A Comparative Severity Assessment of Impaired Balance due to Cerebellar Ataxia using Regression Models

Thang Ngo¹, Student Member, IEEE, Lahiru L. Abeysekara², Member, IEEE, Pubudu N. Pathirana³, Senior Member, IEEE, Malcolm Horne⁴, Laura Power⁵, and David J. Szmulewicz⁶

Abstract—Cerebellar ataxia (CA) refers to the impaired balance and coordination resulting from injury or degeneration of the cerebellum. Testing balance is one of the simplest means of assessing CA. This study compares instrumented assessment and clinical assessment scales of the balance test called Romberg’s test. Inertial Measurement Unit (IMU) data were collected from a sensor attached to their chest of 53 subjects while they performed the test. The corresponding clinical scores were also tabulated. Using this data, 99 features were extracted to quantify acceleration, tremor and displacement of body sway. These features were filtered to identify the subset that better characterizes the distinctive behavior of CA subjects. Elastic Net Regression model resulted a greater agreement (0.70 Pearson coefficient) with the clinical SARA scores. The overall results indicated that data from a single IMU sensor is sufficient to accurately assess balance in CA. The significance of this study is that evaluation of balance using Recurrence Quantification Analysis produces a comprehensive framework for the assessment of CA.

I. INTRODUCTION

Cerebellar Ataxia (CA) refers collectively to problems with balance, coordination and speech that results from injury or degeneration of the cerebellum. Clinically the presence and severity of CA is based on observations of a patient’s performance during a number of simple tasks designed to emphasize the disordered movement. This assessment is quantified using rating scales of the performance of these tests (Scale for Assessment and Rating of Ataxia: SARA) [1]. This assessment is subjective and depends heavily on the clinicians experiences.

CA due to neurodegenerations are relatively uncommon with only a few thousand people of all ages affected in Australia. Australia low population density (3 per sq. km - 2019) combined with the rare disease make it arduous to monitor people with a progressive CA. Rehabilitation may improve postural disorders due to CA [2] and clinical trials also need objective measures of progression. A cloud-based would provide a reliable measure of severity of CA and change over time. Sensors make regular objective monitoring possible using wireless and cloud-based processing.

Apart from clinical scales, postural stability can also be evaluated using a force platform that measure the variability of centre of foot pressure (CoP). However, a limitation of posturography is that the torso is multisegmented and hence not resembling an inverted pendulum always. While wearable sensors have been used to study degenerative CA, the studies [3], [4] are preliminary and limited either by feature extraction techniques or by limitations in correlation between clinical scores from individual tests. This study focuses on providing a reliable score for the severity of the subject similar to the clinical scores.

An IMU-based motion measurement system, such as the instrumented BioKin sensor system and mobile app, provides an affordable, easy to deploy and practical solution to quantify CA in the clinic or at home. Data analysis is performed in the cloud and the results of analysis are sent a mobile via an app. An objective severity score can be used to monitor the disease progression of an individual more effectively. Instrumented measurement eliminates the need for the presence of an experienced clinicians and reduces the subjectiveness of the assessment. Applied machine learning-based algorithms may possess the ability to detect CA at an early stage or certain unique characteristics that may not be clinically observable. Machine-based system may play a significant role in the effective assessment of CA possibly in non-clinical environments. More consistent and less variable assessment will improve rehabilitation programs, time and cost of clinical assessments and disease modifying therapies.

II. MATERIAL

A. Participants

<table>
<thead>
<tr>
<th>TABLE I: Demographic Data of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxia Patients</td>
</tr>
<tr>
<td>Number of subjects (people)</td>
</tr>
<tr>
<td>Ages (years)</td>
</tr>
<tr>
<td>Gender (M/F)</td>
</tr>
<tr>
<td>Dominant limb (left/right)</td>
</tr>
<tr>
<td>Height (centimetre)</td>
</tr>
</tbody>
</table>

Participants were 53 people diagnosed with Cerebellar Ataxia. Demographic details are shown in Table I. This study was approved by the ethics committee of the Human
III. METHOD

A. Signal Pre-processing

Interpolation, fixed-size overlapping, sliding window and low-pass filtering techniques were applied for raw data in the pre-processing stage. Nonlinear techniques typically require 1000 samples [6] to reach a significant level of reliability. Romberg’s data recorded for 30 seconds at 50 Hz resulted in 1500 samples. In occasions where the subject lost balance causing early termination of the test, providing less than 1000 samples were collected, the data from those subjects were interpolated.

B. Feature Engineering

Features of interest were extracted using recurrence quantification analysis (RQA), multiscale entropy (MSE), harmonic ratio (HR), and index of harmonicity (IH) as stability measures. Variability measurements encompass mean, standard deviation (SD), coefficient of variation (CV), the first and second dominant frequency (F1, F2) and powers at the first and second dominant frequency (PO1, PO2). These measurements were subsequently extracted and transferred to the respective feature selection algorithms.

1) Recurrence Quantification Analysis: RQA technique has applications in numerous disciplines [7]. In balance assessments, RQA has mostly been applied to data from standing/sitting force platforms [8]. We implemented RQA with the IMU wearable sensor signals in assessment of CA. Features extracted from recurrence plot include recurrence percentage, determinism, trend, laminarity and entropy.

