Diagnosis of Cerebellar Ataxia Based on Gait Analysis Using Human Pose Estimation: A Deep Learning Approach

Hisham Khalil[1], Ahmed Mohamed Saad Emam Saad[2], Uswah Khairuddin[1]

Abstract— **Human gait analysis has been one of the primary procedures for diagnosis in modern healthcare applications for various diseases. Instead of using expensive wearable sensors on patients, this research aims to assist in gait analysis and classification for medical diagnoses using computer vision solely. A long short-term memory (LSTM) neural network based on MediaPipe Pose for video-based human gait analysis is proposed to assist in diagnosing patients with neurodegenerative diseases, particularly cerebellar ataxia. The kinematic parameters were extracted from the pose estimation model on captured gait videos before deriving the spatiotemporal parameters for quantitative gait analysis. Data augmentation is applied to increase dataset size, and five-fold cross-validation is performed to verify the suitability of the developed dataset for training deep neural networks. The selected LSTM model achieves a testing accuracy of 99.8% with very high precision and recall metrics for ataxic and normal gait classes. The proposed methodology can be applied in broader applications for remote rehabilitation and patient monitoring.**

Clinical Relevance— **The developed system can assist physicians in diagnosing cerebellar ataxic patients and monitoring gait rehabilitation process remotely via camera vision.**

I. Introduction

The current research in the rehabilitation and healthcare sectors is heading towards utilizing state-of-the-art technologies to further their scope and applications, where there is a reduction in human involvement. Diseases are typically detected and rated by conducting a clinical assessment session by qualified physicians on the respective patients. However, recent advancements in machine learning led to its massive implementation in the healthcare sector [1], causing the introduction of telehealth [2] and automated diagnostic systems [3]. Commonly used human data for machine learning-based healthcare assessment and rehabilitation monitoring are gait data. Human gait analysis assists diagnosing systems to identify human movement patterns that appear abnormal and associate them with a disease.

Quantitative methods in human gait analysis are commonly implemented to assist in diagnosing neurodegenerative diseases that cause gait and posture instabilities [4]. However, traditional motion capture methods based on wearable sensors or camera setups enacted previously for gait assessment are possessing limitations such as the confinement to laboratory environments [5], high cost, and the difficulty of application on patients [6]. Furthermore, research in human motion science for healthcare requires extensive procedures to consider the publication of sensitive data from patients. Currently, new trends in motion analysis tools rely on video-based solutions that are easy and convenient to implement in remote scenarios, and the state-of-the-art technology for such purpose is *human pose estimation*.

This research aims to implement human pose estimation for gait analysis to develop a deep learning model to facilitate the diagnosis of neurodegenerative diseases, particularly cerebellar ataxia. The pose estimation framework implemented is MediaPipe Pose [7] due to its high efficiency with low-end devices. Human joint motion data are extracted from MediaPipe Pose applied on videos of ataxic and normal gait, and gait parameters are computed from the joint data to provide additional descriptive dimensionality to the neural network for gait classification. A long short-term memory (LSTM) network is used as its architecture highly performs classification on time series data [8], i.e., multivariate gait sequences. The trained LSTM model provided a reliable diagnosis of cerebellar ataxia from raw videos of humans walking as input.

The rest of the paper is organized as follows. Section II mentions some related literature work and summarizes the contributions of this research. Section III describes the methods of dataset development and neural network training for developing the cerebellar ataxia diagnosing framework. Section IV shows and discusses the results of testing the LSTM model from the cross-validation and optimization stages. Section V states the conclusion and future work for this research.

II. Related Work

There have been many works of gait analysis and classification in literature applying several types of motion capture systems. Lempereur *et al.* [9] implemented Vicon MX motion capture system and force platforms on children with gait disorders from neurological and orthopaedic diseases using their proposed DeepEvent network. Other diseases such as osteoarthritis have been explored in [10], where the authors used pre-trained convolutional neural network (CNN) models on gait datasets consisting of deep features aggregated from gait images based on transfer learning. Woo *et al.* [11] implemented Automated Machine Learning

This work was supported by The Japanese Chamber of Trade and Industry, Malaysia (JACTIM).

