2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Diagnosis Cerebellar Ataxia using Deep Learning with Time Series Transformed Image

Thang Ngo[1], *Student Member, IEEE*, Dinh C. Nguyen[2], *Student Member, IEEE*,
Pubudu N. Pathirana[3], *Senior Member, IEEE*, Malcolm Horne[4], Laura Power[5], and David J. Szmulewicz[6]

***Abstract*— Cerebellar ataxia (CA) is defined by disrupted coordination of movement suffering from disease of the cerebellum. It reflects fragmented movements of the eyes, vocal, upper limbs, balance, gait, and lower limbs. This study aims to use a motion sensor to form a simple yet effective CA quantitative assessment framework. We suggest a pendant device to use a single kinematic sensor attached to the wearer's chest to investigate the balance capability. Via a standard neurological test (Romberg's standing), the device may reveal an early symptom of Cerebellar Ataxia tailoring toward rehabilitation or therapeutic program. We adopt a transformed-image based approach to leverage the advantage of state-of-the-art deep learning models into diagnosis CA. Three transform techniques are employed including recurrence plot, melspectrogram, and Poincaré plot. Experiment results show that melspectrogram transform technique performs best in implementation with MobileNetV2 to diagnose CA with an average validation accuracy of 89.99%.**

I. INTRODUCTION

Cerebellar Ataxia (CA) is a neurological symptom defined by a specific problem with balance and coordination. The clinical setting involves ataxic people to perform many CA tasks as described in the Scale for Assessment and Rating of Ataxia (SARA) [1] to reveal any sign of disordered movement. This assessment scheme is affected heavily by the clinician's experience and the inherent subjectiveness of human involving process.

Defined as a rare symptom, research in CA has a long history in developing an objective assessment scheme. Advancement obtained recently in wearable sensor and cloud computing have encouraged many attempts to develop supportive systems for tele-assessment and tele-rehabilitation CA. These systems promise to provide a reliable and non-subjective scheme for measuring the severity of CA and tracking the rehabilitation progress. Cloud and wearable setting would provide a continual monitoring capability that may reveal new CA research insights.

Recently, there have been many new approaches and efforts to build new supportive assessment systems. Literature witnesses a wide range of designs from neuron imaging or external sensory systems. Brain scanning images used Magnetic Resonance Imaging to discriminate three CA phenotypes SCA2, SCA6, AT with an error rate of 13.75% [2]. External sensory systems utilize a broad range of sensors to capture ataxic behaviours. Eyes' iris movement [3] was captured by mobile phone camera to obtain a diagnostic sensitivity of 0.84 and specificity of 0.77. In limbs, balance, gait, vocal domains, researchers employed kinematic sensors, infrared camera, motion camera, googles, balance mat, EMC, wireless sensing to collect motion or motion-related data [4], [5], [6], [7]. In these studies, the application of machine learning limits within the frame of traditional approaches where collected data are extracted with features of interest and used to train conventional ML models. These approaches are known as hand-crafted methodology where the feature extraction is the most important factor that has been conducted manually.

The application of deep learning constrains to image-based data resources. To our best knowledge, the work in [2] is the only research conducted neural network on brain scanning images. In this study, we propose a framework by which IMU and other time-series can train the deep learning models through image transform techniques. These techniques have been applied widely and presented high feasibility in another research [8]. Not exclusive to time series data, we expect to apply the suggested scheme and leverage deep learning power in many other existing systems.

II. MATERIAL

A. Participants

TABLE I: General Information of Participants

	Ataxic Individuals	Healthy Controls
Number of participants	53	24
Age (years)	61.0 ± 11.7	50.4 ± 19.9
Gender (Male/Female)	31/22	6/18
Dominant limb (left/right)	8/45	6/18
Height (centimetre)	170.2 ± 9.12	167.18 ± 9.23

Fifty-three CA individuals and 24 healthy controls participated in the trial. Table I shows the demographic details. Ethical approval was submitted and given by the ethics committee of the Human Research and Ethics Committee, Royal Victorian Eye and Ear Hospital, East Melbourne,

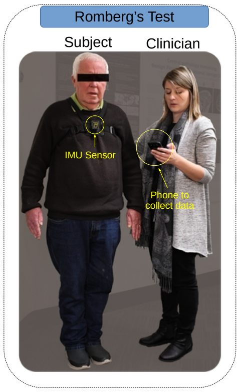
[1,2,3]T. Ngo, C. D. Nguyen and P. N. Pathirana are with the School of Engineering, Deakin University, Waurn Ponds, VIC 3216, Australia tdngo@deakin.edu.au; cdnguyen@deakin.edu.au; pubudu.pathirana@deakin.edu.au

[4]M. Horne is with the Florey Institute of Neuroscience and Mental Health, Parkville, VIC 3052, Australia malcolm.horne@florey.edu.au.

