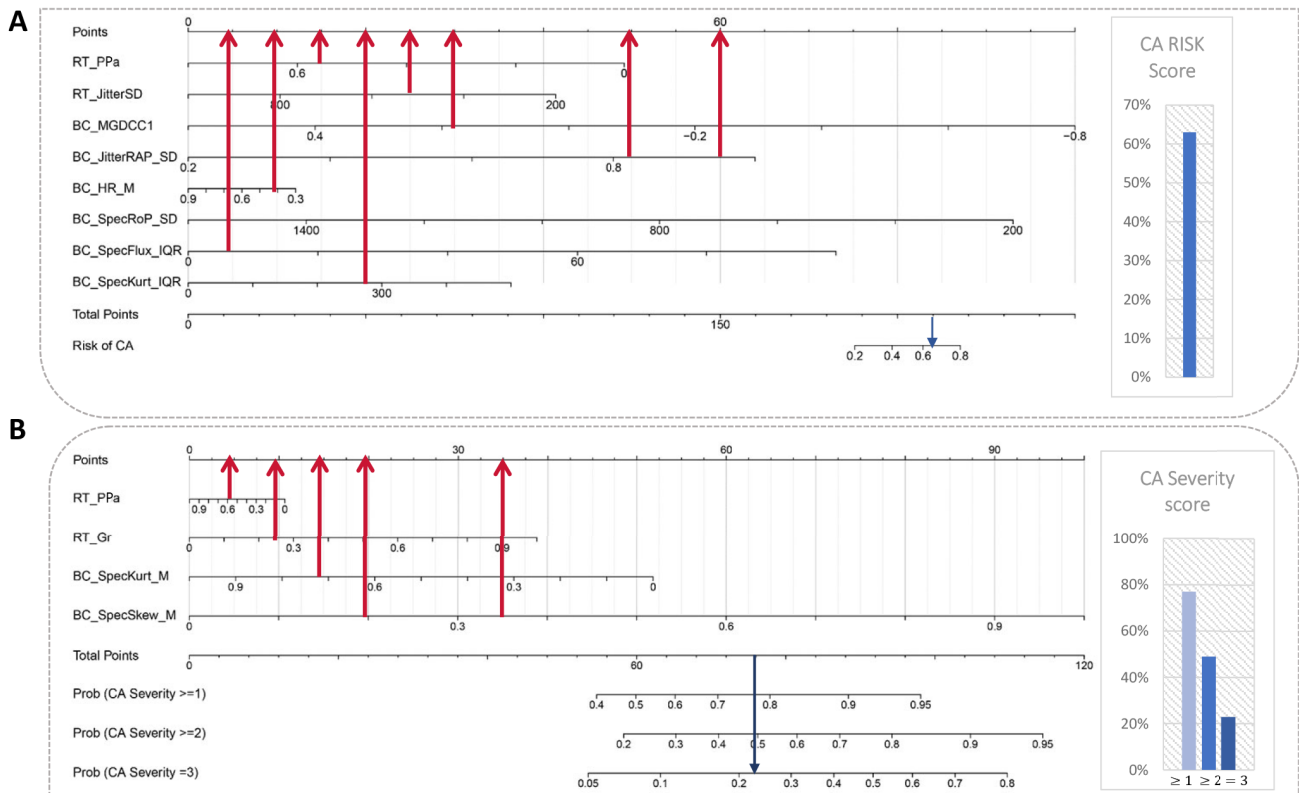


Fig. 4. Nomogram scoring charts for predicting the probability of **A.** Ataxic Speech risk and **B.** Ataxic Speech Severity. To use the nomogram scoring chart, an individual subject’s value from the test dataset was presented on each variable axis, and a vertical line (in red) is plotted upward to calculate the number of “Points” received corresponding to each variable value. The sum of these scores is located on the total points axis and draws a line vertically down (in blue) to find the probability of Ataxic Speech risk or Ataxic Speech severity greater than or equal to a specific category. We demonstrate the use of the chart with an empirical example. **A.** A subject from the validation cohort with speech features located on the variable axes of the chart corresponds to the following values on axis Points, as such (a) $RT_PPa = 15, RT_JitterSD = 20, BC_MGDCC1 = 30, BC_JitterRAP_SD = 50, BC_HR_M = 60, BC_SpecRoP_SD = 10, BC_SpecFlux_IQR = 5, BC_SpecKurt_IQR = 20$, which adds up Total Points (250) and their Ataxic Speech risk probability is more than 0.63; indeed the subject is with ataxia as confirmed by medical diagnosis. **B.** A subject from the validation cohort with speech features located on the variable axes, $RT_PPa, RT_Gr, BC_SpecKurt_M, BC_SpecSkew_M$; will sum up to Points (5+10+15+20+35) and Total Points (80), and their probability scores to be in CA Severity $\geq 1, \geq 2, \geq 3$ are 0.78, 0.50, 0.23 respectively; their SARA score is 2 as confirmed by clinician’s assessment. The higher colour intensity bar in the CA severity score plot indicates a higher severity of Ataxic Speech.