[1]H. Khalil and U. Khairuddin are with Center for Artificial Intelligence and Robotics, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia hishamkhalil.eng@gmail.com, uswah.kl@utm.my

[2]A. M. S. E. Saad is with the Department of Computing Sciences, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412 USA asaad1@islander.tamucc.edu

(AutoML) and dynamic time warping (DTW) for analysis of gait and physical fitness characteristics in senior citizens with and without diabetes. They also developed a home-based monitoring system based on MediaPipe Pose for rehabilitation assistance. Wang *et al.* [12] developed a skeleton detector and gait assessment framework called SAIL, in which they utilized annotated Timed "Up & Go" (TUG) test videos for gait inputs and support vector machine (SVM) as their classifier.

LSTM is used in some research works for human gait classification. Zhao *et al.* [13] used LSTM to assist clinical physicians in identifying gait disturbances instead of subjectively evaluating them. Sadeghzadehyazdi *et al.* [14] demonstrated an LSTM-CNN network model that classifies nine different pathological gait types on a publicly available dataset consisting of spatiotemporal and body joint features extracted from Kinect.

Neurodegenerative diseases have been the most commonly investigated diseases in research for abnormal gait detection and assessment applications. Sato *et al.* [15] implemented pose estimation for Parkinsonian gait quantification, providing evidence that their methods can be applied in home-based environments where virtual therapy sessions can be more convenient. Similar to the scope of this research, in [16]–[18], the authors proposed methods for classifying different types of ataxia. Cerebellar ataxic gait identification was also explored in [19], utilizing SVM, CNN, and Naïve Bayes' classifiers to diagnose ataxic and normal gait using their developed dataset from frequency components of sensor-based accelerometric signals. Other neurodegenerative diseases such as dementia [20], levodopa-induced dyskinesia [21], idiopathic normal pressure hydrocephalus [22], and Huntington disease [23] were also investigated.

Based on the literature review, the targeted disease in this research, cerebellar ataxia, is not sufficiently implemented in correspondence with works on other neurodegenerative diseases based on machine learning-assisted diagnostic tools. Moreover, human pose estimation, was not applied in the general theme of cerebellar ataxic gait classification. Therefore, this research's contributions to address the mentioned research gaps are as follows:

1) Developing a cerebellar ataxic gait motion dataset that implies additional contribution to neurodegenerative disease diagnosis based on human gait analysis using pose estimation models.
2) Building a neurodegenerative gait classification model that generalizes assessment for diseases that cause gait instability, which is beneficial for telemedical and rehabilitation applications.
3) Experimenting a recently developed pose estimation model, i.e., MediaPipe Pose, that validates its applications in the healthcare assessment domain.

III. Methodology

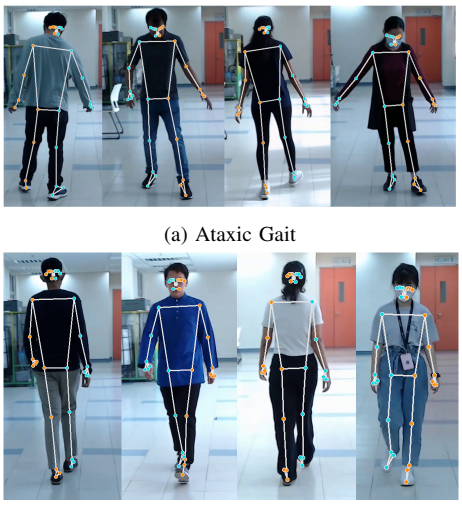
The proposed methodology for dataset and classifier development is as follows. Firstly, three-dimensional (3D) landmark, or joint, coordinates are extracted after using gait videos as input to MediaPipe Pose. Then, data augmentation is applied to the extracted joint coordinates to increase the dataset size for ataxic and normal gait classes. Spatiotemporal and kinematic parameters are computed from the extracted coordinate features. LSTM is used for gait classification, differentiating between ataxic and normal gait. Lastly, model cross-validation, evaluation, and selection are conducted. Fig. 1 illustrates MediaPipe Pose detection on captured videos, showing examples of back and forth walking sequences for each type of gait, while Fig. 2 shows the overall framework of gait classification based on pose estimation.

