Multimodal Dataset and Evaluation Benchmarks for Pathological Gait Classification with Consumer-Grade Devices

Benjamin Lins[1,2], Adrian Ulges[1][0009-0001-1915-2464], and Biying Fu[1,*][0000-0003-4848-1256]

1 Hochschule RheinMain – University of Applied Sciences and Arts, 65195 Wiesbaden, Germany
2 University Lübeck, 23562 Lübeck, Germany
∗ Corresponding author: biying.fu@hs-rm.de

Abstract. Instrumented gait analysis (IGA) utilizes motion data to diagnose gait abnormalities, evaluate treatment efficacy, and inform rehabilitation strategies; however, its resource-, time-, and labor-intensive nature motivates the development of scalable, low-cost consumer-device solutions for remote and routine gait assessment. Thus, we aim to automatically classify six gait classes — five pathological patterns (paraparesis, hemiparesis, Duchenne, Trendelenburg, and foot drop) and normal gait — using a multimodal setup of commercial sensors. In cooperation with three physiotherapy training centers, we collected recordings of these six classes from recruited physiotherapy trainees, capturing smartphone and smartwatch inertial sensor data, acoustic step signals, and markerless pose estimates extracted from video. We evaluated both unimodal and multimodal approaches and performed late fusion via majority voting. Unimodal classification achieved accuracies up to 77 %, with pose landmark data yielding the best performance; combining multiple modalities increased accuracy to 89.4 %. We provide meaningful baseline results on the dataset, which we make public as a robust reference for future comparisons and methodological developments. Although limitations remain regarding dataset scope and transferability to clinical patient populations, the results underscore the potential of cost-effective everyday sensors for abnormal gait classification and their utility in supporting therapists and fostering future research in this domain.

Project Page & Dataset: github.com/Lins-Benjamin/MPGD

Keywords: Gait Classification · Activity Recognition · Human Sensing.

1 Introduction

Clinical instrumented gait analysis is currently performed using a combination of laboratory-based measurement systems designed to quantify human movement and detect muscle weakness [1] with high precision. The standard approach is centered on motion capture laboratories, often referred to as gait labs. While this laboratory-based approach is considered the gold standard for gait assessment [31], it is resource-intensive [12], environment-specific, and limited to short observation windows, which has driven growing interest in wearable and ambulatory gait analysis solutions.

Wearable devices, such as smartphones represent a cost-effective alternative for use in telemedicine and home-based rehabilitation programs [3]. Their adoption opens new possibilities for gait analysis outside specialized laboratory settings. By reducing cost, infrastructure requirements, and operational complexity, wearable sensors have the potential to make instrumented gait analysis more accessible and scalable, thereby enabling more frequent use in everyday clinical and real-world contexts.

A range of scientific papers have been published on this topic [21,19,27,6], which focus preliminary either on the use of specific low-cost, but body-worn sensors or on vision-based but markerless position detection. So far, good results have been achieved for the extraction of spatio-temporal parameters of gait analysis and gait classification using various machine learning approaches [26,11]. However, the combination of multiple recording modalities for gait pattern recognition has been little researched to date.

For these reasons, this work investigates the classification of six different gait patterns (paraparesis, hemiparesis, Duchenne, Trendelenburg, foot drop, and normal gait) using four different sensor modalities. The sensor modalities include non-customized and off-the-shelf sensors such as a smartphone, a smartwatch, acoustic data recording foot steps, and pose landmark data. Since there is currently no publicly available dataset that combines these recording modalities for gait analysis, we acquired a novel dataset specifically targeted for this purpose. The data collection was carried out in cooperation with three physical therapy training centers from Germany in order to record gait patterns through demonstrations by trainees. Out of all possible fine-grained gait patterns, the gait patterns discussed in this paper were selected from textbooks on physical therapy training[20], in close consultation with the participating institutions and based on how realistically they can be demonstrated by physical therapy trainees.

The collected data was then used to train deeplearning-based models that classified gait patterns. The focus was on various sequential architectures such as long-short-term memory (LSTM) [14] structure and graph-neural network architectures like ST-GCN [33] for pose data. The combination of the different modalities was investigated using a late fusion approach with majority voting. A sliding window method was used to preprocess the individual recordings, dividing them into sub-sequences.