screening [52], the results suggest that the feature set from Experiment 1 is preferable, particularly given its greater parsimony and efficiency in classification. These findings corroborate the necessity for task-specific optimized feature sets, reinforcing our assertion that separate models are warranted for diagnosis versus severity prediction of ataxic speech. This result aligns with similar observations from previous studies in other domains of Cerebellar ataxia [53], [54], and Friedrich

ataxia [55], where diagnostic and severity prediction tasks have yielded different sets of optimal features.

C. Nomogram Scoring Charts

The nomogram scoring charts for predicting the probability of Ataxic Speech risk and Ataxic Speech Severity for an individual subject from the test dataset were presented in Figure 4.

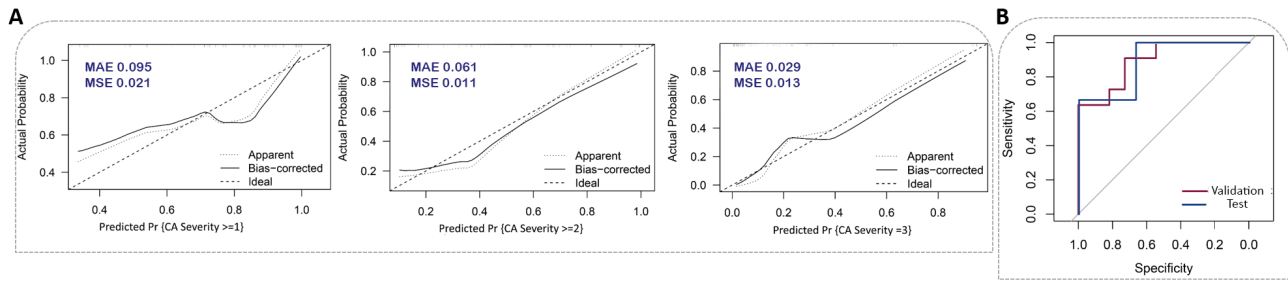


Fig. 5. **A.** The Calibration Curves for the probability of Ataxic Speech Severity prediction nomogram for the validation dataset. The x-axis showed the predicted probability of Ataxic Speech severity. The y-axis showed the actual probability of Ataxic Speech severity. The solid line showed the efficiency of the nomogram. A line close to the diagonal dotted line indicates good predictive capability. **B.** The area under the curve (AUC) for the receiver operator characteristic (ROC) plot of Ataxic Speech Severity prediction for the validation and test datasets are 0.84 and 0.74 respectively.