A. Dataset Development

1) Gait Video Collection: To the best of the authors' knowledge, there are no publicly available video datasets for cerebellar ataxic gait at the time of conducting this research. Therefore, a total of 20 volunteering university students, 15 male and 5 female students with mean age of 24 $\pm$ 3 years old, participated in this study. Each participant recorded two video sessions: one walking normally in front of the camera and one imitating the ataxic gait pattern, each lasting about 30 seconds to 1.5 minutes. 40 gait videos, 20 for ataxic gait and 20 for normal gait, were recorded and used as an input to MediaPipe Pose (see Fig. 1). Their participation and raw motion data publication consent was taken, and they were rewarded for participation motivation.

2) Feature Extraction: MediaPipe Pose contains a repository of 33 full-body joint landmarks, and each landmark is represented in the 3D Cartesian space as (x,y,z) coordinates. The 3D coordinates represent the approximate world coordinate frame in meters (m) with respect to the origin at the center between the hips.

Only the upper body joint coordinates are required for spatiotemporal and kinematic feature calculation. Table I states the extracted joint landmarks from MediaPipe Pose for



(a) Ataxic Gait

(b) Normal Gait

Fig. 1. MediaPipe Pose detection on a sample of captured videos for ataxic and normal gait.

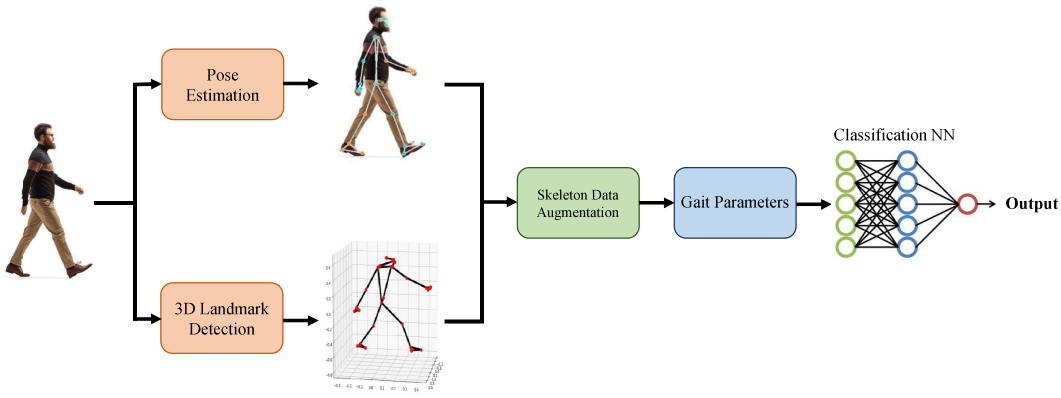


Fig. 2. Process flow of the proposed gait diagnosis framework.

TABLE I

SIGNIFICANT JOINTS AND BODY SEGMENTS IDENTIFIED FOR GAIT ANALYSIS. SEGMENTS ARE OBTAINED BY COMPUTING THE NORM OF THE VECTOR FORMED BY EACH PAIR OF JOINTS MAPPED AS (i, j).

Joint Index	Joint Name	$(i, j) \in$ Segment	Segment Name
1	*left shoulder*		
2	*right shoulder*	(1, 3)	left upper arm
3	*left elbow*	(3, 5)	left lower arm
4	*right elbow*	(1, 7)	left torso
5	*left wrist*	(7, 9)	left thigh
6	*right wrist*	(9, 11)	left shank
7	*left hip*	(2, 4)	right upper arm
8	*right hip*	(4, 6)	right lower arm
9	*left knee*	(2, 8)	right torso
10	*right knee*	(8, 10)	right thigh
11	*left ankle*	(10, 12)	right shank
12	*right ankle*		

Algorithm 1 Bilateral symmetry data augmentation method

Input: 1 × n-dimensional vector of **V**
Output: m × n-dimensional matrix of **A**

```
count = 1;
A = zeros(m, n);
for q = 1:1:p do
    comb = nchoosek(1:1:p, q);
    for b = 1:1:size(comb, 1) do
        A(count, :) = V;
        temp1 = V(comb(b, :));
        temp2 = V(comb(b, :) + i);
        A(count, comb(b, :)) = temp2;
        A(count, comb(b, :) + i) = temp1;
        count = count + 1;
    end for
end for
```

gait analysis and maps the body segments associated with the extracted landmarks, which are used in the data augmentation algorithm.