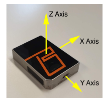
[5]L. Power is with Balance Disorders and Ataxia Service, Royal Victorian Eye and Ear Hospital (RVEEH), East Melbourne, VIC 3002, Australialaura_power@live.com.au

[6]D. J. Szmulewicz is with the Florey Institute of Neuroscience and Mental Health, Parkville, VIC 3052, Australia, the Balance Disorders and Ataxia Service, Royal Victorian Eye and Ear Hospital (RVEEH), East Melbourne, VIC 3002, Australia, and the Cerebellar Ataxia Clinic, Alfred Hospital, Prahran, VIC 3004, Australia dsz@me.com

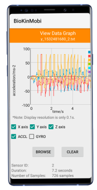
978-1-7281-1178-0/21/$31.00 ©2021 IEEE



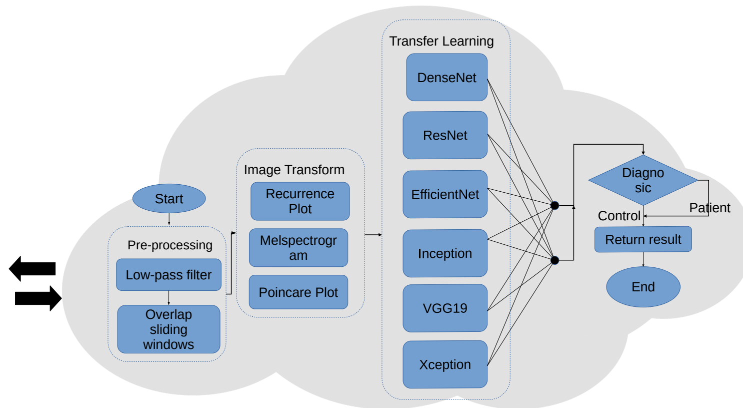
(a) Romberg's Stance Test



(b) IMU sensor



(c) Phone transfering data



(d) Cloud-based Processing

Fig. 1: IMU sensor as a supportive system to assess CA. (a) Romberg's standing test with an ataxic individual wore an IMU sensor. (b) Biokin kinematic motion capture sensor. (c) Mobile application as a gateway to record and upload raw data to a cloud computing. (d) Cloud-based services analyze motion data and return the diagnostic result to clinicians.

Australia (HREC Reference Number: 11/994H/16). Each subject has been provided a written consent which were collected before the trial.

B. Experimental Protocol

Subjects in Romberg balancing trial wore a single Biokin motion sensor on their chest [9]. The sensor was attached by an elastic strap wrap around their chest and over the shoulder. IMU Accelerometer sensor was sampled at 50 Hertz/second compensasing between low frequency human activity and amount of samples required for the algorithm. Subjects were instructed to stand upright on their feet placed closed together as illustrated in Fig. 1. The protocol ideally would last 30 seconds; however, the supporting clinician may stop the test at any time if severe subjects express risks of falling.

III. METHOD

A. Signal Pre-processing

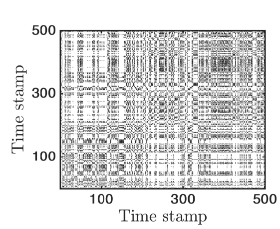
Raw data was first pretreated by interpolating, low-pass filtering, and sliding fixed size overlapping window. As the recurrence plot is suggested to be significant reliability when data length is more than 1000 samples [10]. We applied linear interpolation for mid-severe and high-severe CA individuals if their recorded data were shorter than 1000 samples. Sliding windows were employed to be able to select the most reliable frame. Matlab (R2020b, MatWorks, USA) and Tensorflow2 have been employed in this research.