1) *To Stratify Ataxic Speech Risk:* In the test dataset, the nomogram was used to evaluate the risk of Ataxic Speech for all subjects; the probability indicated by the nomogram was then compared with the probability of the regression model. For the test set, the AUC for the nomogram was 0.95 (95% CI, 0.88-1, $P < 0.0001$) and the AUC for the model was 0.97 (95% CI, 0.90-1, $P < 0.0001$; **Figure 3(A)**), indicating that the logistic regression model and nomogram performed very well in predicting the risk of Ataxic Speech.

a) *Calibration curves:* Evaluation of the nomogram model on the validation set yielded a C-index value 0.98 (95% CI, 0.89-1, $P < 0.0001$, which was identical to the AUC of the regression model (**Figure 3(A)**). The calibration curve for the validation dataset showed a strong agreement with a predictive probability ranging from 0 to 0.7 between nomogram predictions and actual observations (**Figure 3(C)**). Data for the test group yielded a C-index value of 0.96 (95% CI, 0.89-1, $P < 0.0001$), approximately similar to its AUC (**Figure 3(A)**). The MAE and MSE values for the validation dataset and test dataset were (0.092, 0.03) and (0.12, 0.01), respectively. Further, the HL test demonstrated no significant statistical difference between the calibration curves and the ideal curves in both the validation and testing cohorts (**Figure 3(C, D)**)).

2) *To Predict the Probability of Ataxic Speech Severity:* Following the regression analysis in Section III-B.3, the same four acoustic features were incorporated into the nomogram scoring chart to predict the ordinal probabilities. Only Ataxic Speech data were used to construct and calibrate the nomogram. On the basis of the individual scores of these four variables, a user may calculate the total score and obtain a particular probability of Ataxic Speech severity. We illustrate the use of the ordinal nomogram scoring chart, detailed scores of all variables and their interpretation using a participant from the test dataset in **Figure 4**. According to the nomogram, $BC_SpecSkew_M$ had the greatest influence on estimating the probability of the severity of Ataxic Speech, followed by the features $BC_SpecKurt_M$, RT_Gr and RT_PPa . The mean areas ROC curve of the regression models resulting in the Multiclass ROCs in the validation and testing groups were 0.84 and 0.74, respectively (**Figure 5(B)**).

a) *Calibration curves:* The C-index of the nomogram model in the validation group was 0.81 (**Table IV**). In the validation dataset, calibration plots showed better consistency between the nomogram projections and the actual observations for Ataxic Speech Severity ≥ 2 , $= 3$ than Ataxic Speech Severity ≥ 1 (**Figure 5(A)**).

IV. DISCUSSION

Ataxic Speech is a crucial aspect of the clinical manifestation of CA, yet automated methods for Ataxic Speech evaluation are currently lacking. Thus, this study aimed to establish a comprehensive and objective evaluation of Ataxic Speech, providing a probability of its presence and severity using a nomogram connected to the SARA. A total of 126 participants, including 65 with CA and 61 controls, were recruited to obtain speech samples. The algorithm utilized feature selection and elastic-net regularization to develop regression models that distinguish Ataxic Speech from controls and detect severity, followed by nomogram-based reports to predict Ataxic Speech probability and severity. It is important to note that identifying the presence of speech ataxia in individuals with CA and categorizing it from controls does not constitute a clinical diagnosis. Instead, it enables discrimination between controls and those with Ataxic Speech with a wide range of severity levels. Diagnosis of ataxia usually occurs early in the disease when symptoms are mild, and the current tool is not intended for clinical diagnosis. Instead, it could provide insights into developing tools that can differentiate early or emerging Ataxic Speech from unaffected individuals' speech.

In this implementation, we extensively investigated a variety of speech features that have been previously reported as relevant to Ataxic Speech, including time, spectral and cepstral measurements, to extract 101 acoustic features. This allowed us to explore a wide range of acoustic and phonetic characteristics that may be related to Ataxic Speech and select the most informative features. This study also investigated non-linear speech features to determine their relevance to Ataxic Speech specifically. While the non-linear acoustic features have been used previously in the analysis of pathological speech [26], [27], their capability to characterise Ataxic Speech has not been explored. Their effectiveness and complementary behaviour in differentiating individuals with ataxia from controls and objectively estimating Ataxic Speech severity were studied here. While the feature sets FS1 (RT_PPa , $RT_JitterSD$, BC_MGDCC1 , $BC_JitterRAP_SD$, BC_HR_M , $BC_SpecRoP_SD$, $BC_SpecFlux_IQR$, $BC_SpecKurt_IQR$) and FS2 (RT_PPa , RT_Gr , $BC_SpecKurt_M$, $BC_SpecSkew_M$) for diagnosing and predicting the severity of speech ataxia, are indeed distinctive for Ataxic Speech, some of them, such as the spectral and classical features, has also been found to be altered in other types of dysarthria, such as hypokinetic