3) Data Augmentation: Data augmentation is a technique implemented in this research for increasing the dataset size to improve the model training performance since there are no sufficient data to achieve high classification accuracy. Algorithm 1 is based on the pseudocode proposed in [24], in which the authors applied human bilateral symmetry data augmentation on skeleton data obtained from Kinect.

The $1 \times n$ vector $\mathbf{V}$ is the feature vector, and $m \times n$ is the augmented matrix $\mathbf{A}$ of feature data. p represents the number of symmetric pairs of segments, and i represents the difference in index value between each pair of symmetric body segments in the feature vector.

$n = 10$ is the number of segments, and $m = 2^n - 1 = 2^5 - 1 = 31$ is the number of combinations of generated synthetic augmentations to $\mathbf{V}$. The number of symmetric pairs is $p = 5$, and the difference of symmetric segment pair index value is $i = 5$. A generated matrix $\mathbf{M}$ is a concatenation of $\mathbf{V}$ (1×10) and $\mathbf{S}$ (31×10) with a total dimension of 32×10.

The data augmentation method is applied to each of the 40 videos, where each $\mathbf{M}$ represents an augmentation of a single time series data point. The produced dataset for each video is concatenation of n $\mathbf{M}$ matrices, where n is the number of initially extracted frames from each video. The re-organized dataset is represented in the following equation:

$$\mathbf{D} = \{\{\mathbf{M}_{1,1}, \cdots, \mathbf{M}_{1,n}\}, \cdots, \{\mathbf{M}_{32,1}, \cdots, \mathbf{M}_{32,n}\}\}, \quad (1)$$

where the length of $\mathbf{D}$ is 32. The next step is to scale the new augmented vector of segment lengths with respect to the segment lengths from the initial pre-augmented dataset. The scaling is performed using the following equation:

$$\mathbf{S}_{v,scaled} = \frac{\mathbf{S}_v \times L}{L_0}, \quad (2)$$

where L_0 is the original segment length and $L = \|\mathbf{S}_v\|$ is the current augmented segment length. A uniformly distributed value of random noise between 0 and $\frac{\sigma_L}{4}$ is added to avoid redundancy in each $\mathbf{S}_{v,scaled}$, where σ_L is the sample standard deviation of all the lengths of that particular segment in the whole augmented $\mathbf{D}$ dataset. To obtain the joint coordinate

from $\mathbf{S}_{v,scaled}$, the following equation is used:

$$\begin{bmatrix} x & y & z \end{bmatrix}^{\mathsf{T}}_{new} = \mathbf{S'}_{v,scaled} + \begin{bmatrix} x & y & z \end{bmatrix}^{\mathsf{T}}_{original}, \quad (3)$$

where $\mathbf{S'}_{v,scaled}$ denotes a scaled and noised vector. The resulting dataset is concatenated with the original joint coordinate dataset. The final augmented matrix containing all the coordinates has a shape of 32×36, where 32 results from the 31 variations plus the original variation and 36 is the number of obtained coordinate components (3 axes $\times$ 12 joints).

4) Gait Feature Engineering: Kinematic and spatiotemporal parameters are calculated from the extracted joint coordinates from MediaPipe Pose. The kinematic parameters are the angles of six joints in the upper body: left and right shoulders, hips, and knees. The spatiotemporal parameters associated with ataxic gait are derived from [25], [26]. Table II states the computational formulas for the selected kinematic and spatiotemporal features. The final dataset[1] is composed of 36 joint coordinates combined with 6 kinematic and 5 spatiotemporal parameters, making a total of 47 features.

B. Model Training, Validation, and Evaluation

Table III states various tested hyperparameters throughout different trials of training the LSTM model, and Fig. 3 illustrates a simplified graphical representation of the neural network's architecture. The LSTM architecture is developed using PyTorch framework, and the model is trained on NVIDIA GeForce RTX 3070 Graphical Processing Unit (GPU) for computational acceleration.

LSTM trains not on whole time series variables but instead on divided sequences of data. The sequences are labeled and fed to the LSTM network for training. The shape of each sequence is 256 defined time series data points $\times$ 47 features.

[1]The developed dataset for this research can be found at `http://dx.doi.org/10.17632/2vkk2r9tx3.1`.

