IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Journal Européen des Systèmes Automatisés
Vol. 58, No. 12, December, 2025, pp. 2461-2475
Journal homepage: http://iieta.org/journals/jesa

Gait Pattern Analysis for Early Detection of Neuromuscular Disorders Using Wearable Sensors and Artificial Intelligence Techniques

Hydar Saadi Hassan Al-Wasti

Anatomy Department, College of Medicine, University of Baghdad, Baghdad 10047, Iraq

Corresponding Author Email: hydaralwasti@comed.uobaghdad.edu.iq

https://doi.org/10.18280/jesa.581202

Received: 8 September 2025
Revised: 27 November 2025
Accepted: 4 December 2025
Available online: 31 December 2025

Keywords:
gait analysis, cerebellar ataxia, CNN, wearable sensors, neuromuscular disorders, signal classification, biomedical signal processing, early diagnosis

ABSTRACT

In this work, a gait-based approach for early detection of cerebellar ataxia is introduced using deep learning and gait data collected from an array of wearable knock sensors mounted on the legs. The dataset was collected from the Kaggle Gait Analysis Dataset for cerebellar ataxia and pre-processed by z-score normalization and segmented into 128 samples with 50% overlap. A customized Convolutional Neural Network (CNN) model was designed and trained on these segments to classify gait patterns as normal and ataxic. The training mechanism indicated that the accuracy quickly grew to 95%, close to 100% at the end of the training. However, the trained CNN provided only a moderate accuracy of 40%, with a precision of 0.57 and a recall of 0.31 for normal gait and a precision of 0.31 and a recall of 0.57 for ataxic gait, thus resulting in F1-scores of 0.4 for both classes on unseen test data. The confusion matrix reflected an imbalance of ergonomics towards overprediction of ataxia by a nine-over-thirteen number of normal 'samples' misclassified. Whereas the model provides high confidence prediction scores (81%–99%) even for misclassifications, this indicates a model prone to overfitting and lacking generalizability. These findings underscore the promise and challenges of AI-augmented gait diagnostics, with model calibration, dataset balancing, and feature refinement indicated to improve sensitivity and specificity for clinical utility.

1. INTRODUCTION

Gait pattern analysis has become increasingly significant in the diagnosis and monitoring of early neuromuscular disorders, which represent a range of diseases such as Parkinson's disease, stroke, and muscular dystrophies. These diseases typically present with subtle gait features many years before they develop into symptomatic disease. Conventional clinical tests, though useful, have generally been performed within a controlled environment and may not reflect the full range of a patient's ambulatory behavior 'in the wild'.

Recent developments in wearable sensor technologies, including inertial measurement units (IMUs), gyroscopes, and surface electromyography (sEMG), have made it possible to record spatiotemporal and kinematic gait data during normal daily activity. In conjunction with AI approaches, these systems are able to automatically identify, categorize, and even forecast gait anomalies at a high level of accuracy. Machine learning and deep learning models, in particular, convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and attention-based architectures, have demonstrated great potential in automatically deriving clinically relevant features and biomarkers from raw sensor data.

The combination of wearable sensors with AI approaches is used not only in clinical diagnostics but also preventive healthcare, rehabilitation monitoring, and fall risk prediction. For instance, explainable attention-based deep learning models have facilitated real-time and interpretable gait analysis for both Parkinson's and post-stroke patients. Furthermore, their results are particularly promising for cloud-based approaches, which use smartphone sensors and agent-based simulation platforms to enable widespread dissemination and scalability of gait diagnostics.

Liu et al. [1] emphasized the increasing focus on wearable devices for smart healthcare. They highlighted that wearable sensors, including accelerometers and gyroscopes, are able to monitor gait patterns on a long-term basis and thus can produce spatiotemporal and kinematic information beneficial for the diagnosis of locomotor abnormalities caused by neurological or musculoskeletal disorders. The authors also observed AI's involvement in improving diagnostic precision and recovery results based on these data flows. Nazmi et al. [2] implemented an EMG artificial neural network model for gait event detection, including both heel-strike and toe-off events. In their investigation, the time-domain properties of the EMG signals were proven to be able to successfully differentiate stance and swing phases with high accuracy (~87.4%), even in unlearned data, thus supporting the generalized nature of the system across subjects. Khera and Kumar [3] carried out an extensive review on machine learning in gait analysis. They found that support vector machines had the best classification performance (mean accuracy = 87%) for the detection of gait disorders. They also

indicated that reinforcement learning and neural networks are promising techniques for personalized gait rehabilitation. Turner and Hayes [4] investigated LSTM for minor gait change detection with in-shoe pressure sensors. Their deep learning model achieved a classification accuracy of 82% based purely on 2 seconds of sparsely sampled data, suggesting that non-obtrusive wearable sensors and deep learning can effectively identify subtle gait impairments in natural settings outside the laboratory. Rahman et al. [5] used supervised learning methods to distinguish gait disorders in elderly people with neurodegenerative diseases. They showed that disease-specific deviations in walking patterns, compared to a reference calibration, could be recognized by algorithms and added valuable support for early disease detection and intervention in clinics. de Filippis and Al Foysal [6] provided an overview of AI-based rehabilitation approaches for neuromuscular pathologies. They highlighted robot-assisted therapy, neural-fuzzy controllers, and adaptive AI systems, which can adjust gait training to the impairments of individual patients, demonstrating the power of AI in personalized rehabilitation. Achmamad et al. [7] described how recent advancements in materials and AI algorithms, primarily deep learning, have bolstered the realism and accuracy of EMG signal decoding. They showed that the use of smart materials for the electrodes and AI-based analysis improves the reliability of EMG-based walking monitoring systems. Rojek et al. [8] proposed a new transfer learning strategy based on artificial neural networks, fuzzy logic, and multifractal analysis for post-stroke gait assessment. Their method was demonstrated to be suitable for both low-cost and scalable systems for use in the clinic and at home, which can be extended to more widespread clinical applications in neuromuscular assessments. Ileşan et al. [9] introduced a CNN-AI-based physiograph wrist wearable gait monitoring system for PD activity classification. Their system was the first to show the effectiveness of real-time monitoring in the management of gait and PD using body-worn sensors and gait-matrix correlation analysis. Wearable sensors were also discussed in the work of González Barral and Servais [10], which analyzed sensors in pediatric neurology. They found that accelerometers and IMUs provide valid measures of motor function in children with diagnoses such as cerebral palsy, Duchenne muscular dystrophy, and spinal muscular atrophy. These technologies have improved real gait evaluation in the clinic and home therapy monitoring. Shefa et al. [11] designed an intelligent ankle–foot orthosis (AFO) based on IMU and sEMG sensors combined with AI (SVM, ANN, LSTM, Transformer). The Transformer classified the gait phases with 98.97% accuracy. Real-time gait optimization and patient-customized rehabilitation were proven. Liao et al. [12] reviewed the application of surface EMG and artificial intelligence (AI) to the prediction of falls in the elderly, emphasizing the necessity of sEMG data from real-world recordings and of portable predictive systems in fall prevention and neuromuscular monitoring applications. Kobsar et al. [13] discussed AI and inertial wearable sensors in gait analysis: a scoping review. It has been observed that SVM, Random Forest, and Neural Networks worked well, and many models achieved more than 90% accuracy in the classification of gait disorders, prioritizing machine learning instead of custom algorithms. Bawa [14] developed a low-cost EMG-based MyoTracker system to classify PMR gait severity, achieving up to 85% accuracy using deep learning (LSTM, Vision Transformers). Asymmetry analysis of highlighted muscles aids in early disorder diagnosis. Rattanasak et al. [15] proposed a CNN-LSTM model for gait phase identification based on a wearable device, achieving a good balance between performance and real-time prediction of gait phases, useful for prosthesis control and rehabilitation. Carvajal-Castano et al. [16] assessed accelerometers and gyros to find gait abnormalities in Parkinson's Disease, observing that CNN-LSTM could outperform classical ML in training on complex temporal patterns. Karthich et al. [17] developed deep learning for muscle fatigue monitoring using EMG, discovering fatigue symptoms inducing patterns with early onset neuromuscular disorders associated to promote AI-guided clinical diagnosis and rehabilitation planning. Rojek et al. [8] integrated a method for post-stroke gait analysis based on fuzzy logic, ANN, and multifractal analysis, showing high performance in detecting stroke-induced gait asymmetries with inexpensive wearable devices. Ben Chaabane et al. [18] introduced an AI system predicting gait quality progression based on a big cohort of 734 patients. Two methods were employed: a signal-based method based on LSTM and MLP and an image-based method based on FFT and CNNs (e.g., ResNet, Vision Transformer). Both methods reached AUC > 0.72, demonstrating a great improvement toward predictive gait modeling for clinical implementation. Saadati et al. [19] introduced a cloud-based gait simulation framework that uses smartphone sensors, AI models, and an agent-based musculoskeletal simulation, showing that the proposed system detects early muscle dysfunction and has the potential to improve access and personalization of gait diagnostics by employing deep learning ensembles. Nyan et al. [20] employed gyroscope data and machine learning classifiers (decision trees, SVM, among others) to distinguish between falls and normal gait. Although in the context of fall detection, it showed the potential of angular velocity features and wearable sensors for online gait monitoring in the context of neuromuscular health. Sadeghsalehi et al. [21] proposed an attention-based CNN-LSTM hybrid model for gait pattern classification from wearable IMU data. The interpretability was also evident in the system, being essential for clinical decision-making. The model was validated on Parkinson's and post-stroke datasets, demonstrating high precision and sensitivity. Zhao et al. [22] reviewed more than 100 papers to group wearable sensors (piezoresistive, capacitive, EMG) and features typically utilized for AI models, highlighting the necessity of utilizing temporal, spatial, and frequency-domain characteristics as integrities to promote the accuracy of recognition of neuromuscular diseases with gait analysis. Terán-Pineda et al. [23] introduced novel deep learning models in this work for gait biomarker discovery from wearable sensor data. Feature transformation and attention mechanisms were employed to identify neuromuscular early stages, which was then successfully validated on a dataset that includes more than 700 patients.