dysarthria in Parkinson's disease [56]. However, it is important to note that the combination of these features, rather than any individual feature, provides the best separation accuracy between ataxia and controls, suggesting that these features are uniquely informative for Ataxic Speech and capture specific acoustic and phonetic properties that are characteristic of Ataxic Speech, such as changes in pitch, timing, and spectral properties.

Also, the fact that the features leading to the best separation between ataxia and controls are completely different from those selected for the prediction of Ataxic Speech severity may seem counterintuitive. However, this can be explained by the fact that the two tasks have different goals. The goal of the first task (separation between ataxia and controls) was to identify the most distinctive features that can differentiate Ataxic Speech from normal speech. The goal of the second task (prediction of Ataxic Speech severity) was to identify the features that are most strongly associated with the severity of Ataxic Speech, regardless of whether they are also altered in other types of dysarthria. Therefore, it is possible that some of the features that are distinctive for Ataxic Speech do not necessarily correlate with the severity of the disorder, and vice versa.

Two separate repeated C-V syllable paradigms were assessed with the expectation that variations in Ataxic Speech features might be found in the two tasks. The rationale was that in each speech task, different features of Ataxic Speech may be more pronounced and a cross-task comparison may be helpful in isolating and documenting aspects of abnormalities. The comparative plots for Model 1 (All features from FS 1), Model 2 (RT features from FS 1), Model 3 (BC features from FS 1) demonstrated in the decision curve analysis confirmed this rationale. The Model 1 performance was superior to Models 2 and 3 as its net benefit surpassed the net benefit of the other two models across the threshold probability range (0.1-0.85). This decision curve analysis affirmed the clinical usefulness of our selected model.

In regression models, the actual 'effect' determines the value of the outcome and more specifically, with logistic regression, it is the 'log-odds ratio'. We introduced nomogram-based scoring models and charts in our study to accentuate this idea and translate it to an easily interpretable visual score. Two predictive nomograms were developed using the independent factors from feature sets FS 1 and FS 2 to generate indicators for estimating the risk of Ataxic Speech and its severity, respectively. In order to predict disease severity, as well as repeatability and reliability in different populations, severity scoring systems should be valid, calibrated and discriminated against. The supporting predictive scoring chart in our study will provide a combined quantitative tool for clinicians to assess the risk of Ataxic Speech and the individual probability of its severity. Further, to test and validate the prognostic accuracy of the nomogram model, adequate discrimination and calibration were performed. Our assessment models performed better than the objective assessments of Ataxic Speech from previous literature as depicted in Table II.

In conclusion, this study demonstrates the effectiveness of incorporating the complementary behaviour of objective

speech measures extracted from the four domains (time, spectral, cepstral and non-linear dynamics) in differentiating individuals with ataxia from controls and objectively estimating Ataxic Speech severity. Additionally, the use of a combination of speech features from different speech tasks highlighted specific aspects of Ataxic Speech less easily identified by a single task; and improved the identification of Ataxic Speech and the estimation of its severity. The findings show that the automated analysis of meaningful acoustic features from recordings of the two repeated C-V syllable paradigms selected (RT and BC) can be a reliable tool for monitoring CA-associated vocalisation deficits. Furthermore, nomogram based scoring charts assist in offering an accurate individualized prediction of the presence of Ataxic Speech and its severity while highlighting the important prognostic information that can be gleaned from simple speech tests. We believe that this technique can be further developed and translated for other motor evaluations as well as into the entire spectrum of motor speech disorders manifested by other neurodegenerative disorders. The presented pilot study results offer new possibilities for future research on motor speech disorders, ranging from conventional laboratory-based analyses; and monitoring the impact of therapy and progression of longitudinal disease to high-throughput screening possibilities.