TABLE II

SPATIOTEMPORAL AND KINEMATIC PARAMETERS FOR BOTH DATASETS: ATAXIC AND NORMAL GAIT. IC IS INITIAL CONTACT AND FC IS FINAL CONTACT.

Parameter	Formula
Base Width [m]	$\lvert left\ ankle_x - right\ ankle_x \rvert$
Feet Clearance [m]	$\lvert left\ ankle_y - right\ ankle_y \rvert$
Stride Length [m]	$\lvert left\ ankle_z - right\ ankle_z \rvert$
Stride Time [s]	$time^{FC} - time^{IC}$
Stride Speed [m/s]	$\frac{Stride\ Length}{Stride\ Time}$
Shoulder Abduction Angle [rad]	$\arccos \frac{\overrightarrow{upper\ arm} \cdot \overrightarrow{torso}}{\lVert \overrightarrow{upper\ arm} \rVert \cdot \lVert \overrightarrow{torso} \rVert}$
Hip Flexion Angle [rad]	$\arccos \frac{\overrightarrow{torso} \cdot \overrightarrow{thigh}}{\lVert \overrightarrow{torso} \rVert \cdot \lVert \overrightarrow{thigh} \rVert}$
Knee Flexion Angle [rad]	$\arccos \frac{\overrightarrow{thigh} \cdot \overrightarrow{shank}}{\lVert \overrightarrow{thigh} \rVert \cdot \lVert \overrightarrow{shank} \rVert}$

TABLE III

TESTED HYPERPARAMETERS THROUGHOUT MODEL VALIDATION AND OPTIMIZATION STAGES. SELECTED HYPERPARAMETERS LEADING TO SATISFACTORY RESULTS ARE IN **BOLD**.

Hyperparameter	Tested Values
Epochs	**120**, 150, 160, 200, 250
Batch Size	**64**, 128
Layers	3, **4**
Hidden Units	128, **256**
Optimizer	**Adam**
Learning Rate	0.001, **0.0001**
Dropout Rate	0, 0.5, **0.75**
Weight Decay	0, **0.00001**

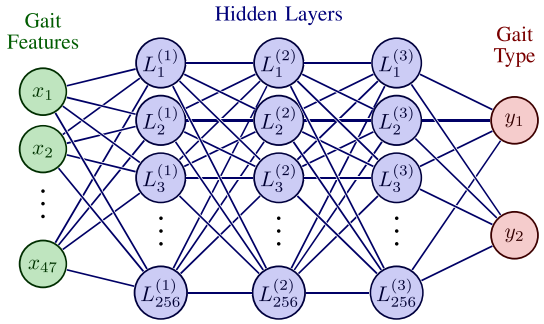


Fig. 3. Selected LSTM neural network visualization. L represents LSTM cells, parentheses superscripts represent hidden layer index, and subscripts represent the index of each unit within the layer.

k-fold cross-validation is performed to determine the best configuration of training and testing splits. The validation method implemented is k-fold cross-validation, which is described by the following representation [11]:

$$\mathcal{A}^* \in \arg\min_{A \in \mathcal{A}} \frac{1}{k} \sum_{i=1}^{k} \mathcal{L}\left(A, \mathcal{D}_{\text{train}}^{(i)}, \mathcal{D}_{\text{val}}^{(i)}\right). \quad (4)$$

The whole testing set is used as the validation set. The k-fold cross-validation model determines the model $\mathcal{A}^* \in \mathcal{A}$ with the optimized composition of training and validation data. $\mathcal{D}$ is the dataset, one each for training and validation, and $\mathcal{L}$ is the cross entropy loss function. The number of folds considered is $k = 5$. The allocated samples are 32 videos for training and 8 videos for validation evenly split among ataxic and normal gait classes.

Evaluation methods of the LSTM classifier are the confusion matrix and three metrics that are derived from it, which are recall, precision, and accuracy [27].