The aim of this work is to develop, implement, and evaluate methods for multimodal analysis of six pathological gait classes based on commercial and off-the-shelf hardware, and also support future research in this field. Both unimodal classifiers for individual sensor types and fusion approaches across multiple modalities are considered. Particular attention is paid to investigating model performance, the comparability of different modalities, and the evaluation of the various possible combinations of modalities on the benchmark dataset. The main contributions of our work can be summarized in the following:

- Our primary contribution is to evaluate the task of gait analysis using widely available commercial and off-the-shelf sensors (such as smartphones and smartwatches) rather than high-precision clinical instruments.
- An additional key contribution is the compilation of a novel public available dataset containing recordings of six distinct gaits collected from trainees at three physiotherapy training centers in Germany — which can foster future research in this area.
- Finally, our work provides a proof of concept demonstrating the technical feasibility of a minimal, cost-effective multimodal gait analysis system.

2 Related Works

This section reviews existing work on pathological gait classification with sensors. While comprehensive multimodal classification approaches remain largely unexplored, prior studies have addressed related aspects, including on-body sensing, remote gait detection, and the challenge of limited data availability.

2.1 On-Body Detection of simulated Pathological Gait Patterns

Lee et al. [15] employed custom body-worn inertial measurement unit (IMU) sensors to distinguish between normal and hemiplegic gait, placing sensors at the wrist and lumbar region (L3–L4). Their approach achieved 100 % classification accuracy and was clinically validated. Ferreira et al. [8] conducted a pilot study on Trendelenburg gait using wearable IMU sensors positioned bilaterally over the anterior superior iliac spines. Data were collected from 13 participants during treadmill walking using a custom belt integrating a microcontroller and two IMUs.

Gait phase detection using IMU data has also been explored. In [26], gait events such as foot-off and foot-contact were identified using LSTM networks applied to ankle-mounted sensors. Similarly, Gao et al. [9] investigated the detection of hemiplegic, tiptoe, and cross-threshold gait patterns using a CNN–LSTM fusion model with IMUs placed on the waist and ankles in a cohort of 25 healthy participants.

Overall, existing studies predominantly focus on binary gait classification and rely on specialized hardware with fixed sensor placements. Such constraints limit scalability and real-world applicability. In contrast, our approach leverages off-the-shelf wearable devices and avoids rigid sensor configurations, enabling more practical and widely deployable gait analysis.

2.2 Remotely Detecting simulated Pathological Gait Patterns

Remote detection of simulated pathological gait patterns has predominantly relied on vision-based approaches. An early study by Nieto-Hidalgo et al. [18]

extracted gait characteristics from 2D RGB video sequences using a silhouette-based analysis framework. Rather than targeting specific pathological gait classes, their work focused mainly on distinguishing normal from abnormal gait patterns. The findings demonstrated the potential of vision-based methods to support applications in rehabilitation, sports performance analysis, and the assessment of motor development in children.

Hii et al. [11] proposed an automated gait analysis framework based on a markerless pose estimation approach using MediaPipe Pose. Their results demonstrated strong performance in extracting most temporal gait parameters with four calibrated and synchronized cameras for healthy subjects. The proposed method shows promise for assessing the effectiveness of rehabilitation or training interventions. Yamamoto et al.[32] proposed the use of a single RGB camera–based pose estimation approach, utilizing OpenPose, to replace foot-mounted IMU sensors for measuring ankle joint kinematics under various gait conditions. The study was conducted with 16 healthy young adult male participants.

Spilz et al. [28] introduced a multimodal dataset that integrates IMU measurements with marker-based motion capture data. Their setup employs ten cameras arranged and synchronized with four body-worn IMU units to record detailed movement patterns. This configuration provides high-precision data suitable for training machine-learning models for gait classification. However, the complexity and infrastructure requirements of the system limit its practicality outside of controlled laboratory environments.

2.3 Limitation of Data Availability

Existing wearable-based gait datasets primarily focus on Parkinson's disease, including the WearGait-PD dataset [5], a multi-pathology IMU dataset [29], and the GaitMotion dataset [35]. Vision-based pathological gait datasets remain comparatively scarce due to hardware complexity. Notable examples include the Kinect-based dataset by Jun et al. [13], covering six gait types (normal, antalgic, stiff-legged, lurching, steppage, and Trendelenburg), and the GAIT-IT dataset [2], which captures simulated gait abnormalities only using markerless video. More recently, Ranjan et al. [24] released a large-scale video dataset comprising 1,874 sequences of normal and pathological gait collected in unconstrained settings.

Despite these efforts, Prasanth et al. [22] noted in their systematic review of wearable sensor–based gait analysis that most studies lack validation on real pathological populations. This significant gap between methodological development and real-world validation underscores a critical limitation in the current literature. Addressing this lack of representative pathological data is essential to advance reliable and clinically meaningful gait analysis, thereby motivating the development of the present dataset.