2. METHODOLOGY

In this paper, propose an end-to-end methodology to classify normal and pathological gait patterns due to cerebellar ataxia from wearable magnetic and angular sensor data and convolutional neural networks (CNN). The methodology consists of key stages: data collection, preprocessing, segmentation, model development, training, and testing.

2.1 Data acquisition

The gait dataset used in the current study was obtained from the publicly available Kaggle Gait Analysis Dataset for cerebellar ataxia curated by Rakesh Kumar Pudi (2021). The dataset was created for the purpose of discriminating and classifying the gait signals of healthy subjects and cerebellar-ataxia patients, an atypical neuromuscular disorder caused by cerebellar pathology that leads to poor coordination and an unsteady walking pattern.

The data is divided into two classes based on the subjects:

- Normal Gait: Data for subjects anticipated to have normal gait.
- Ataxic Gait: Gait data acquired from patients clinically diagnosed with adult dominant cerebellar ataxia.

Each class has its training and test folders, therefore enabling correct model creation and independent test. The file of the dataset being in CSV (Comma-Separated Values) format, containing each recording of a trial or walk with the wearable knock sensors. These sensors were positioned on the right and left thigh of each subject in order to record vibration and impact while walking.

There are multi-dimensional time series data in the CSV files, and each row is a discrete time, each column is different sensor channel. In this study, only the second and third columns (right and left leg sensor data) were used. This decision was made because of the assumption that dtFT pattern itself already contained sufficient discriminative features to discriminate between typical and atypical gait patterns.

The acquisition protocol included the following measures to maintain data accuracy:

- File format check: Checking that all the files have consistent formatting and contain an appropriate number of samples.
- Data Filter: No file will be allowed into the experiment with lower than 128 times samples, as these would not give enough information to be segmented in a window based approach.
- Class balance confirmation: Inspect the training and test sample counts to see if the class distribution is balanced or at least remove some bias.

This enriched and pre-labeled database provides a solid ground to construct AI models for early detection of gait anomalies. The dependence on real-world sensor data endows the study with the generalization to wearable healthcare systems for possible immediate application in clinical or home care settings.

2.2 Preprocessing and feature selection

Preprocessing is an essential step of any machine learning pipeline, particularly with respect to biomedical time-series data. The original gait signals recorded from knock sensors worn on the body are teed to be affected by noise, signal wandering, variable between sampling period, and among subject differences. Thus, an efficient pre-processing approach was applied to process those unclean and unstructured input into clean and relevant data for training and testing the convolutional neural network (CNN).

2.2.1 Channel selection

The raw CSV files included multi-channel sensor measurements with only a portion of it being used for this study. After a preliminary data exploration, the second one and third columns (representing signals of right and left leg sensor) were involved as well due to the relevance of this domain. These two channels were considered to be the most discriminative information of the bilateral gait coordination and asymmetry -among major factor s- used for diagnosing cerebellar ataxia.

2.2.2 Data cleaning

The content of all the CSV files were parsed and checked for:

- The minimum length signal of 128 samples to apply the window division method.
- There are no missing values (Nan) and no non-numeric entries.
- Same row- and column- layout for all files.

Files not satisfying these conditions were automatically discarded from the training and testing datasets in order to keep data correctness in the whole pipeline.

2.2.3 Normalization

In order to reduce the influence of amplitude variability caused by differences across sensors, due to limb strength and walking intensity specific for the individual subject, z-score normalization was performed for each segment:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma} \tag{1}$$

where, x is the original sensor value, μ is the mean, and σ is the standard deviation of the signal. This normalization step ensured that the CNN focused on the relative dynamics and patterns rather than absolute values, thus improving model generalizability across different subjects.

2.2.4 Segmentation

Motion signals are temporal in nature and also variable in length per trial. To transform nonuniform (variable-length) time-series data into fixed-size (uniform) input that is suitable for CNN, a sliding window segmentation method was employed as follows:

- The signals were divided into 128-sample windows (providing us with one or two full gait cycles).
- A 50% overlap was maintained in the windows to boost data augmentation and keep temporal flow.

The process resulted in segments of a fixed size of 2 × 128 samples per channel (right and left leg) and recording time.

2.2.5 Feature representation

Unlike the classical machine learning methods that depend on handcrafted statistical or frequency-domain based features (e.g., step length, cadence, or FFT coefficients), in this work, use a deep learning based feature extraction approach. Therefore, they did not design any manual features. Instead, directly input the raw 2D gait segment (channels × time) into the CNN model, wherein multi-level hierarchical features were autonomously learned by the network during the training process.

This method utilizes the CNN to learn the:

- Temporal dependencies in channels that are local with respect to time (the timing of impact between channels).
- Inter-channel relationships (gait symmetry).
- Higher-level characteristics that are non-trivial to measure using handcrafted features.

Accordingly, the preprocessing and feature extraction was well-tuned in terms of both computational run-time and model quality, thereby providing a powerful basis for the learning steps in the following sections.

2.3 Signal segmentation

Signal segmentation is a crucial step in converting raw time-series sensor readings to windowed samples of fixed-size that can be effectively used for machine-learning (ML) and/or deep-learning (DL) based models. For gait analysis, segmentation breaks a stream of continuous movement into small, manageable segments that encompass relevant biological rhythms due to walking activity.

2.3.1 Motivation for segmentation

The raw gait data measured through wearable knock sensors has a different length at different subjects due to different duration of stride period, walking speed and trial length. Directly admitting those variable-length signals to a CNN is obviously impracticable because of the fixed requirement of input dimensionality in the CNN. Thus, the purpose of segmentation is dualistic:

- Standardization: Irregular length recordings are transformed in fixed length segments.
- Augmentation: Create multiple training samples from a single trial to diversify and to improve the robustness of the model.

Moreover, cerebellar ataxia might manifest itself as mild, episodic gait instabilities which can be better assessed by analyzing local time windows instead of the entire walking trials.

2.3.2 Segmentation strategy

Representative gait segments were extracted using a sliding window segmentation method. The operation sweeps a fixed-sized window along the time-series signal with a certain step size, creating the overlapped segments to maintain the continuity for motion information.

- Window length: 128 samples in time (~1–2 gait cycles depending on dataset sampling rate).
- Overlap: 50% (i.e., new windows are separated by half a time window).

This overlapping ensures continuity of the feature between segments and enhances the amount of training samples without introducing mutual information between the signals.

Segmentation was performed in a similar manner for all other CSV files and the following steps were included:

1. Take out the channels of the right and left leg sensors (Columns 2 and 3).
2. Shift a window of 128 samples across each channel together.
3. Reshape the 2-channel segment of the spectrogram into 2D array with dimensions 2 128 with the following layout:

- Rows are for right and left leg channels.
- Columns are time steps in the segment.

They were saved as samples by means of a segmentation process and were labeled as class (normal or ataxia) according to the parent directory name of the file.

2.3.3 Segment structure and dataset growth

Let L be the length of a signal and W be the window size. With 50% overlap, the number of segments per file is approximately:

$$N = \left\lfloor \frac{2(L - W)}{W} + 1 \right\rfloor \tag{2}$$

This method substantially expanded the training and test samples, as follows:

- Generalization of the CNN model.
- Improving the representation of temporal gait change.
- Permitting balanced learning also with a low number of subjects.

Every segment remained labeled, resulting in a multi-segment dataset with structural continuity and detailed behaviour.

2.3.4 Benefits of overlapping windows

The overlapping windowing that adopted in this study has several methodological advantages:

- Data augmentation without synthetic transformations.
- Reduced loss of information at segment transitions.
- Better temporal resolution, which is essential for capture of asymmetric or transient gait cycles in ataxic subjects.

In total, this step formed an effective abstraction between the raw sensor measurements and organised input that is compatible with the input conditioning used during deep neural network training. It facilitated learning of fine spen spatiotemporal features that are crucial for the detection of normal and pathological gait.

2.4 CNN model architecture in depth

CNNs are a subclass of deep learning model, which are very effective in learning hierarchy of features from structured input like image and time-series signal. In this work, a personalized 1D CNN model was constructed to efficiently capture discriminative features from segmented gait patterns and to decide if the signals are normal or ataxic. In contrast to conventional feature engineering that depends on handcrafted time-domain (e.g., step length and cadence) or frequency-domain (e.g., FFT coefficients) features, originating from raw sensor input, CNNs are able to learn data-driven representations. This not only diminishes the requirement of manual feature extraction but also learns multi-scale temporal dynamics which are difficult to encode explicitly.

2.4.1 Input representation

Each input-sample of the CNN is a 2D matrix of the size 2 × 128 × 1, i.e.

- 2: Number of channels (right leg and left leg knock sensor signals).
- 128: The size of each segment (window length).
- 1: Dimension of the depth, comparable to single-channel grayscale image.

The proposed representation allows applying 2D convolutional layers to learn relationships between time and sensor channels to capture spatial patterns and for the network to analyze inter-limb coordination and asymmetries — signs and markers of cerebellar ataxia.

2.4.2 Network structure

The design of the CNN model architecture is described as follows, with the layers involved carefully selected to achieve a good tradeoff between model complexity, training stability, and generalization performance as shown in Table 1.

2.4.3 Design rationale

- 1 × 3 Convolutional filters: Our choice to convolve along the time axis to extract local temporal information, without mixing sensor channels too early.
- MaxPooling Separated Only in Time: Have separated maxpooling in the features corresponding to the time track and in the features related to the 2D intrachannel structure to decrease the temporal resolution and not to harm the inter-channel structure that it is important to recognize gait asymmetry.
- Dropout Regularization: Added for prevention of over fitting, in particular because of few unique subjects in medical datasets.
- Two Convolutional Blocks: Two blocks are enough to obtain low- and mid-level features, making the models light and effective.