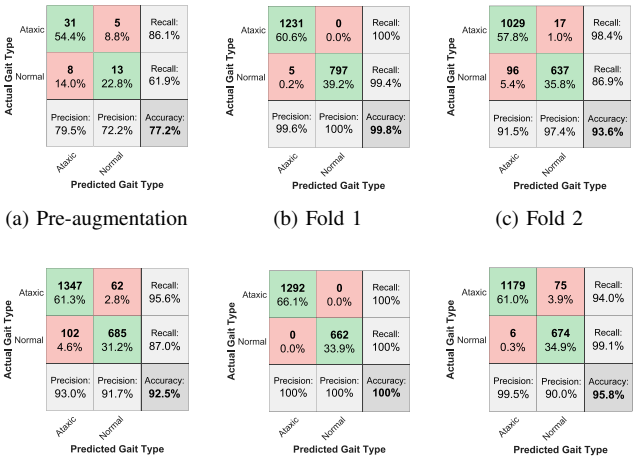
IV. RESULTS AND DISCUSSION

To validate the effect of data augmentation on model performance, the pre-augmented dataset was used for training using the same hyperparameters and the training/validation split as the first fold of cross-validation. In fold 1, each person is generating data for the two classes, and both ataxic and normal gait datasets are included in the training and validation sets, testing the hypothesis that the classification

performance shall be better than other randomly generated training folds. Folds 2 to 5 were developed by random generation of dataset splits.

Fig. 4 shows the confusion matrices for the pre-augmentation and the five-fold cross-validation stages. To interpret comparable results between all five folds, initial hyperparameters were set and maintained constant. The confusion matrix for the pre-augmentation stage in Fig. 4a indicates that the model under fitted the smaller dataset. Moreover, the model tended to classify validation examples as ataxic gait more than as normal gait, which is evident from the normal gait class's lower precision and recall results. In comparison, the classification results of the five folds showcase higher accuracy results, supporting the hypothesis that data augmentation improves performance drastically.

Folds 1 and 4 resulted in the best classification results among all the folds. This validates the hypothesis of the positive effect of including the same person's ataxic and normal gait datasets in the training set, which was the case for fold 1. However, fold 4, as shown in Fig. 4e, still achieved a higher and maximum accuracy of 100%. Moreover, folds 2, 3, and 5 achieved highly accurate results, indicating the suitability of the dataset for training LSTM networks. Even though fold 4 led to maximum classification results, hyperparameter tuning was implemented on the best-performing fold since the initial hyperparameters did not include regularization, which could hypothetically lead to poor performance on unseen data due to high model complexity. Figs. 5 and 6 show the validation accuracy and loss graphs with respect to training epochs and the confusion matrix of the selected LSTM model with the optimized hyperparameters, respectively. Fold 4 was selected as the model to be optimized because the model achieved the best accuracy in the cross-validation stage. To evaluate different hyperparameters on the classification results, four iterations of hyperparameter tunings were conducted to obtain highly performing results with the inclusion of regularization terms



(a) Pre-augmentation (b) Fold 1 (c) Fold 2

(d) Fold 3 (e) Fold 4 (f) Fold 5

Fig. 4. Confusion matrices of the pre-augmentation and five-fold cross-validation stages.

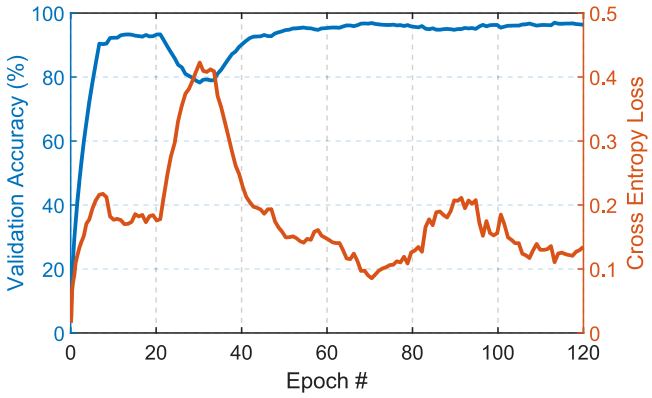


Fig. 5. Validation accuracy and loss curves of the selected LSTM model.

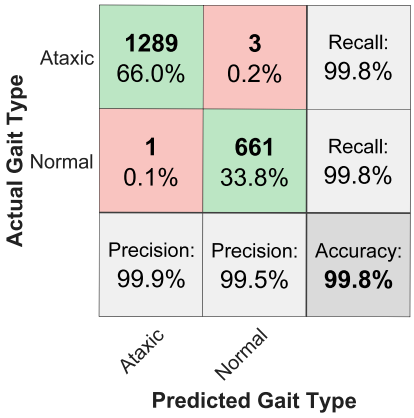


Fig. 6. Confusion matrix of the selected LSTM model.

in the LSTM network architecture.