ACKNOWLEDGMENT

The authors would like to thank the patients from Royal Victorian Eye and Ear Hospital (RVEEH), St Andrews Place, East Melbourne, for their participation in their clinical trials, they make this important research possible.

REFERENCES

- [1] W. Ziegler and H. Ackermann, "Subcortical contributions to motor speech: Phylogenetic, developmental, clinical," *Trends Neurosciences*, vol. 40, no. 8, pp. 458–468, Aug. 2017.
- [2] J. R. Duffy, *Motor Speech Disorders-E-Book: Substrates, Differential Diagnosis, and Management*. Amsterdam, The Netherlands: Elsevier, 2012.
- [3] B. Brendel et al., "Comparing speech characteristics in spinocerebellar ataxias type 3 and type 6 with friedreich ataxia," *J. Neurol.*, vol. 262, no. 1, pp. 21–26, Jan. 2015.
- [4] H. Ackermann and I. Hertrich, "Speech rate and rhythm in cerebellar dysarthria: An acoustic analysis of syllabic timing," *Folia Phoniatrica et Logopaedica*, vol. 46, no. 2, pp. 70–78, 1994.
- [5] R. D. Kent, J. F. Kent, J. R. Duffy, J. E. Thomas, G. Weismer, and S. Stuntebeck, "Ataxic dysarthria," *J. Speech, Lang., Hearing Res.*, vol. 43, no. 5, pp. 1275–1289, 2000.
- [6] K. Rozenstoks, M. Novotny, D. Horakova, and J. Ruzs, "Automated assessment of oral diadochokinesis in multiple sclerosis using a neural network approach: Effect of different syllable repetition paradigms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 32–41, Jan. 2020.
- [7] H. Ackermann and I. Hertrich, "Dysarthria in Friedreich's ataxia: Timing of speech segments," *Clin. Linguistics Phonetics*, vol. 7, no. 1, pp. 75–91, Jan. 1993.
- [8] W. Ziegler and K. Wessel, "Speech timing in ataxic disorders: Sentence production and rapid repetitive articulation," *Neurology*, vol. 47, no. 1, pp. 208–214, Jul. 1996.
- [9] E. Schalling, B. Hammarberg, and L. Hartelius, "A longitudinal study of dysarthria in spinocerebellar ataxia (SCA): Aspects of articulation, prosody, and voice," *J. Med. Speech-Lang. Pathol.*, vol. 16, no. 2, pp. 103–118, 2008.
- [10] B. Kashyap, M. Horne, P. N. Pathirana, L. Power, and D. Szmulewicz, "Automated topographic prominence based quantitative assessment of speech timing in cerebellar ataxia," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101759.