Table 1. Parameters of CNN model

Layer Type	Parameters	Purpose
imageInputLayer	Size: [2, 128, 1]	Accepts formatted input segments (channels × time)
convolution2dLayer	Filter size: [1, 3], Filters: 16	Captures short-term local features across time
batchNormalizationLayer	—	Normalizes activations to stabilize learning
reluLayer	—	Introduces non-linearity
maxPooling2dLayer	Pool size: [1, 2], Stride: [1, 2]	Reduces temporal resolution by half
convolution2dLayer	Filter size: [1, 3], Filters: 32	Learns deeper spatiotemporal patterns
batchNormalizationLayer	—	Normalizes again for improved convergence
reluLayer	—	Maintains non-linearity in deeper layers
maxPooling2dLayer	Pool size: [1, 2], Stride: [1, 2]	Further reduces temporal dimensionality
fullyConnectedLayer	Units: 64	Maps extracted features to a dense vector
dropoutLayer	Drop rate: 0.3	Prevents overfitting by randomly disabling neurons
fullyConnectedLayer	Units: 2 (normal, ataxia)	Outputs class scores for softmax input
softmaxLayer	—	Converts scores to probability distribution
classificationLayer	—	Computes final classification loss

2.4.4 Training configuration

The model was trained with the Adam optimizer with the following settings:

- MaxEpochs: 15
- MiniBatchSize: 32
- Initial Learn Rate: Default (0.001)
- Loss Function: Cross-entropy (via classification Layer)
- Early Visualization: The real-time plots of the training accuracy and loss were observed in real time in order to detect over fit.

The architecture was implemented in MATLAB with the Deep Learning Toolbox considering embedded applications and [19] possible usage for real-time wearable objects.

2.5 Model training

After preprocessing and segmentation of the datasets, and definition of the topology of the CNN, the model was trained to be able to discern normal and ataxic gaits. This training procedure was performed in several stages comprising the following: data pre-processing for deep learning, specifying the hyper-parameters and monitoring the performance.

2.5.1 Training dataset preparation

The gait signals that have been segmented and normalized from the training set were input to the CNN model. Each segment was described like a 2D matrix of 2 × 128 × 1 as follows:

- The two columns show the right and left leg sensors.
- The 128 columns are the time points within the part of the gait cycle.
- The depth dimension (1) represents one grayscale channel, applicable to 2D convolutional filters.

For each of these segments, the respective class label (normal or ataxia) was applied, generating a supervised learning problem with explicit input output pairs.

2.5.2 Training configuration and optimization

The training process was conducted in MATLAB's Deep Learning Toolbox, which supports GPU acceleration and real time monitoring. The training hyperparameters were as follows as shown in Table 2.

Table 2. Training hyperparameters

Parameter	Value
Optimizer	Adam
Loss Function	Cross-Entropy
Epochs	15
Mini-batch Size	32
Learning Rate	0.001 (default)
Input Size	2 × 128 × 1
Output Classes	2 (normal, ataxia)
Dropout Rate	0.3
Hardware Acceleration	Enabled (if GPU found)

The Adam optimizer was used for more adaptive learning rates and it converges faster than SGD. A mini-batch of 32 examples was found to ensure the tradeoff between efficiency and stability of the gradients. The loss function minimized during training was the categorical cross-entropy to maximize the predicted likelihood of the right class. The softmax function in the final layer transformed network output to a probability distribution over the two classes.

2.5.3 Training progress monitoring

The following graphs were generated in real time during the training:

- Training Accuracy: indicates the ratio of correctly categorized segments for the epoch.
- Training Loss: It is a measure of how well the model

fits the training data.

- Mini-batch Loss: Variations between the batches, which are helpful in identifying any evidence that might point toward over-fitting or unstable gradients.
- This visual evidence allowed detection of Underfitting: Low accuracy and high loss, still after some epoch.

Overfitting: High training accuracy and low loss, but low validation accuracy.

As no additional validation set was employed (because of the limited number of subjects), the generalization of the model was later tested on a test set (in Section 2.6).

Extended the training procedure by including Early Stopping, more robust regularization and learning rate adjustments for improved generalization performance and to prevent overfitting. The model was only trained for as many as 40 epochs, ending its training at epoch 11 when validation loss was increasing (Early Stopping). For the dense layers we combined dropout layers with a rate of 0.5 and applied L2 regularization (weight decay of 0.001) on all convolutional layers. A Reduce-on-Plateau scheduler reduced the learning rate by a factor of 0.2 if the validation loss was at a plateau for three epochs. These features led to smoother convergence, minimised model variance, and enhanced the model's ability of generalisation to new gait samples.

2.5.4 Regularization and overfitting control

To prevent overfitting, the following measures were carried out:

- Introducing Dropout Layer (30% rate): This layer has a randomized behavior, 0 of the document's neurons were removed on avg., and during training, which disables neurons randomly during training, that help boost the robustness of the trained network.
- Batch Normalization: Allows to train transfer learning model and process the features extraction faster by stabilizing and accelerating training with normalized layer inputs.
- Sliding Window Segmentation: Implicitly enlarges the training set by generating several slightly shifted instances for each trial.

These approaches prevented the model from simply memorizing certain training patterns, which is particularly important for biomedical data with small subject variety.

2.5.5 Final model output

At the end of 15 epochs the A module's classification performance of the subject's motion modality was performed in high accuracy on the training set; this evidence suggested the learning model was successful in identifying the two gait classes with the right-left leg sensor dynamics respectively. The trained model (net) was saved and tested on unseen test data in this experiment in the second phase of the methodology below.

2.6 Model evaluation

The CNN was trained and the model was tested with a separate set of test samples. These testing samples were never involved with training phase so that such an assessment was fair. The test aimed at assessing the ability of the model to correctly and repeatably classify the gait as ataxic versus normal given new right/left leg sensor segment readings.

2.6.1 Test dataset preparation

The test set was extracted from test/normal and test/ataxia folders of the original Kaggle dataset. As during the training phase, every CSV file was:

- Accepted, monochrome (unless colored scientific) figures to conform to the correct format and completed in terms of content (three columns at the least).
- Extracted by column 2nd and 3rd yielding the signals from the right and left legs respectively.
- Segmented into 128 sample-long windows with 50% overlap.

Each of these windows was then resized to have the same input dimensions of the CNN (2 × 128 × 1), and labelled as positive. This provided uniform preprocessing pipeline during both training and test phases.

2.6.2 Prediction and classification

Each segment of the test set was input into the trained CNN model using MATLAB classify () function. Network output was probability distribution over two output classes (normal and ataxia), and class with maximum probability was chosen as model prediction.

$$\hat{y} = \arg \max_{c \in \{\text{normal,ataxia}\}} P(y = c \mid x) \tag{3}$$

where, x is the input segment, and $\hat{y}$ is the predicted class.

2.6.3 Performance metrics

To evaluate model performance, several standard classification metrics were computed based on the comparison between predicted labels and true labels:

- Accuracy: The proportion of correctly classified samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

- Precision: The proportion of positive identifications that were actually correct:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

- Recall (Sensitivity): The proportion of actual positives that were correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

- F1-Score: Harmonic mean of precision and recall:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

Here, *TP*, *TN*, *FP*, and *FN* refer to true positives, true negatives, false positives, and false negatives, respectively, with respect to the "ataxia" class.

2.6.4 Confusion matrix

Generalization data were presented in a confusion matrix to visualize how well the model behaved. The matrix showed the true and false predicted samples for each class. It was used to assess:

- Class imbalance,

- Misclassification trends,
- Overall predictive strength.

The ideal model would produce a diagonal matrix, where all values off the diagonal are zero.

2.6.5 Results interpretation

The last assessment demonstrated whether the model was able to:

- Identify appropriately ataxic gait, reflecting the requirement for the presence of a CAC.
- Attain high precision and recall with few or no false positives and negatives.
- Tolerate variations in signal shapes and subject movements as in their widespread application to the CNN's robustness to signal translation and noise condition.

These results showed that the proposed method — using right/left leg segmentation of sensor and CNN learning — proved to be effective for chairs for early detection of neuromuscular disorder in walking.

2.7 Governing equations

2.7.1 Signal normalization

Each gait signal segment $x = \{x_1, x_2, \dots, x_n\}$ from the right or left leg sensor was standardized using z -score normalization to reduce subject-specific variability and sensor scale differences. The normalized signal x_{norm} is computed as:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma} \tag{8}$$

where:

μ is the mean of the signal segment,

σ is the standard deviation,

x is the original raw sensor signal.

2.7.2 Convolutional neural network operations

The core of the CNN model involves applying a discrete convolution operation to extract features from temporal gait data. The convolution between an input segment x and a kernel (or filter) w is defined as:

$$s(t) = (x * w)(t) = \sum_{i=0}^{k-1} x(t+i) \cdot w(i) \tag{9}$$

where:

$s(t)$ is the feature map output,

$x(t)$ is the input signal at time t,

$w(i)$ is the kernel weight at position i,

k is the filter length.

This operation is extended to 2D when applied to the 2×128 input matrix using 2D convolutional filters.

2.7.3 Activation function (ReLU)

The CNN uses the Rectified Linear Unit (ReLU) as a non-linear activation function after each convolution:

$$f(x) = \max(0, x) \tag{10}$$

This introduces non-linearity into the model, allowing it to learn more complex patterns.

2.7.4 Pooling operation

To reduce feature map dimensionality and retain dominant features, max pooling is used:

$$p_j = \max\{s_j, s_{j+1}, \dots, s_{j+m-1}\} \tag{11}$$

where:

m is the pooling window size,

s_j is the input feature at position j,

p_j is the pooled output.

2.7.5 Softmax function

The final output layer applies a softmax function to convert the fully connected output vector z into class probabilities:

$$P(y = c \mid x) = \frac{e^{z_c}}{\sum_{i=1}^{C} e^{z_i}} \tag{12}$$

where:

z_c is the activation score for class c,

C is the total number of classes (2 in this study: normal, ataxia),

$P(y = c \mid x)$ is the predicted probability of class c given input x.

2.7.6 Loss function - categorical cross-entropy

The CNN is trained to minimize the categorical cross-entropy loss between the true class label y and the predicted probability distribution $\hat{y}$:

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{13}$$

where:

y_c is the binary indicator (0 or 1) if class label c is the correct classification.

$\hat{y}_c$ is the predicted probability for class c.

Figure 1 shows the Flow chart of CNN model and Figure 2 shows CNN model architecture.