B. Image Transform Techniques

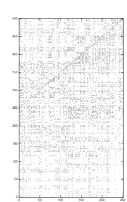
1) Recurrence Plot: A recurrence plot (RP) is an advanced method to analyse the nonlinear behaviour embedded in a data [10]. The recurrence plot visualise a square matrix, in which the matrix elements correspond to the closeness of a state concerning other states of a dynamical system. The first step in constructing the recurrence plot is to define the time delay $R^{\delta,\tau}(m)$ of the original signal.

$$R^{\delta,\tau}(m) = \begin{bmatrix} x^0(m) \\ x^1(m+\tau) \\ \vdots \\ x^\delta(m+\delta\tau) \end{bmatrix}, \; 1 \leq m \leq N - \delta\tau \quad (1)$$

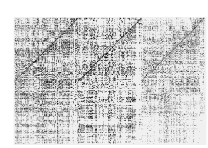
where $x^0(n)$ is the accelerometer with n samples in range $[1, \ldots, N]$, N is the data length and $x^\delta(m)$ is the time delayed signal where $x^0(n) = x^\delta(m+\delta\tau)$ for $n = m+\delta\tau$. δ and τ are pre-defined dimension and time delay parameters, respectively. The second step is to calculate distances between pairwise time delay signals using the Euclidean distance and RP is a heat map representing inter-distance among time delay signals. Fig. 2 shows the original recurrence plot of a typical CA individual. In this work, we modify the original recurrence plot by removing the upper half and rotating the lower triangle to eliminate redundant information. This modification reduces the size of input figures by half so that lowering complexity of DL structure.



(a) Recurrence Plot, Medial-Lateral axis, Accelerometer.



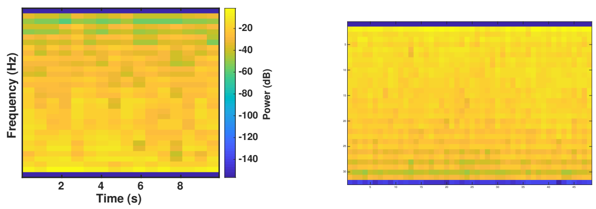
(b) Cut and Rotated Recurrence Plot.



(c) Merging three axes horizontally, Accelerometer.

Fig. 2: Recurrence plot used to feed into Deep learning models. (a) Original recurrence plot from accelerometer (medial-lateral axis). (b) Modified plot to reduce redundant information. (c) Merging of three axes to form the final plot to train the models.

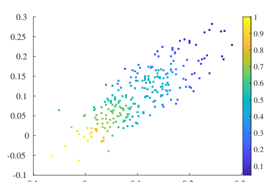
2) Melspectrogram: A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time [11]. A spectrogram is generated by a bank of 32 band-pass filters. Fig. 3 illustrates melspectrogram of a representative person.



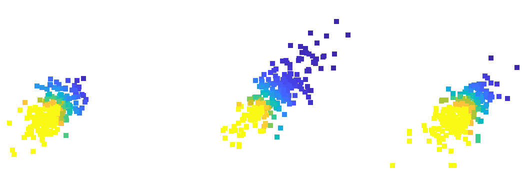
(a) Melspectrogram, Medial-Lateral axis, Accelerometer. (b) Input feed into Deep learning: Merging three axes horizontally, Accelerometer.

Fig. 3: Melspectrogram image transform from imu signal (accelerometer, medial-lateral axis). (a) Original melspectrogram with bank filters. (b) Stacking three axes melspectrograms into one figure to prepare for training the deep neural networks.

3) Poincaré plot: A Poincaré plot is a measure of RP employed to evaluate self-resemblance within processes [12]. It quantifies the correlation between two consecutive data points in IMU signal and visualizes geometrically in a 2D figure. Fig. 3 plots a presenting person's kinematic data. By stacking three axes accelerometers horizontally, we construct an input figure to feed into DL models.



(a) Poincaré plot, Medial-Lateral axis, Accelerometer.



(b) Merging three axes horizontally, Accelerometer.

Fig. 4: Poincaré Plot transformed image from accelerometer (medial-lateral axis). (a) Poincaré Plot of one channel of imu sensor. (b) Plot from three axes place horizontally for deep learning training and testing.