Existing literature predominantly relies on body-worn sensors or single modality. Datasets are either restricted or include only simulated gait patterns from healthy volunteers. These factors thus limit ecological validity and sensor diversity. In contrast, our work highlights the necessity — and takes initial steps — toward richer datasets focusing on more diverse gait classes and a broader, multimodal sensor configuration, enabling a more comprehensive and cost-efficient investigation of gait analysis in realistic clinical and everyday settings.

3 Methodology

Here, we introduce our methodology by first explaining the dataset generation process and then followed by our model training.

The targeted five pathological gait classes (plus a healthy normal gait) are explained (as Def. [20]) in the following:

- **Paraparesis**: Neurologically mediated partial weakness of both legs, leading to slow or stiff gait patterns.
- **Hemiparesis**: Neurologically mediated weakness on one side of the body, causing asymmetric walking.
- **Duchenne gait**: Trunk leaning toward the stance leg, commonly due to hip abductor weakness.
- **Trendelenburg gait**: Pelvic drop on the swing side commonly caused by weak hip abductors.
- **Foot drop**: Inability to lift the forefoot during swing, often compensated by high stepping.

Healthy participants can approximate the visual appearance of these gait patterns for experimental purposes, but be careful that such simulations do not fully capture the underlying biomechanical and neurological impairments of true pathological gait.

3.1 Dataset Creation

The collected data was stored completely anonymously. Each participant was informed upfront about the exact acquisition process, data processing and data storage, and signed a consent form before the recordings began. As the study posed minimal risk and complied with applicable university ethical guidelines, formal review by an institutional ethics committee was not required. The procedures involved no physical or psychological harm and posed no risk of discrimination to participants.

Setup For the sake of reproducibility of the recorded data, the setup is explained in more detail. The setup consists of a laptop as master, two smartphones, and a smartwatch.

- A Google Pixel Watch 2 (smartwatch) is worn on the left hand wrist and records acceleration, gyroscopic, and acoustic data from the user's steps.

– A Samsung Galaxy S21 (smartphone) is placed in the right pants pocket and records acceleration, gyroscopic, and orientation data.
– A Google Pixel 8 (smartphone) is used as a stationary camera, which the laptop uses to automatically generate marker-free pose landmark data via Google Media Pipe Pose Estimation [10]. This device is placed on a remote position and not body-worn. A tripod of 1.065m height is used.

The decision to use pose data rather than RGB images was motivated not only by computational considerations but also by privacy concerns. In consultation with the physiotherapy training center, we therefore agreed to restrict visual analysis to pose and skeletal representations, which preserve essential kinematic information while minimizing personally identifiable visual features.

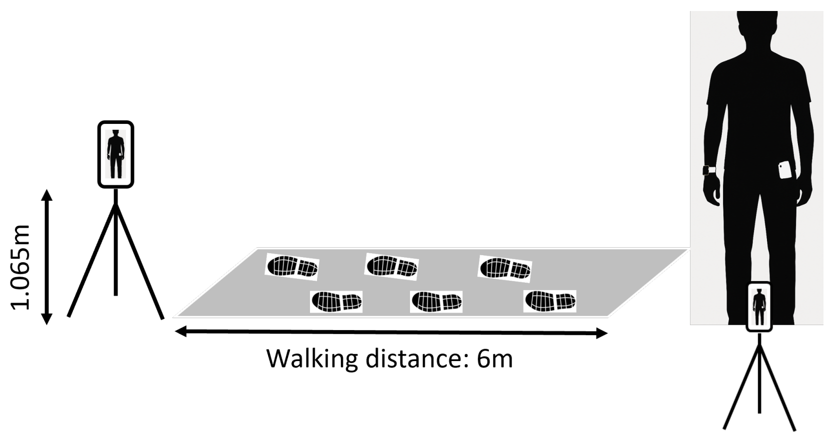

Fig. 1. illustrates our experimental setup for the data collection process. It uses a smartphone camera for pose estimation, a wrist-worn smartwatch for collecting both dynamic arm movements as well as the individual foot steps using acoustic, and another body-worn smartphone sensor to acquire leg movement during walking.