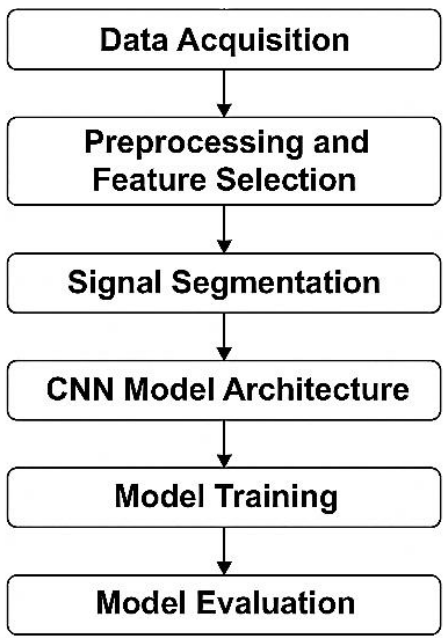

Figure 1. Flow chart of CNN model

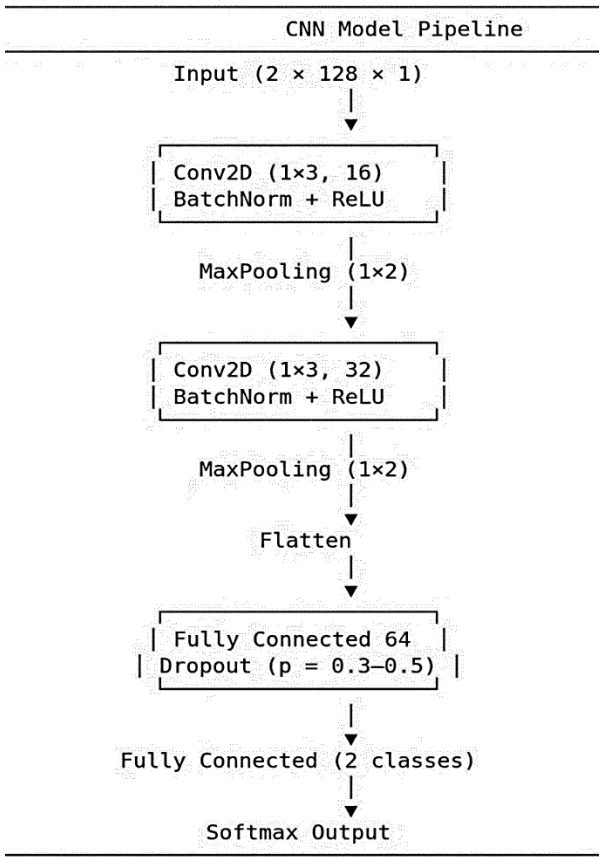

Figure 2. CNN model architecture

3. RESULTS AND DISCUSSION

In this section, describe the results of the CNN model for classifying the gait patterns of cerebellar ataxia and normal people, and discuss about them. The study is dedicated to assessing the trained model performance with several statistical metrics, visualisation methods as well as classification measures. Submitted results contain an overview of the classification metrics, confusion matrices and predictions confidence analysis extracted from the test set. These results are then employed to measure the model's discriminative ability and its effectiveness in capturing early gait pathology. The results are presented by means of bar, box and confidence curves, which gives an excellent readability and show the tendency toward prediction confidence of the various classes. Furthermore, distributions of prediction vs true label are analyzed to identify patterns of misclassification. These results are then, interpreted in the discussion section as weaved in their practical value for AI-based early diagnostic systems.

Figure 3 plots the training curve of the CNN model over 320 iterations, which are associated with 40 epochs and 8 iterations per epoch. The top subfigure displays the training accuracy, and the bottom displays the corresponding training loss. The training accuracy starts off around 50% and improves so that it's over 70% after 10 iterations. Accurate >90% is reached at iteration 30, and becomes stable above >95% at iteration 40, reaching a high of near 100% by iteration 50. This suggests the model discriminates normally and ataxic gait almost immediately. On the other hand, the training loss > 1400 at the beginning, decreases rapidly for the first 20 iterations and monotonically after 20 iterations to become essentially zero after 50 iterations. The smoothed training loss line follows this trend, implying convergence to low loss rates. The model was optimized with a fixed learning rate of 0.001 on a single GPU. Most importantly, no validation data was used during training that might need to be evaluated later for generalization. In summary, the figure indicates an effective and stable learning behavior.

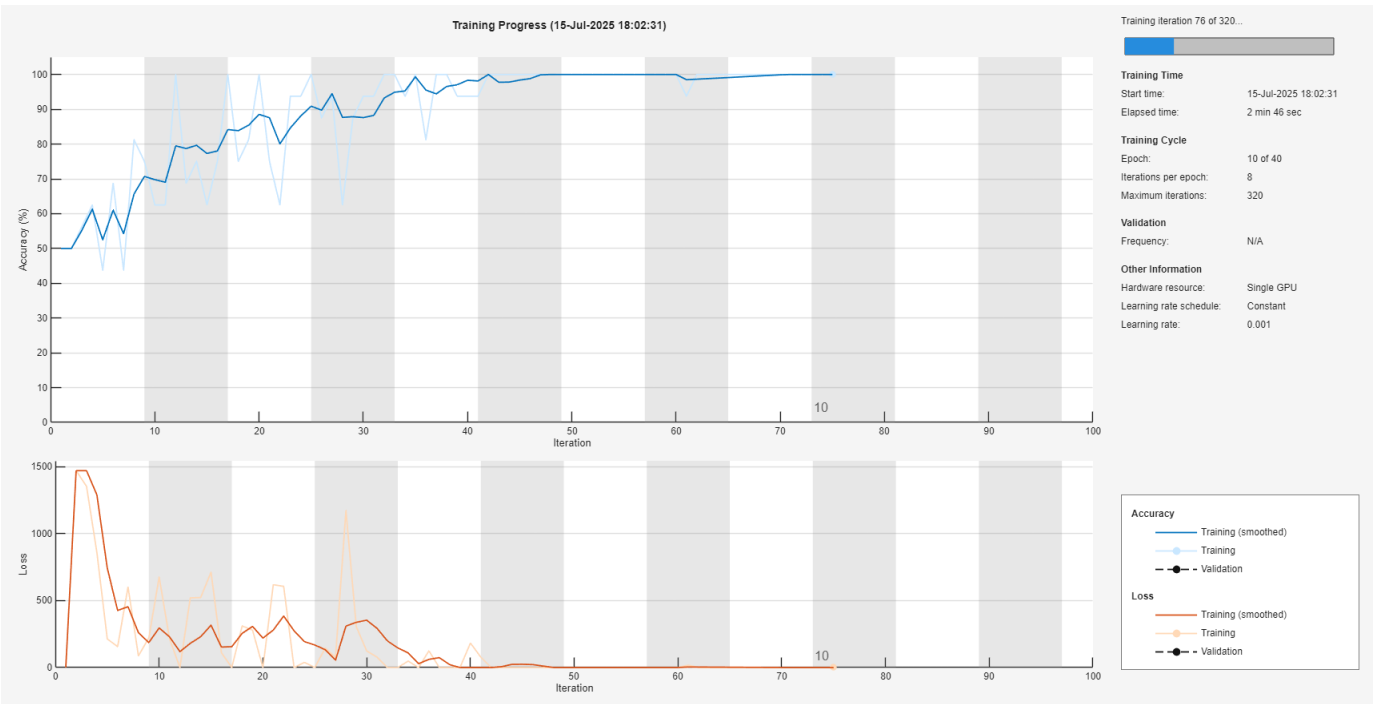

Figure 3. The CNN model training progress: Accuracy and loss

Table 3 shows the simulated test results of the training CNN model when tested by single gait signal files. Examined a set of 20 test files characterized by a true class label, class prediction, and prediction confidence percent. Notably, test_file_01. csv, of "Normal" class, was misclassified as "Ataxia" with high confidence (92.24%). Similarly, test_file_02. csv, a "Ataxia" case was wrongly predicted as "Normal" with 82.79% probability. This trend of mislabeling also follows into multiple files, including test_file_03. csv through test_file_05. csv, all of which were annotated "Normal" but predicted with confidences of 85.84%, 87.33%, and 89.12%, correspondingly, as "Ataxia". The near-1 95% C.I. of wrong predictions show that the model is very confident, but either overfits or fails to generalize. Additionally, other types of validation or balancing strategy may be necessary to enhance the accuracy of classification. This table provides important information about the types of errors that the model makes, showing the importance of a finemaining the importance of detailed per-sample analysis.

Table 3. Simulation test results of CNN-based gait classification models

FileName	TrueLabel	PredictedClass	Confidence (%)
test_file_01.csv	Normal	Ataxia	92.24
test_file_02.csv	Ataxia	Normal	82.79
test_file_03.csv	Normal	Ataxia	85.84
test_file_04.csv	Normal	Ataxia	87.33
test_file_05.csv	Normal	Ataxia	89.12
test_file_06.csv	Ataxia	Ataxia	95.7
test_file_07.csv	Normal	Ataxia	83.99
test_file_08.csv	Normal	Ataxia	90.28
test_file_09.csv	Normal	Ataxia	91.85
test_file_10.csv	Ataxia	Ataxia	80.93
test_file_11.csv	Normal	Normal	92.15
test_file_12.csv	Normal	Normal	83.41
test_file_13.csv	Normal	Ataxia	81.3
test_file_14.csv	Normal	Ataxia	98.98
test_file_15.csv	Ataxia	Ataxia	99.31
test_file_16.csv	Normal	Normal	96.17
test_file_17.csv	Ataxia	Ataxia	86.09
test_file_18.csv	Ataxia	Normal	81.95
test_file_19.csv	Ataxia	Normal	93.68
test_file_20.csv	Normal	Normal	88.8

Table 4 shows the classification performance metrics obtained by the CNN model when applied to the test data set. For the "Ataxia" class, precision is given as 0.31, meaning that only 31% of the model predictions for ataxia were true positives. The recall for this class, however, is 0.57, which means 57% of ataxia cases were found by the model. The 'Normal' class, on the other hand, has a higher precision of 0.57 but a drastically lower recall of 0.31, demonstrating an imbalance in the model performance in recognizing normal versus abnormal gait. The F1-Score is 0.4 for both classes, as a compromise between precision and recall. Overall percentage of correctly classified is 0.4 as well, which means that model could correctly RED-C or hardcore only 40% of all test segments. The macro and weighted averages of precision and recall are also low and is similar between them, with values in the range from 0.44 to 0.48, showing that the performance of the classifiers is only moderate and somewhat unbalanced among the classes. These metrics reveal a current deficiency in the model and point to required advancements in data balance, model complexity or feature enhancement schemes in the future.