The hyperparameters stated in Table III were experimented during the four optimization iterations, and the selected hyperparameters for the optimal model are highlighted in bold. The validation accuracy curve in Fig. 5 shows stability and conversion to a very high accuracy rate. The smoothed loss curve in Fig. 5, however, shows slightly unstable performance, especially during the first 40 epochs. This behavior is expected in neural network performance on large datasets. The actual minimum cross entropy loss was reached at the 108^{th} epoch from 120 total number of epochs with a value of 0.0077.

The confusion matrix in Fig. 6 shows very high accuracy of 99.8% on the testing set. Results from precision and falsely classified sequences indicate that the model's bias is slightly leaning towards ataxic gait. This result could be due to the larger portion of ataxic gait sequences in the training set. However, the selected model misclassified three ataxic gait sequences as normal gait types, which is more than the single misclassification of the normal gait sequence. Generally, bias towards one class can be minimized by balancing the dataset size among all classes.

V. Conclusion

In this research, a human pose estimation-based classification framework using LSTM neural network that can

diagnose cerebellar ataxic and normal gait is proposed. The dataset for training the LSTM network is developed by acquiring videos of people imitating both ataxic and normal gait patterns and applying data augmentation and feature engineering techniques. The proposed model achieved gait type prediction accuracy of 99.8%. Emerging machine learning solutions can assist patients possessing neurodegenerative diseases to experience high-performing and reliable healthcare systems that operate efficiently and remotely, especially when physical diagnosis is infeasible. Moreover, the development of a human motion dataset for cerebellar ataxic gait and a very highly accurate disease diagnosis classifier based on such dataset is the major contribution of this research. Based on the authors' knowledge, there is no ataxic gait dataset based on pose estimation in the literature, which its creation advocates this research's novelty.

Challenges encountered in this research are the uncertainty of the model's performance on real-world data and occlusion during pose detection. Future implementations to resolve these limitations include the acquisition of videos consisting of varying specifications such as camera angle, resolution, and gait pace to produce a more generalized model and minimize occlusion. Further testing of a variety of neural network types and architectures, e.g., CNN, could yield more robust models. Furthermore, techniques such as domain adaptation could provide significantly enhanced performances when testing the model on unseen data from varying distributions. Future applications of this research include the deployment of the trained LSTM model on a mobile application that can be utilized by clinical physicians with their assigned patients virtually.

ACKNOWLEDGMENT

The authors would like to thank JACTIM for sponsoring the project titled "Detection and Classification of Cerebellar Ataxic Gait from Videos Based on Human Pose Estimation" under JACTIM Research Proposal Competition 2021.

REFERENCES

[1] P. Esmaeilzadeh, "Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–19, 2020.

[2] F. De Marchi, E. Contaldi, L. Magistrelli, R. Cantello, C. Comi, and L. Mazzini, "Telehealth in Neurodegenerative Diseases: Opportunities and Challenges for Patients and Physicians," *Brain Sciences*, vol. 11, no. 2, p. 237, 2021.

[3] A. Stanley and J. Kucera, "Smart Healthcare Devices and Applications, Machine Learning-based Automated Diagnostic Systems, and Real-Time Medical Data Analytics in COVID-19 Screening, Testing, and Treatment," *American Journal of Medical Research*, vol. 8, no. 2, pp. 105–117, 2021.

[4] G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. R. D'Orazio, "Human Gait Analysis in Neurodegenerative Diseases: A Review," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 229–242, 2021.

[5] H. Khalil, E. Coronado, and G. Venture, "Human Motion Retargeting to Pepper Humanoid Robot from Uncalibrated Videos Using Human Pose Estimation," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, 2021, pp. 1145–1152.

[6] P. Chinmilli, S. Redkar, W. Zhang, and T. Sugar, "A review on wearable inertial tracking based human gait analysis and control strategies of lower-limb exoskeletons," *Int. Robot. Autom. J*, vol. 3, no. 7, p. 00080, 2017.

[7] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device Real-time Body Pose tracking," *arXiv preprint arXiv:2006.10204*, 2020.