- [11] S. Jannetts and A. Lowit, "Cepstral analysis of hypokinetic and ataxic voices: Correlations with perceptual and other acoustic measures," *J. Voice*, vol. 28, no. 6, pp. 673–680, Nov. 2014.
- [12] S. Luna-Webb, "Comparison of acoustic measures in discriminating between those with Friedreich's ataxia and neurologically normal peers," 2015.
- [13] B. Kashyap, P. N. Pathirana, M. Horne, L. Power, and D. Szmulewicz, "Quantitative assessment of speech in cerebellar ataxia using magnitude and phase based cepstrum," *Ann. Biomed. Eng.*, vol. 48, no. 4, pp. 1322–1336, Apr. 2020.
- [14] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," CUIDADO 1st Project Rep. 54.0, pp. 1–25, 2004.
- [15] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust ASR," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, May 2004, p. 1-193.
- [16] E. J. Fine, C. C. Ionita, and L. Lohr, "The history of the development of the cerebellar examination," in *Seminars in Neurology*, vol. 22, no. 4. New York, NY, USA: Thieme Medical Publishers, 2002, pp. 375–384.
- [17] S. E. Jelliffe and W. A. White, *Diseases of the Nervous System: A Text-Book of Neurology and Psychiatry*. Philadelphia, PA, USA: Lea & Febiger, 1923.
- [18] K. M. Rosen, J. E. Folker, A. P. Vogel, L. A. Corben, B. E. Murdoch, and M. B. Delatycki, "Longitudinal change in dysarthria associated with friedreich ataxia: A potential clinical endpoint," *J. Neurol.*, vol. 259, no. 11, pp. 2471–2477, Nov. 2012.
- [19] J. Rusz, T. Tykalová, G. Salerno, S. Bancone, J. Scarpelli, and M. T. Pellecchia, "Distinctive speech signature in cerebellar and parkinsonian subtypes of multiple system atrophy," *J. Neurol.*, vol. 266, no. 6, pp. 1394–1404, Jun. 2019.
- [20] J. Rusz et al., "Characteristics of motor speech phenotypes in multiple sclerosis," *Multiple Sclerosis Rel. Disorders*, vol. 19, pp. 62–69, Jan. 2018.
- [21] G. Noffs et al., "Acoustic speech analytics are predictive of cerebellar dysfunction in multiple sclerosis," *Cerebellum*, vol. 19, no. 5, pp. 691–700, Oct. 2020.
- [22] E. Storey, M. Bahlo, M. Fahey, O. Sisson, C. J. Lueck, and R. J. M. Gardner, "A new dominantly inherited pure cerebellar ataxia, SCA 30," *J. Neurol., Neurosurg. Psychiatry*, vol. 80, no. 4, pp. 408–411, Dec. 2008.
- [23] D. D. Pothier, J. A. Rutka, and P. J. Ranalli, "Double impairment: Clinical identification of 33 cases of cerebellar ataxia with bilateral vestibulopathy," *Otolaryngol.-Head Neck Surg.*, vol. 146, no. 5, pp. 804–808, May 2012.
- [24] D. J. Szmulewicz, "Combined central and peripheral degenerative vestibular disorders: CANVAS, idiopathic cerebellar ataxia with bilateral vestibulopathy (CABV) and other differential diagnoses of the CABV phenotype," *Current Otorhinolaryngol. Rep.*, vol. 5, no. 3, pp. 167–174, Sep. 2017.
- [25] D. J. Szmulewicz, C. A. McLean, H. G. MacDougall, L. Roberts, E. Storey, and G. M. Halmagyi, "CANVAS an update: Clinical presentation, investigation and management," *J. Vestibular Res.*, vol. 24, nos. 5–6, pp. 465–474, Dec. 2014.
- [26] P. Henríquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Diaz-de-Maria, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 6, pp. 1186–1195, Aug. 2009.
- [27] J. R. Orozco-Arroyave, J. D. Arias-Londono, J. F. Vargas-Bonilla, and E. Nöth, "Analysis of speech from people with parkinson's disease through nonlinear dynamics," in *Proc. Int. Conf. Nonlinear Speech Process.* New York, NY, USA: Springer, 2013, pp. 112–119.
- [28] U. Australia et al., "Australian code for the responsible conduct of research," Nat. Health Med. Res. Council, Australia, Tech. Rep., 2018.
- [29] T. Schmitz-Hübsch et al., "Scale for the assessment and rating of ataxia: Development of a new clinical scale," *Neurology*, vol. 66, no. 11, pp. 1717–1720, Jun. 2006.
- [30] S. W. Ekanayake, A. J. Morris, M. Forrester, and P. N. Pathirana, "BioKin: An ambulatory platform for gait kinematic and feature assessment," *Healthcare Technol. Lett.*, vol. 2, no. 1, pp. 40–45, Feb. 2015.
- [31] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for secure EHRs sharing of mobile cloud based E-health systems," *IEEE Access*, vol. 7, pp. 66792–66806, 2019.