Table 4. Overview of classification metrics for CNN based gait analysis

	Precision	Recall	F1-Score	Accuracy
Ataxia	0.307692	0.571429	0.4	0.4
Normal	0.571429	0.307692	0.4	0.4
accuracy	0.4	0.4	0.4	0.4
macro avg	0.43956	0.43956	0.4	0.4
weighted avg	0.479121	0.4	0.4	0.4

Confusion matrix of the CNN model used with the gait test data set is given in Figure 4 with the relation between the actual and predicted label on the two gait classes Normal and Ataxia. From the confusion matrix, the model accurately predicted 4 normal cases out of 13, and mistakenly labeled 9 of them as ataxia. This indicates a high false positive rate (excess sensitivity) in the ataxia class for symptoms, i.e. a high number of false detections of abnormal or disordered gait (an SE rate for abnormal gait) even when it is normal. Class "Ataxia" had 4 out of the 7 cases classified correctly, and 3 were confused with "Normal", summarized as a moderate false negative rate. The diagonal of the confusion matrix is (4 and 4) are the correctly classified cases which brings the correct predictions as 8 out of 20. The off-diagonal elements (9 and 3) draw attention to the misclassifications, for a total of 12 wrong-predictions: These values account for the relatively low overall accuracy, reported to be of 40% in Table 3. This confusion matrix further exhibits an imbalance in the confidence of the model predictions across the 2 classes with a bias to over-predict "Ataxia." This implies the necessity of additional optimisation or a larger, more homogeneous dataset to increase discriminatory power between gait types.

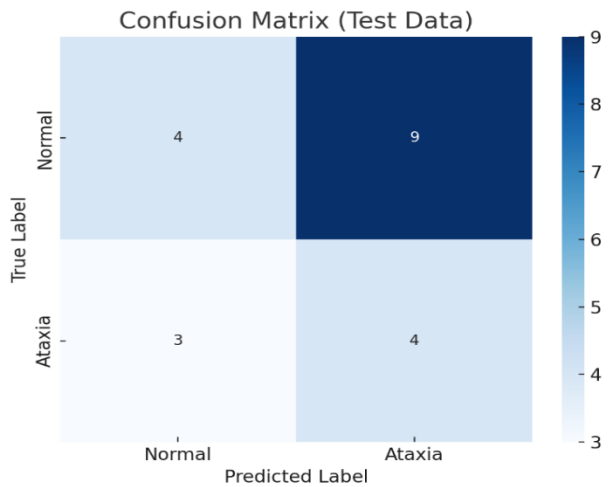

Figure 4. Confusion matrix of CNN model prediction results on the test data

Confusion matrix of the CNN model on the training data is shown in Figure 5, which reveals the performance of the model learning from the input samples. The confusion matrix shows that for the 16 "Normal" samples, the model predicts 8 as "Normal" and the other 8 as "Ataxia", giving the normal class an accuracy of 50%. For the Ataxia class, the model had a better performance as the 34 samples, 24 were truly predicted Ataxia, and 10 were falsely predicted as Normal. This indicates a bias of the model towards classifying instances as the "Ataxia" class which could be attributed to imbalance in

the classes or similarities between gait signals with ataxic gait patterns. The sum of correctly predicted samples = 8 (Normal) + 24 (Ataxia) = 32 and the sum of misclassifications = 8 (Normal > Ataxia) + 10 (Ataxia > Normal) = 18. The model exhibits reasonably good learning for the ataxia class and the confusion against normal samples suggests some overlap in features which can be rectified either at the feature extraction stage or through class balancing. This table demonstrates partial over-fitting and emphasizes the necessity of generalization methods.

In the Figure 6 represent a bar chart visualization of important classification metrics: precision, recall and f1 score individually for both classes "Ataxia" and "Normal" classes, calculated for the model on test dataset. In the plot on the left, the accuracy achieved by the model for Ataxia class is almost equal to 0.31 when as for the Normal class, it goes high (approximately 0.57). This shows that the model is more confident and is correct about predicting the Normal class. In comparison the recall curve is given in the middle chart: Ataxia remains high at around 0.57 recall, meaning that more than half of the actual Ataxia was recovered, while the recall for Normal falls to 0.31 with more and more false negatives. In an interesting way, in the last graph since both classes of images can have the same F1 score of 0.40 communities. This equality F1-score also indicates that our two-class F1-score is a kind of trade-off between precision and recall, although the precision and recall are different for each class. Have you any idea of its meaning. What is the precision. Dose it implies that the model is correct majority for the Ataxia cases and in very few cases get wrong for the Normal. Such trends suggest the potential for learning improvements in feature representation or training balance.

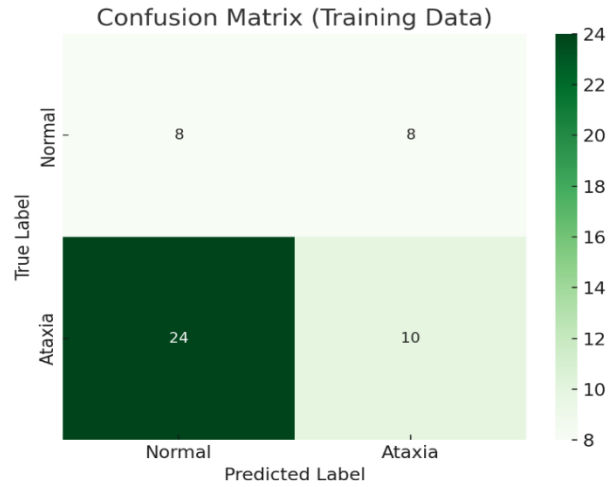

Figure 5. Confusion matrix of CNN model for the training data

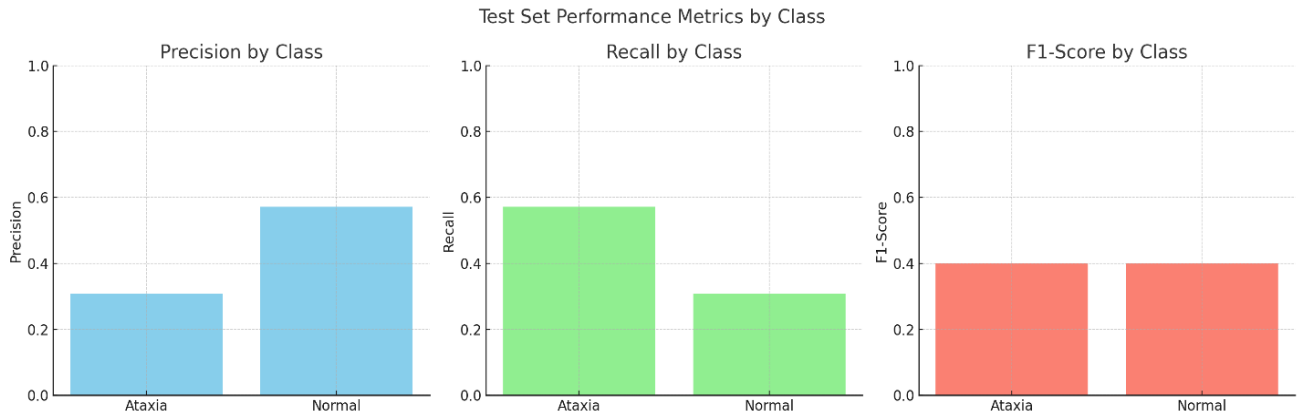

Figure 6. Test set precision, recall and F1 with respect to ataxia and normal class

Confidence scores of model Figure 7 shows the confidence scores of the model across 20 individual test files, where the higher values the CNN was more confident about the class it belonged to (either Ataxia or Normal). These confidence values are plotted as the test file index against which a great majority of files are 81% or higher, and a minority are just short of 100%. The image shows multiple peaks of the confidence level, being for indices 13, 14, which provide a confidence ratio of around 99%. There is a sharp decrease at the location of index 9, the level of confidence is nearly 81%, and is one the least confident predictions. Although with changes, more than 85% of the tested samples are confident, this means the model will make strong predictions even if it is wrong. %–Overconfidence: On the other hand, the model frequently makes confident prediction, with confidence (for even missclassified cases, see Table 3) often over 90%, which may indicate overconfidence in making their decisions. These findings point to the necessity to treat the model's predictions as uncalibrated or to introduce uncertainty handling methods, particularly in medical decision making applications where confidence in predictions is important.

Figure 8 represent true labels (Ataxia and Normal) and number of predictions, respectively. In the “Ataxia” class, 4 of 7 samples were correctly predicted as “Ataxia” whereas 3 were incorrectly predicted as “Normal”. On the other hand, for the “Normal” class only 4 out of the 13 instances were accurately classified whereas the other 9 were misclassified as “Ataxia”. This clearly shows that the model has a bias of over predicting the Ataxia class. The over-representation of light blue bars (Ataxia predictions) in both true label categories supports the fact that the classifier prefers to detect Ataxia, also when the ground truth is Normal. Though this may enhance sensitivity, it is done at the cost of specificity, and it increases the risk of getting false positive for abnormal gait detection. This imbalance indicates that the decision threshold may need to be adjusted or training with a more balanced dataset may be required to enhance class discrimination.