C. Deep Learning

1) Deep Learning Models:

MobileNetV2: Mobilenet is a lightweight convolutional architecture that contains 53 layers deep. The model utilises depthwise separable convolutions in which it fundamentally implements a single convolution on each of the three colour channel instead of merging and flattening them.

ResNet101V2: ResNet is a convolutional network that contains 101 layers deep. The kernel of ResNet introduces an "identity shortcut connection" in which it bypasses one or several layers. This technique helps overcome the notorious vanishing gradient problem of deep networks where the model's performance becomes saturated or even begins degrading quickly.

DenseNet201: DenseNet stands for Densely Connected Convolution Neural Networks (CNN). In its original manner, CNN has one layer connected only to its adjacent layers. DenseNet is a variant whereas its layer will be connected to all other layers in a feedforward manner.

EfficientNetB7: EfficientNet is a convolutional architecture that has been built to design a new baseline network. EfficientB7 is a network obtained by scaling up the baseline model. The network has obtained utmost accuracy on the ImageNet dataset while being smaller and faster than the best existing convolution neural network.

InceptionV3: Inception-v3 is a CNN comprising 48 connected layers. The model has been trained on the ImageNet database which contains more than 10^6 images. The training dataset contains 1000 object categories such as animals, pencil, keyboard, and household objects. With the diversified dataset, the Inception network has been capable of extracting features tailored to a broad array of images. The model is constructed from both blocks with pooling, concats, dropouts, and fully connected layers.

Xception: Xception is a CNN with 71 connected layers, an extreme edition of Inception model with a modified depthwise separable convolution. Xception architecture depends purely on depthwise separable convolution layers.

VGG19: VGG19 is an alternation of a very deep convolutional network which has been used in large-scale image recognition. It comprises 16 convolution layers, 3 fully connected layers, 5 maxpool layers and 1 output softmax layer. The model has been trained using an image database of 14,197,122 images.

2) Transfer Learning: Transfer learning is a technique to re-utilize knowledge obtained in previous assignment to resolve relevant ones [13]. Transfer learning has been used extensively in clinical research where human data was limited to collect. In this work, we remove the top layer of the pre-trained models and added two other layers of a global average pooling 2D and two nodes in the output presenting healthy controls and patients. The pre-trained models would help to extract features or texture out of the CA dataset.

IV. Diagnostic Results

Deep learning models are compared based on their validation accuracy over 50 epochs. We also considered the training time, which related tightly to the architecture of DL models. Table II reported performances of DL models over three image transform techniques. MobileNetV2 obtained its highest diagnostic result of 89.99% accuracy in accordance with melspectrogam. Fig. 5 illustrated the training and validation

TABLE II: Validation Accuracy of Deep Learning Model on Three Transform Techniques

	Transform technique		
Model	Recurrence	Melspectrogram	Poincaré
MobileNetV2	83.33	**89.99**	80.00
ResNet101V2	**86.67**	76.67	83.33
DenseNet201	80.00	68.33	76.67
EfficientNetB7	66.67	76.67	73.33
InceptionV3	83.33	69.99	80.00
Xception	76.67	60.00	80.00
VGG19	63.33	66.67	63.33

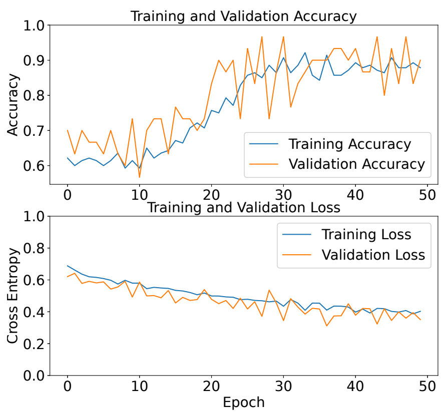


Fig. 5: Validation Accuracy of 89.99% of Melspectrogram on MobileNetV2.

accuracy as well as the loss of MobileNetV2. Compared to the traditional ML approach, the image-transform based scheme suggested in this work yielded a higher classification accuracy against 88.24% reported in [10]. The ResNet101V2 reached 86.67% accuracy, but the structure was heavier with a higher number of training parameters and so as training time. Application of DL eliminated the feature extraction step, which in turn reduced the engineering labour works. The suggested methodology was not limited to kinematic but could be applied on many different data type such as audio ataxic speech.