[8] T. D. Pham, "Time–frequency time–space LSTM for robust classification of physiological signals," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[9] M. Lempereur *et al.*, "A new deep learning-based method for the detection of gait events in children with gait disorders: Proof-of-concept and concurrent validity," *Journal of Biomechanics*, vol. 98, p. 109490, 2020.

[10] M. A. Khan *et al.*, "Human gait analysis for osteoarthritis prediction: a framework of deep learning and kernel extreme learning machine," *Complex & Intelligent Systems*, pp. 1–19, 2021.

[11] Y. Woo *et al.*, "Classification of Diabetic Walking for Senior Citizens and Personal Home Training System Using Single RGB Camera through Machine Learning," *Applied Sciences*, vol. 11, p. 9029, 2021.

[12] Y. Wang *et al.*, "SAIL: A Deep-Learning-Based System for Automatic Gait Assessment From TUG Videos," *IEEE Transactions on Human-Machine Systems*, vol. 52, pp. 110–122, 2022.

[13] A. Zhao, L. Qi, J. Dong, and H. Yu, "Dual channel LSTM based multi-feature extraction in gait for diagnosis of Neurodegenerative diseases," *Knowledge-Based Systems*, vol. 145, pp. 91–97, 2018.

[14] N. Sadeghzadehyazdi, T. Batabyal, and S. T. Acton, "Modeling spatiotemporal patterns of gait anomaly with a CNN-LSTM deep neural network," *Expert Systems with Applications*, vol. 185, p. 115582, 2021.

[15] K. Sato, Y. Nagashima, T. Mano, A. Iwata, and T. Toda, "Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning–based 2D pose estimator," *PloS one*, vol. 14, no. 11, p. e0223549, 2019.

[16] R. Jaroensri *et al.*, "A Video-Based Method for Automatically Rating Ataxia," in *Proceedings of the 2nd Machine Learning for Healthcare Conference*, 2017, pp. 204–216.

[17] O. Vyšata *et al.*, "Classification of Ataxic Gait," *Sensors*, vol. 21, p. 5576, 2021.

[18] Q. Zhang, X. Zhou, Y. Li, X. Yang, and Q. H. Abbasi, "Clinical Recognition of Sensory Ataxia and Cerebellar Ataxia," *Frontiers in Human Neuroscience*, vol. 15, 2021.

[19] A. Procházka *et al.*, "Deep Learning for Accelerometric Data Assessment and Ataxic Gait Monitoring," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 360–367, 2021.

[20] K.-D. Ng, S. Mehdizadeh, A. Iaboni, A. Mansfield, A. Flint, and B. Taati, "Measuring Gait Variables Using Computer Vision to Assess Mobility and Fall Risk in Older Adults With Dementia," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–9, 2020.

[21] M. H. Li, T. A. Mestre, S. H. Fox, and B. Taati, "Vision-based assessment of parkinsonism and levodopa-induced dyskinesia with pose estimation," *Journal of NeuroEngineering and Rehabilitation*, vol. 15, p. 97, 2018.

[22] S. Jeong, H. Yu, J. Park, and K. Kang, "Quantitative gait analysis of idiopathic normal pressure hydrocephalus using deep learning algorithms on monocular videos," *Scientific Reports*, vol. 11, p. 12368, 2021.

[23] S. Zhang, S. K. Poon, K. Vuong, A. Sneddon, and C. T. Loy, "A Deep Learning-Based Approach for Gait Analysis in Huntington Disease," in *Studies in Health Technology and Informatics*, vol. 264, 2019, pp. 477–481.

[24] B. Kwon and S. Lee, "Human Skeleton Data Augmentation for Person Identification over Deep Neural Network," *Applied Sciences*, vol. 10, no. 14, p. 4849, 2020.

[25] S. Summa *et al.*, "Spatio-temporal parameters of ataxia gait dataset obtained with the Kinect," *Data in Brief*, vol. 32, p. 106307, 2020.

[26] E. Buckley, C. Mazzà, and A. McNeill, "A systematic review of the gait characteristics associated with Cerebellar Ataxia," *Gait & Posture*, vol. 60, pp. 154–163, 2018.

[27] M. Vihinen, "How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis," in *BMC Genomics*, vol. 13, no. 4, 2012, pp. 1–10.