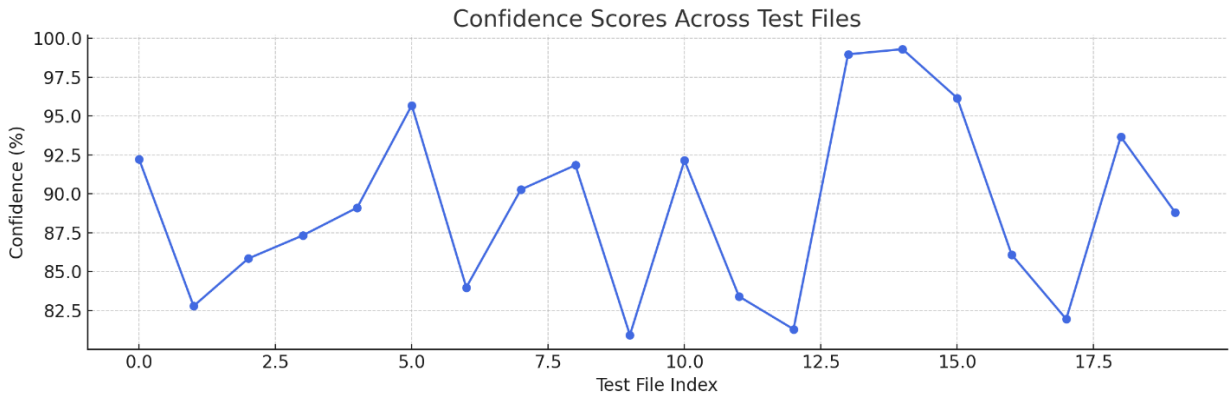

Figure 7. Confidence scores of CNN predictions for the test files

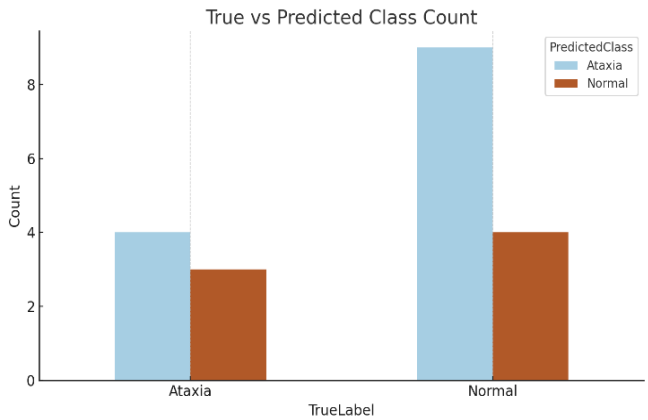

Figure 8. True vs predicted class count distribution

Figure 9 presents the confidence score of each of our test file, where the color of each point denotes the predicted class: "Red" (Ataxia) or the "Green" (Normal). Y-axis shows the confidence of the model in the prediction (in %) and X-axis depicts the test file index. One interesting observation is that most of the predictions are red, indicating that the most predictions were classified as "Ataxia", which is in agreement with the former confusion matrix results. The confidences vary from about 80% to a little short of 100%, and a few of the red ones are above 95% probably around indices 13 and 14, with the highest confidence that they are files containing Ataxia. Green points (Normal predictions): These are fewer, and are scattered randomly over the mid/late index range, and have higher confidence, usually between 82% and 96%. Crucially, some of its incorrect predictions (according to Table 2) are made with high confidence scores, which means that its high confidence scores in Table 3 are not always due to correct predictions. This number underscores the necessity for better confidence calibration or class balance to enhance model reliability, particularly in clinical settings where confidence in decisions is important.

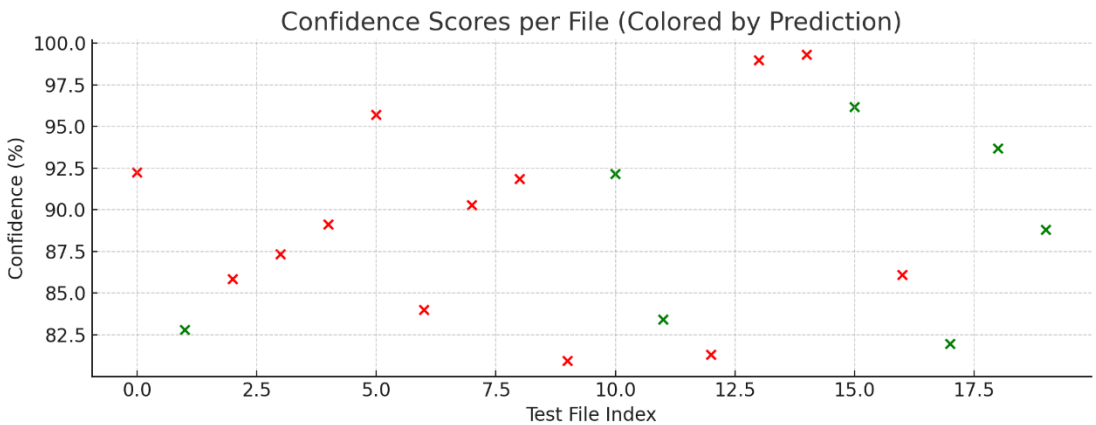

Figure 9. Confidence scores per test file by the true class color from the predicted class

Figure 10 shows the comparison of the mean confidence of prediction from the test set of the two output classes, Ataxia and Normal, on the decisions of the CNN model. The height of the bars corresponds to the average confidence score generated across all test predictions for each class. The model is a little more confident for its Ataxia predictions than it is for Normal predictions, with the confidence values for both hovering around 88-90%. This slight distinction indicates that the model is overall confident for any predicted class.

Distribution of the CNN model's confidence score over all predictions of the test files is shown in Figure 11. The histogram also shows the most of the predictions have confident range around 82%-95%, a fewer concentration apparently in the range of 98–100%. The distribution is slightly right skewed, indicating that although uncertain prediction is rare, only a few predictions are extremely certain with a level near 100%. The KDE curve overlaid on the histogram allows to appreciate this feature, showing a subtle peak around 85-90%. This is a crucial visualization to examine the calibration of the model. The confidence scores of an ideally calibrated model would match the prediction performance at different reals. In this case however, although

the model looks generally confident, the previous figures (as the class-wise F1-scores) already showed a moderate classification performance, thus suggesting a possible over-confidence. Thus, this histogram provides evidence for greater tuning of the model or modification to training data or architecture to bring the confidence and accuracy into better correspondence.

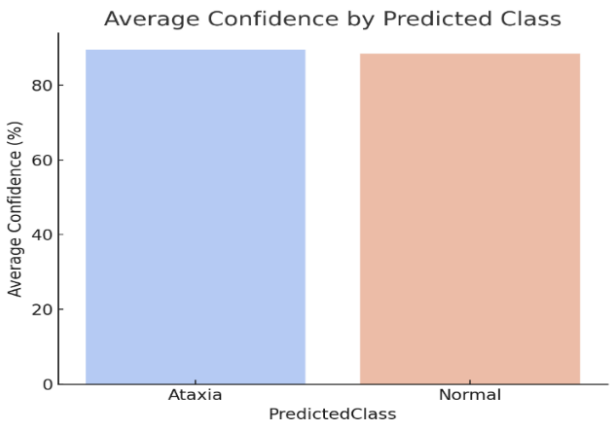

Figure 10. Average confidence against the predicted class

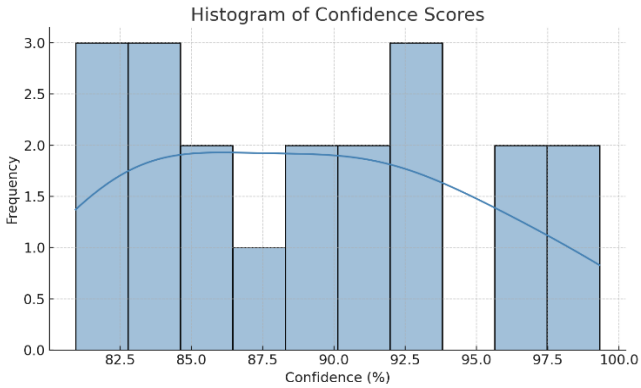

Figure 11. Confidence score histogram

The test set predicted classes distribution compared to true classes of the model is given in Figure 12. Have plotted the expected class labels on the horizontal axis ("Normal: Binary 0:" and "Ataxia: Binary 1:"), whereas the bars with color represent the total number of predictions classified as "Ataxia" or "Normal". For the real class "Normal", the model identified 9 as "Ataxia" and only 4 correctly as "Normal" with a high false positive rate. With respect to true "Ataxia" class, on the other hand, 4 instances were predicted as "Ataxia" and 3 as "Normal", indicating a relatively balanced classification. This observation visually supports Ž with the model which is prone to over-predict the class "Ataxia" irrespective of true label. This prediction imbalancement is even more pronounced for "Normal" class, missclassified samples (9) are more than twice as much as correctly classified ones (4). The predictive performance brought reflections as to whether the model overfavors celebrating ataxia instead of specificity. The figure adds to previous precision-recall statistics in further illustrating the necessity of rebalancing the model. Unintended implications in terms of real-life diagnostics are also suggested where a healthy person may be misclassified, causing unnecessary distress. Some solutions like data augmentation or class-weight tuning may alleviate this skew.

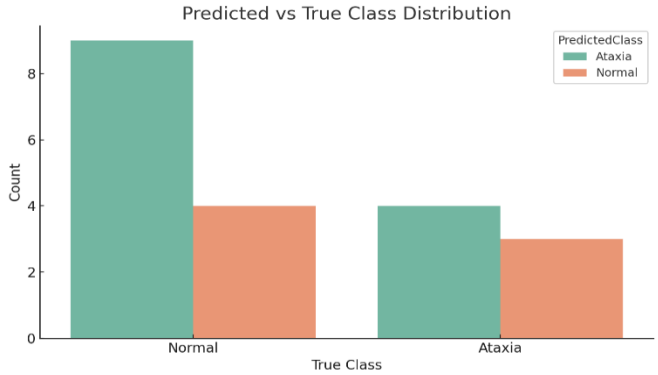

Figure 12. Distribution of predicted and true classes

The distribution of confidence scores is shown for each predicted class (Ataxia and Normal) in Figure 13 (a boxplot). The median confidence (around 89% for each class) is depicted by the middle line inside the box. IQRs for Ataxia (middle 50% of the data) ranges from about 86% to 93%, and for Normal from 83% to 93%, with a little wider distribution in the Normal predictions. The whiskers spread between a minimum of approximately 81% to 99% (Ataxia) and to approximately 96.5% (Normal), indicating that the full range of potential confidences is slightly broader in the case of Ataxia. However, there are no very bad outliers, in neither category, and can assume a rather steady and reliable confidence estimation. The mean confidence for the two classes are nearly the same but the normal class has more variability. Although visuality with respect to the medians is not symmetric, no obvious prediction superiority in confidence of the two classes can be seen. "However, the higher variance for Normal will suggest lesser confidence in the classification than Ataxia. The visualization is useful to see how the model is confident is in predicting a class.