A laptop acts as master to synchronize the sensor inputs and stores the collected data. The exact structure of the setup used for data collection is shown in Figure 1 and explained in more detail below. The body-worn smartwatch on the left wrist is used to collect acoustics beyond the dynamic movement of the arm. This makes the flooring material important. It was vinyl flooring in our study due to hygiene requirements mandated in medical facilities. Participants wear sneakers, sports shoes, or sandals (no flip-flops) that do not affect the gait patterns. The body-worn smartphone sensor in the right pant's pocket aims to collect leg movement during walking further supplementing the arm movement. The smartphone on the tripod is used as a camera for video recording. The height of the tripod was chosen to ensure repeatability. This height allows the entire person to be captured in the center of the image both in the starting position and during recording. The walking distance used for the recordings was 6 m and was completed as a round trip. The distance to be covered was limited to 6 m because the accuracy of the pose estimation decreases with increasing distance. Also aligned with the 4-meter walk test, which includes a 2-meter start-up phase for gait assessment [17], thus we selected a 6-meter walking distance as a practical and suitable setup for the physiotherapy training environment. The frontal or rear orientation of the person to the camera was chosen for practical reasons, as a shot from the profile view was difficult to implement due to spatial limitations. The front and rear view, on the other hand, can be easily achieved on a 6 m long stretch, such as a long, narrow corridor, which is common in a clinical or institutional setting.

Test Subjects The test subjects were physiotherapy trainees from three different physiotherapy educational centers in Germany. This group of people was chosen mainly for technical reasons. As physiotherapy students study gait analysis in detail and learn to recognize gait patterns by demonstrating them to each other. It is ensured that the trainees have already learned the gait classes and they were instructed to perform them as natural as possible. A total of 87 individuals from three different physiotherapy training centers were recorded. The test subjects were between 18 and 25 years old, 49 of whom were male and 38 female.

Recording process The data collection process consisted of the following steps:

1. Attaching the sensors on each participant.
2. Discussing the gait pattern to be demonstrated with the test subject.
3. Assuming the starting position: the test subject stood with their back to the camera so that their entire body was centered in the image from head to toe.
4. Start of recording. The start of walking was signaled by vibration of the smartwatch.
5. Covering a distance of 6 m, then turning around and walking back the same distance to the starting point.
6. End of recording when crossing the starting point.
7. Manually entering the demonstrated gait pattern.

A complete recording process for all six gait patterns, including attaching the sensors, took approximately 10 minutes. Data collection took place over five individual days, each lasting three to eight hours. With 20 test subjects, only the normal gait pattern was recorded. With the remaining 67 test subjects, the recording process described above was carried out for all six gait patterns. Figure 2 illustrates an exemplarily pose landmark recordings of a normal walk.

Recorded Data An initial total of 430 recordings were collected. Following data cleaning and anonymization, a final total of 407 recordings were retained for further processing. Each recording comprises up to five core files, detailed below:

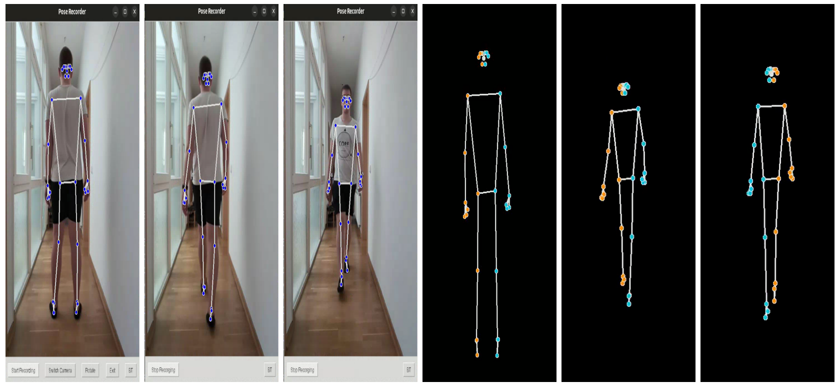

Fig. 2. A marker-free pose landmark recording of a normal walk. Skeleton is overlayed with 33 nodes and edges connecting the joints.

1. `metadata.json` stores the recording ID and the respective gait pattern class.
2. `pose.json` contains pose landmark data derived from MediaPipe, recorded at approx. 25 Hz.
3. `phone.json` contains the body-worn smartphone's acceleration, gyroscope, and orientation data, recorded at approx. 60 Hz.
4. `watch.json` contains the body-worn smartwatch's acceleration and gyroscope data, recorded at approx. 25 Hz.
5. `audio.m4a` contains acoustic data on step sounds and arm movements from the smartwatch (included where privacy constraints permit).

Optionally, a visualization video for each recording can be generated via a provided Python script. In addition, a file was created that uniquely assigns each recording to a person, enabling person-dependent data splitting for model training in the next step.