- [32] F. Haider, S. de la Fuente, and S. Luz, "An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 272–281, Feb. 2020.
- [33] R. J. Moran, R. B. Reilly, P. de Chazal, and P. D. Lacy, "Telephony-based voice pathology assessment using automated speech analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 468–477, Mar. 2006.
- [34] S. Skodda, W. Visser, and U. Schlegel, "Short- and long-term dopaminergic effects on dysarthria in early Parkinson's disease," *J. Neural Transmiss.*, vol. 117, no. 2, pp. 197–205, Feb. 2010.
- [35] K. Umashathy, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 421–430, Mar. 2005.
- [36] L. Daudet, N. Yadav, M. Perez, C. Poellabauer, S. Schneider, and A. Huebner, "Portable mTBI assessment using temporal and frequency analysis of speech," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 2, pp. 496–506, Mar. 2017.
- [37] C. Moers, B. Möbius, F. Rosanowski, E. Nöth, U. Eysholdt, and T. Haderlein, "Vowel- and text-based cepstral analysis of chronic hoarseness," *J. Voice*, vol. 26, no. 4, pp. 416–424, Jul. 2012.
- [38] B. Brendel et al., "Friedreich ataxia: Dysarthria profile and clinical data," *Cerebellum*, vol. 12, no. 4, pp. 475–484, Aug. 2013.
- [39] H.-F. Liu, Z.-H. Dai, W.-F. Li, X. Gong, and Z.-H. Yu, "Noise robust estimates of the largest Lyapunov exponent," *Phys. Lett. A*, vol. 341, nos. 1–4, pp. 119–127, Jun. 2005.
- [40] L. E. Melkumova and S. Y. Shatskikh, "Comparing ridge and LASSO estimators for data analysis," *Proc. Eng.*, vol. 201, pp. 746–755, Jan. 2017.
- [41] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005.
- [42] F. Tomaschek, P. Hendrix, and R. H. Baayen, "Strategies for addressing collinearity in multivariate linguistic data," *J. Phonetics*, vol. 71, pp. 249–267, Nov. 2018.
- [43] S. Buti et al., "Validation of the GRade, age, nodes and tumor (GRANT) score within the surveillance epidemiology and end results (SEER) database: A new tool to predict survival in surgically treated renal cell carcinoma patients," *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, Sep. 2019.
- [44] W. Yanling et al., "Radiomics nomogram analyses for differentiating pneumonia and acute paraquat lung injury," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Oct. 2019.
- [45] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [46] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, vol. 398. Hoboken, NJ, USA: Wiley, 2013.
- [47] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Monthly Weather Rev.*, vol. 78, no. 1, pp. 1–3, Jan. 1950.
- [48] Z. Zhang et al., "Decision curve analysis: A technical note," *Ann. Translational Med.*, vol. 6, no. 15, p. 308, Aug. 2018.
- [49] F. Pilotto and S. Saxena, "Epidemiology of inherited cerebellar ataxias and challenges in clinical research," *Clin. Transl. Neurosci.*, vol. 2, no. 2, Jul. 2018, Art. no. 2514183X18785258.
- [50] M. S. Salman, "Epidemiology of cerebellar diseases and therapeutic approaches," *Cerebellum*, vol. 17, no. 1, pp. 4–11, Feb. 2018.
- [51] M. E. Siddall, "Parsimony analysis," in *Techniques in Molecular Systematics and Evolution*. Basel, Switzerland: Birkhäuser, 2002, pp. 31–54.
- [52] L. D. Maxim, R. Niebo, and M. J. Utell, "Screening tests: A review with examples," *Inhalation Toxicology*, vol. 26, no. 13, pp. 811–828, Nov. 2014.
- [53] T. Ngo et al., "Balance deficits due to cerebellar ataxia: A machine learning and cloud-based approach," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1507–1517, May 2021.
- [54] R. Krishna, P. N. Pathirana, M. Horne, L. Power, and D. J. Szmulewicz, "Quantitative assessment of cerebellar ataxia, through automated limb functional tests," *J. Neuroeng. Rehabil.*, vol. 16, no. 1, pp. 1–15, Dec. 2019.
- [55] R. Krishna, P. N. Pathirana, M. K. Horne, L. A. Corben, and D. J. Szmulewicz, "Quantitative assessment of friedreich ataxia via self-drinking activity," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 1985–1996, Jun. 2021.
- [56] B. Karan, S. S. Sahu, J. R. Orozco-Arroyave, and K. Mahto, "Hilbert spectrum analysis for automatic detection and evaluation of Parkinson's speech," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102050.