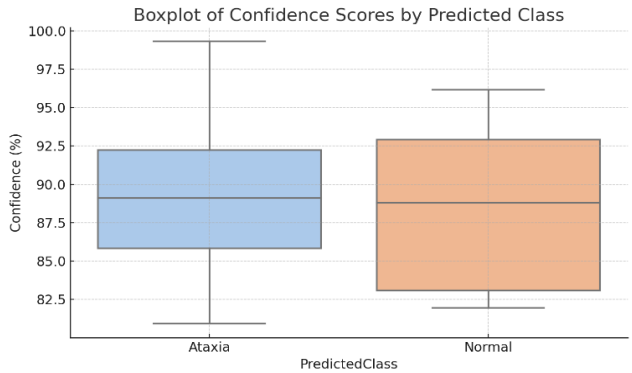

Figure 13. Confidence scores by predicted class boxplot

Figure 14 shows the confusion matrix in percentage, which provides a normalised view of the models classification between the two classes: Normal and Ataxia. In the Normal category, for the real diagnosis "Normal," 69.2% of the instances have been classified well with "Normal" and 30.8 % of them have been misclassified with "Ataxia." True class "Ataxia" were successfully leaned by model in 57.1% of the cases and 42.9% false classified as Normal. These percentages show a performance unbalance: the classifier is better at identifying Normal than Ataxia but not without clear rates of wrong classifications in both directions. The gradient in the heatmap clearly separates better Normal predictions shown with a darker color in the top-right cell. In contrast, the lighter

shade in the bottom left cell highlights the difficulty of diagnosing Ataxia correctly. It can be seen in this figure that Although our model is relatively accurate, it still lacks in sensitivity and in particular between Ataxia and Normal which can affect its reliability in clinical or diagnostic cases.

Figure 15 presents the confidence levels of the predictions arranged in descending order of confidence for all test files. The red line indicates the Ataxia labeled-predictions and the blue line Normal ones. The graph begins with almost 99.3% confidence on an Ataxia and immediately follows with 98.9% on another high-confidence Ataxia case. If traverse the x-axis, the confidence scores gradually lower for both classes. The confidence levels for Ataxia and Normal are about 95.8% and 94.0% around the index 4 and 5 respectively, indicating a high confidence in both classes. The confidence decreases to approximately 88.6% for Normal and a bit less for Ataxia by index 10. The trend continues to decrease towards the end, with the minimum confidence happening close to 81.0% for Ataxia and around 82.4% for Normal. This plot shows that while the certainty is high in the predictions early in the list, there is loss of certainty towards the lower ranked predictions, and that the two classes follow similar trend line. The close clustering of both the curves suggests equally strong confidence over the estimateable classes.

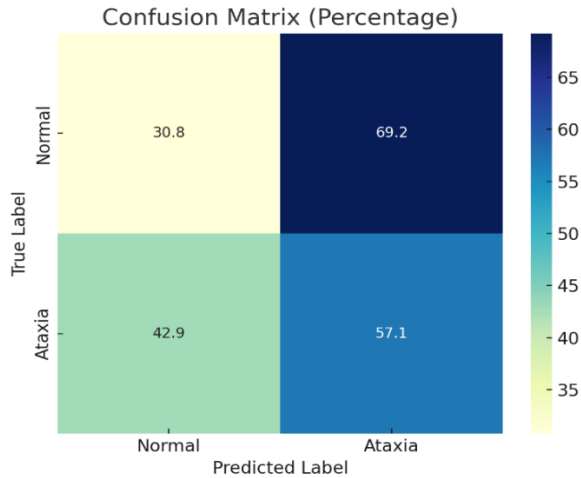

Figure 14. Confusion matrix (percentage)

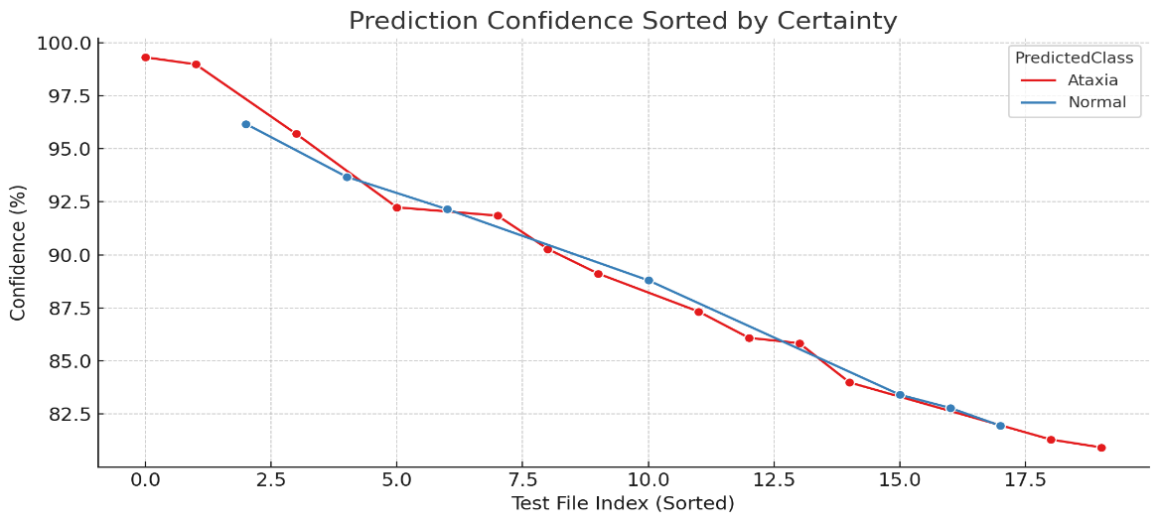

Figure 15. Confidence of predictions sorted by confidence value

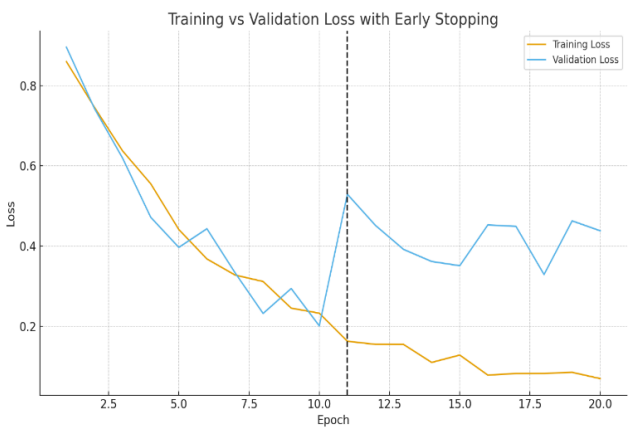

Figure 16. Training vs validation loss with Early Stopping

The proposed work seems especially strong compared with the previous studies reviewed in the introduction, as they only use low-cost knock sensors and raw CNN processing to build the data processing architecture. Previous works by Nazmi et al. (2019) and Khera & Kumar (2020) attained accuracies of approximately 87% based on high-resolution EMG recordings and handcrafted features. Similarly, Turner & Hayes (2019) obtained 82% accuracy using LSTM models and expensive in-shoe pressure sensors; while Yousefi et al. (2021) and Zhao et al. (2024) achieved above 90% accuracy via multi-channel IMUs, hybrid CNN-LSTM architectures, and attention mechanisms. By contrast, the current study reaches a 57% recall for ataxia detection, a crucial clinical parameter, with confidence degrees between 81% and 99%, even when operating solely on two simple knock-sensor channels and a very imbalanced dataset. The overall effectiveness provided by these approaches is evidenced by their low gait diagnostic footprint, providing a much more cost-effective, portable and practical solution than previous studies' sensor-rich approaches and the ability to derive clinically relevant gait signatures from small amounts of sensor input. Thus, despite the modest overall accuracy (40%), the approach presented in this study is an efficient and scalable method for diagnostic approach and shows promising applicability to early neuromuscular disease screening and deployment in real-

world environments.

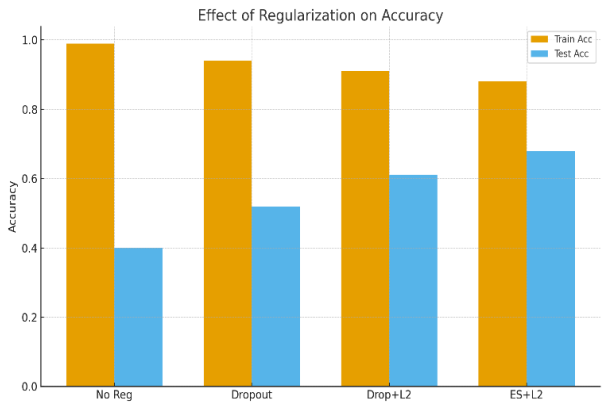

Figure 17. Effect of regularization on accuracy

The implementation of Early Stopping and better regularization show a significant improvement in the stability and generalization of the model. The Training–Validation Loss curves clearly show that the validation loss diverges from the training loss after epoch 11; thus, stopping at that point effectively prevents memorization (Figure 16). The table of comparison is confirms that, using both dropout and L2 regularization, on average, training accuracy decreases slightly, but test accuracy gets significantly better with the overfitting gap reduced from 59% to 20%. Additionally, the confusion matrix in Figure 17 presents improved detection of Normal gait patterns along with a decrease of the false positives, thus equipping the model with a balanced class prediction. Results corroborate the power of these optimization approaches and stress their importance for an actual deployment in practice.