2) Multi-scale Entropy: MSE has been introduced to quantify regularity or complexity in a time series obtained from IMU sensors [9]. MSE evaluates at multiple time scales of the sample entropy. It has been constructed consecutively to form a more coarse-grained time series. MSE is a function of dimension $m$, time scales $t$ and $r$ parameters. MSE was calculated for $t$ values ranging from 1 to 6, $m = 2$ and $r = 0.15$ as suggested by [10]. Value of $r$ is normally in a range of 0.1 to 0.25 [11], [12] and not strictly constrained to a specific value [9].

3) Harmonic ratio and Index of Harmonic: HR and IH measures based upon harmonic theory to quantify the smoothness of the signal. HR quantifies in frequency spectra as defined by the ratio of the sum of the first 20 even to the first 20 odd harmonics as given by the equation 1. IH is essentially the ratio of the DC component to the combined power of the first harmonics [13] as given in equation 2, where $P_i$ is the power spectral density of frequency $i$.

$$HR = \frac{\sum_{i=2,4,6}^{20} P_i}{\sum_{i=1,3,5}^{20} P_i}$$

$$IH = \frac{P_1}{\sum_{i=1}^{5} P_i}$$
Fig. 2: Boxplots of the top four features selected by RFE scheme. (a) Recurrence Rate from accelerometer (medial-lateral axis). (b) Power of the first dominant frequency of velocity (medial-lateral axis). (c) Coefficient of variation of gyro sensor (superior-inferior axis). (d) Signal to noise of velocity signal (anterior-posterior axis).

C. Feature selection

When an external estimation scale assign weights to features (e.g., the coefficients in a linear model), Recursive Feature Elimination (RFE) is used to select features recursively by considering smaller sets from previous set of features. Firstly, the estimator is trained using the initial features. Then the importance of each feature is calculated via an attributed coefficient or a feature importance attribute [14]. Secondly, the least important features are eliminated from the set. This procedure is recursively repeated on the pruned set until the desired number of features are finally attained.

D. Machine Learning

Machine learning methods were implemented on the Romberg’s test data for estimating severity of ataxia due to Cerebellar dysfunction. The generated regression model candidates included Gradient boosting regression (GraBoRe), elastic net [15], least angle regression (Lars) [16], random forest [17], decision tree, linear regression, and logistic regression. Metrics have been used to evaluate regression models including: median absolute error (MAE) with the best value of zero, explained variance regression score function (EVS) with the best possible score of 1, lower values denote lesser correlation with the clinical score, coefficient of determination \( R^2 \), and normalized mean square error (NMSE).

The values reported for each regression performance metric were averaged across all cross-validation folds and repetitions.

IV. RESULTS

A. Feature Selection Scheme

Recursive feature elimination generated twelve features that have accumulative appearance exceeding three over five folds as depicted in Fig. 3 that were considered as valuable features. Three significant observations can be highlighted from this figure: (1) regression model captured features from all three axes to estimate the subject’s severity. This infers
that a complex activity such as truncal human sway use distinct information in all orthogonal axes for a comprehensive description; (2) only the recurrence rate in the Medial-Lateral (ML) direction was selected as opposed to the other two axes - Anterior-Posterior (AP) and Superior-Inferior (SI). The relative instability in the ML axis may be due to stability in the AP direction supported by the feet; (3) the harmonic ratio and entropy were not considered as valuable features while variability features were more important in the formation of the regression model. Fig. 2 shows the variation of eyes-open and eyes-closed of the four main selected features: recurrence rate, power of the first dominant frequency, coefficient of variation and signal to noise.

B. Patient Severity Assessment

The objective of severity assessment is to form a model that produce scores exhibiting well correlation with the SARA scores. Complete correlation is not necessarily the ideal because there may be intrinsic features that are not identified by clinical observation which are nevertheless recognized by a machine learning approach. As well, the range of SARA scores and their non-linearity may reduce the performance of regression models. Before feeding into regressors, we transformed targets to a smoother linear score by fitting the firt-degree polynomial. The correlation score increased by an average of 5% by this technique. Fitting a curve of second-degrees or more did not increase the correlation score. Table II reports the performance of candidate regression models where Elastic Net model produced a good performance with 0.70 correlation and 0.27 NMSE.

With RQA, interpolation of 1000 sample points yielded a significant correlation performance. Interpolating further to 2000 data points, however, did not improve performance of the regression model. Windowing segment of 500 data samples were long enough to obtain 0.70 Pearson correlation. Compensating between the model performance and user friendly aspect, 10 second data length of 500 data samples was decided to perform the test for later development.

C. Discussion

Romberg’s test is an important part of the assessment of balance in CA. The ability conduct assessments in non-clinical settings an allow more regular assessments for patients undergoing rehabilitation therapies. Fig. 3 presented accumulated features over five folds. Top fifteen selected feature were ones being nominated by more than three selectors. Obtaining good results with a small set of features indicate several irrelevant and redundant information were blended in the initial feature set. The proposed feature selection method optimizes the algorithm and reduce the computational overheads.

ACKNOWLEDGMENT

We acknowledge clinical contributions made by Sarah Milne and Mandy Marriott. This work is supported by the Florey Institute of Neuroscience and Mental Health, Melbourne, Australia via the National Health and Medical Research Council (NHMRC) GNT1101304 and APP1129595.

REFERENCES