V. Discussion

Researchers in the literature classified ataxic versus healthy controls with more 80% accuracy and estimate severity with 0.7 correlation with clinical assessment [10]. Occupies its vital role in customizing rehabilitation therapy, phenotype identification remains a challenging question. With promising result from applying DL in CA via image-transform base, we expect to employ further the suggested scheme tackle problems of the severity estimation and phenotype identification. To these complicated problems, DL approach may outperform and provide solutions where conventional ML approaches seem to have already reached their limitations.

VI. Conclusion

This study suggested a scheme to use image transform based on deep learning framework to quantitatively asssess CA. Analysis results showed its possibility to give 89.99% classification result by using a lightweight convolutional architecture with melspectrogram. We noticed that more complex and deeper models such as DenseNet did not guarantee a higher performance. Future works include utilising tailored models for texture dataset and stacking two or more different models with different image input aiming to increase the final performance.

Acknowledgment

The authors acknowledge clinically supports by Sarah Milne and Mandy Marriott. This research is sponsored by the Florey Institute of Neuroscience and Mental Health, Melbourne, Australia via the National Health and Medical Research Council (NHMRC) GNT1101304 and APP1129595.

References

[1] T. Schmitz-Hübsch, S. T. Du Montcel, L. Baliko, J. Berciano, S. Boesch, C. Depondt, P. Giunti, C. Globas, J. Infante, J.-S. Kang *et al.*, "Scale for the assessment and rating of ataxia: development of a new clinical scale," *Neurology*, vol. 66, no. 11, pp. 1717–1720, 2006.

[2] Z. Yang, S. Zhong, A. Carass, S. H. Ying, and J. L. Prince, "Deep learning for cerebellar ataxia classification and functional score regression," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2014, pp. 68–76.

[3] Z. Chang, Z. Chen, C. D. Stephen, J. D. Schmahmann, H.-T. Wu, G. Sapiro, and A. S. Gupta, "Accurate detection of cerebellar smooth pursuit eye movement abnormalities via mobile phone video and machine learning," *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.

[4] 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," *Annals of biomedical engineering*, vol. 48, no. 4, pp. 1322–1336, 2020.

[5] C. Hohenfeld, I. Dogan, R. Schubert, C. Didszun, L. Schöls, M. Synofzik, I. A. Giordano, T. Klockgether, J. B. Schulz, R. Reilmann *et al.*, "Application of quantitative motor assessments in friedreich ataxia and evaluation of their relation to clinical measures," *The Cerebellum*, vol. 18, no. 5, pp. 896–909, 2019.

[6] J. Jin, W. Sun, F. Al-Turjman, M. B. Khan, and X. Yang, "Activity pattern mining for healthcare," *IEEE Access*, vol. 8, pp. 56 730–56 738, 2020.

[7] T. Honda, H. Mitoma, H. Yoshida, K. Bando, H. Terashi, T. Taguchi, Y. Miyata, S. Kumada, T. Hanakawa, H. Aizawa *et al.*, "Assessment and rating of motor cerebellar ataxias with the kinect v2 depth sensor: Extending our appraisal," *Frontiers in Neurology*, vol. 11, 2020.

[8] T. Hur, J. Bang, J. Lee, J.-I. Kim, S. Lee *et al.*, "Iss2image: A novel signal-encoding technique for cnn-based human activity recognition," *Sensors*, vol. 18, no. 11, p. 3910, 2018.

[9] D. C. Nguyen, K. D. Nguyen, and P. N. Pathirana, "A mobile cloud based iomt framework for automated health assessment and management," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 6517–6520.

[10] T. Ngo, P. Pathirana, M. Horne, L. Power, D. Szmulewicz, S. Milne, L. Corben, M. Roberts, and M. Delatycki, "Balance deficits due to cerebellar ataxia: A machine learning and cloud-based approach," *IEEE Transactions on Biomedical Engineering*, 2020.

[11] L. Rabiner and R. Schafer, *Theory and applications of digital speech processing*. Prentice Hall Press, 2010.

[12] T. S. Henriques, S. Mariani, A. Burykin, F. Rodrigues, T. F. Silva, and A. L. Goldberger, "Multiscale poincaré plots for visualizing the structure of heartbeat time series," *BMC medical informatics and decision making*, vol. 16, no. 1, pp. 1–7, 2015.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.