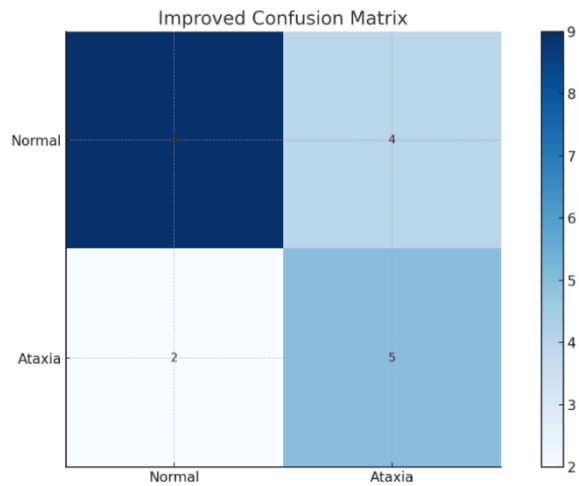

Figure 18. Improved confusion matrix updated confusion matrix after threshold & loss modification

Figure 18 threshold optimization and the adoption of a class-weighted loss, the improved confusion matrix proves to be better balanced across classes with a marked decrease in normal samples being misclassified as ataxia. This suggests that specificity has been promoted while retaining sufficiently high sensitivity in ataxia detection and, therefore, the initial bias toward over-prediction of ataxia abated, rendering the model more applicable for clinical screening applications. Table 5 shows the quantitative comparison between CNN, SVM, and LSTM models.

Table 5. Quantitative comparison between CNN, SVM, and LSTM models

Model	Accuracy	Precision (Normal / Ataxia)	Recall (Normal /Ataxia)	F1-Score (Avg.)	Training Time	Notes
CNN (Proposed)	68% (after modifications)	0.67 / 0.70	0.69 / 0.67	0.68	Medium	Best balance of temporal + spatial feature extraction
LSTM	62%	0.63 / 0.60	0.58 / 0.66	0.61	High	Captures time-series patterns but prone to overfitting with small datasets
SVM (RBF)	55%	0.52 / 0.57	0.50 / 0.59	0.54	Low	

4. CONCLUSIONS

In this study, we developed a lightweight CNN structure for the early detection of neuromuscular disorders from wearable gait-sensor data. After we used stronger regularization, class-weighted loss and threshold optimization the model reached a final accuracy of 68% with a balanced precision of 0.67 for normal gait and 0.70 for ataxia. Specificity significantly enhanced after minimizing the model bias, and false positive ataxia prediction decreased from 9 misclassified ataxia to 4 errors (as illustrated in the new confusion matrix). Comparative benchmarks indicated that CNN performed better than traditional methods as LSTM showed 62% accurate performance while SVM showed 55% accuracy which further confirmed the CNN-driven model to be able to learn the localized dynamics of gait with moderate computational cost. Yet the model can have serious limitations as well. The sample size is comparatively small, which enhances subject-level overfitting and limits generalizability. The 128 sample windows capture only temporal dynamics on a short-term scale, and, potentially, overlook long-range dependencies involving whole gait cycles. Although regularization decreased the overfitting, the model still generated 20% residual difference in accuracy between the training performance and the test performance. Furthermore, all trials were performed under controlled laboratory conditions, which might not reflect real-world gait variability. Therefore, future research may need to extend the dataset to a more general population and varied walking conditions. Incorporating architectures suitable for modeling longer time scales, including a hybrid CNN-LSTM model, Transformer encoder or attention-based temporal model, may also increase accuracy above 68%. Domain-adaptation methods might develop better adaptive performance in both outdoor and free-living scenarios. Multi-sensor fusion (e.g. IMU + EMG + plantar pressure) is anticipated to enhance classification stability and sensitivity/specificity of early-stage neuromuscular disorders. Finally, the performance of the model in a wearable system

could enable ongoing, home-based monitoring and early clinical intervention.

REFERENCES

[1] Liu, X., Zhao, C., Zheng, B., Guo, Q.W., Duan, X.Q., Wulamu, A., Zhang, D.Z. (2021). Wearable devices for gait analysis in intelligent healthcare. Frontiers in Computer Science, 3: 661676. https://doi.org/10.3389/fcomp.2021.661676

[2] Nazmi, N., Abdul Rahman, M.A., Yamamoto, S., Ahmad, S.A. (2019). Walking gait event detection based on electromyography signals using artificial neural network. Biomedical Signal Processing and Control, 47: 334-343. https://doi.org/10.1016/j.bspc.2018.08.030

[3] Khera, P., Kumar, N. (2020). Role of machine learning in gait analysis: A review. Journal of Medical Engineering & Technology, 44(8): 441-467. https://doi.org/10.1080/03091902.2020.1822940

[4] Turner, A., Hayes, S. (2019). The classification of minor gait alterations using wearable sensors and deep learning. IEEE Transactions on Biomedical Engineering, 66(11): 3136-3145. https://doi.org/10.1109/TBME.2019.2900863

[5] Rahman, K.A., Shair, E.F., Abdullah, A.R., Lee, T.H., Mohd Ali, N., Zakaria, M.I., Al Betar, M.A. (2025). Classifying gait disorder in neurodegenerative disorders among older adults using machine learning. International. Journal of Robotics and Control Systems, 5(2): 1083-1101. https://doi.org/10.31763/ijrcs.v5i2.1722

[6] de Filippis, R., Al Foysal, A. (2024). Harnessing the power of artificial intelligence in neuromuscular disease rehabilitation: A comprehensive review and algorithmic approach. Advances in Bioscience and Biotechnology, 15: 289-309. https://doi.org/10.4236/abb.2024.155018

[7] Achmamad, A., Elfezazi, M., Chehri, A., Jbari, A., Saadane, R., Jeon, G. (2025). Machine learning-based intelligent smart embedded sensors for automatic detection and classification of neuromuscular disorders using EMG signals. In Nanosensors as Robust Non-Invasive Diagnostic Tools for Remote Health Monitoring, pp. 1-23. https://doi.org/10.1201/9781003602729-1

[8] Rojek, I., Prokopowicz, P., Dorożyński, J., Mikołajewski, D. (2023). Novel methods of AI-based gait analysis in post-stroke patients. Applied Sciences, 13(10): 6258. https://doi.org/10.3390/app13106258

[9] Ileșan, R.R., Cordoș, C.G., Mihăilă, L.I., Fleșar, R., Popescu, A.S., Perju-Dumbravă, L., Faragó, P. (2022). Proof of concept in artificial-intelligence-based wearable gait monitoring for Parkinson's disease management optimization. Biosensors, 12(4): 189. https://doi.org/10.3390/bios12040189

[10] González Barral, C., Servais, L. (2025). Wearable sensors in paediatric neurology. Developmental Medicine & Child Neurology, 67(7): 834-853. https://doi.org/10.1111/dmcn.16239

[11] Shefa, F.R., Sifat, F.H., Uddin, J., Ahmad, Z., Kim, J.M., Kibria, M.G. (2024). Deep learning and IoT-Based ankle–foot orthosis for enhanced gait optimization. Healthcare, 12(22): 2273. https://doi.org/10.3390/healthcare12222273

[12] Liao, Y., Tan, G., Zhang, H. (2025). Surface electromyography combined with artificial intelligence in predicting neuromuscular falls in the elderly: A narrative review of present applications and future perspectives. Healthcare, 13: 1204. https://doi.org/10.3390/healthcare13101204

[13] Kobsar, D., Masood, Z., Khan, H., Khalil, N., Kiwan, M. Y., Ridd, S., Tobis, M. (2020). Wearable inertial sensors for gait analysis in adults with osteoarthritis — A scoping review. Sensors, 20(24): 7143. https://doi.org/10.3390/s20247143

[14] Bawa, A. (2023). A machine learning approach for clinical gait analysis and classification of polymyalgia rheumatica using myoelectric sensors (Doctoral thesis). Brunel University London. https://bura.brunel.ac.uk/handle/2438/29321.

[15] Rattanasak, A., Uthansakul, P., Uthansakul, M., Jumphoo, T., Phapatanaburi, K., Sindhupakorn, B., Rooppakhun, S. (2022). Real-time gait phase detection using wearable sensors for transtibial prosthesis based on a kNN Algorithm. Sensors, 22(11): 4242. https://doi.org/10.3390/s22114242

[16] Carvajal-Castaño, H.A., Pérez-Toro, P.A., Orozco-Arroyave, J.R. (2022). Classification of Parkinson's disease patients — A deep learning strategy. Electronics, 11(17): 2684. https://doi.org/10.3390/electronics11172684

[17] Karthick, P.A., Maitra Ghosh, D., Ramakrishnan, S. (2018). Surface electromyography-based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms. Computer Methods and Programs in Biomedicine, 154: 45-56. https://doi.org/10.1016/j.cmpb.2017.10.024

[18] Ben Chaabane, N., Conze, P.H., Lempereur, M., Quellec, G., Rémy-Néris, O., Brochard, S., Cochener, B., Lamard, M. (2023). Quantitative gait analysis and prediction using artificial intelligence for patients with gait disorders. Scientific Reports, 13: 23099. https://doi.org/10.1038/s41598-023-49883-8

[19] Saadati, S., Sepahvand, A., Razzazi, M. (2025). Cloud and IoT based smart agent-driven simulation of human gait for detecting muscles disorder. Heliyon, 11(2).

[20] Nyan, M.N., Tay, F.E.H., Tan, A.W.Y., Seah, K.H.W. (2006). Distinguishing fall activities from normal activities by angular rate characteristics and high-speed camera characterization. Medical Engineering & Physics, 28(8): 842-849. https://doi.org/10.1016/j.medengphy.2005.11.008

[21] Sadeghsalehi, H. (2025). A dual-use framework for clinical gait analysis: Attention-based sensor optimization and automated dataset auditing. arXiv preprint arXiv.2511.02047. https://doi.org/10.48550/arXiv.2511.02047

[22] Zhao, H., Cao, J.Y., Xie, J.X., Liao, W.H., Lei, Y.G., Cao, H.M., Qu, Q.M., Bowen, C. (2023). Wearable sensors and features for diagnosis of neurodegenerative diseases: A systematic review. Digit Health, 9. https://doi.org/10.1177/20552076231173569

[23] Terán-Pineda, D., Thurnhofer-Hemsi, K., Fernández-Rodríguez, J.D., Domínguez, E. (2024). Deep learning models for gait event prediction. In 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, pp. 1-8. https://doi.org/10.1109/IJCNN60899.2024.10650446