3.2 Models and Training

The aim of model training is to investigate different approaches to classifying the gait patterns recorded. To this end, the data was first preprocessed by data cleaning, data splitting, and format adjustments. This was followed by the investigation of different classification variants. The main focus was on sequence-based network architectures for timeseries processing. First, the LSTM network architecture was examined on all four sensor modalities. More specialized network structures were then tested to adopt the respective recording modalities. Subsequently, various fusion methodologies for combining the modalities into an overall classifier were examined.

Data Preparation Data preprocessing cleaned up the dataset by removing faulty recordings, dividing it into training, validation, and test data, and converting it into the format required for further use. Faulty recordings were automatically cleaned up when the files were read. This was done by checking the completeness of the recording and querying the gait pattern label. If either the completeness check or the correct label check failed, the recording was not included in the dataset. The gait pattern label check failed if the gait pattern did not match one of the six gait patterns. Of the 430 recordings, 28 faulty recordings were identified and removed through this process. The data was then divided into training, validation, and test data according to a 70 % training, 15 % validation, and 15 % test data. When dividing the data, additional care was taken to ensure that the gait patterns were evenly distributed and that each test subject had multiple recordings. The data split was carried out in such a way that all data for a person only appeared either in training (278), validation (58), or test data (66). In this way, it is guaranteed that no data leakage occurs from the training data to the test data.

Long Short Term Memory An LSTM architecture [14] was chosen to build the classification models for each single modality. The same basic model architecture was used for all modalities. Data preparation was based on a sliding window method, in which the sensor data was segmented into sequences of fixed length L with an overlap of 50% . Each of these sequences formed an input instance for the model. To produce this data format, different preprocessing was required for each modality. For the pose landmark data, the smartwatch data, and the smartphone data, the respective data were concatenated in forms of timeseries per timestamp. The acoustic data from the wrist-worn smartphone collecting footsteps were converted into the correct format using Mel spectrogram [23] transformation.

The input layer accepted sequences of length L with M features per timestamp. This was followed by one or more LSTM layers with N neurons and a dropout of 0.5. The output layer was a dense layer with softmax activation and six output neurons, corresponding to the number of targeted gait patterns here. The training process was carried out with a batch size B and the Nadam optimizer [7].

To optimize the hyperparameter selection of the models, a grid search [16] was performed for each modality. The hyperparameters selected for this purpose and their search spaces were defined as follows:

1. Sequence length/window size L of the input sequences: The starting value corresponded to a duration of one second for each modality. L was then varied in several steps up to a maximum window length of eight seconds.
2. Depth of the neural network (number of LSTM layers, each followed by a dropout layer): $[1, 2, 3]$.
3. Number of neurons N (recurrent units per layer): $[16, 32, 64, 100]$.
4. Batch size for training: $[16, 32, 64, 128]$.

The choice of hyperparameter ranges and the basic structure of the model architecture were based on the recommendations of Reimers and Gurevych [25], who conducted a study of optimal configurations for deep LSTM networks in sequence classification tasks. Parameter tuning is performed on the validation data only, without finetuning on the test data.

In addition, the relevance of the individual sensor data fields and preprocessing methods (scaling of the data) within the modalities was investigated. For the body-worn smartphone data, for example, the influence of orientation data on classification performance was evaluated. Similar partial evaluations were performed for the other modalities to determine the contribution of individual data to the overall classification result.

Furthermore, it was investigated whether a two-stage classification leads to an improvement in the model performance. The two-stage classification was conducted by first performing a binary classification between the physiological and pathological gait. Subsequently, a second classification was performed to distinguish the simulated pathological gait patterns among each other.

The models performed the classification decision of individual sub-sequences for an input modality. In order to obtain an overall classification per recording, the predictions of the sequences are combined into a final classification result using majority voting. To identify the model with the highest accuracy in the majority voting process, the relationship between sequence length (L), classification accuracy on the individual sub-sequences, and accuracy on the overall evaluation of the recordings were examined.

Spatio-Temporal Graph Convolutional Networks Besides generic LSTM models, a specialized network structure that has proven itself in other studies was tested for the pose landmark modality. We selected a graph network to accommodate the inherent topology of the skeletal structure, enabling explicit modelling of joint connectivity and facilitating the integration of spatial and temporal dynamics for gait analysis.

An ST-GCN architecture [34] was used for the pose data derived from MediaPipe [10]. The pose landmarks of the subjects were represented as a graph with 33 nodes connected by edges corresponding to the joint connections. The sensor data was segmented into sequences of fixed length, normalized, and provided to the network in the form (C,T,V), where C is the channels (x, y, z), T is the temporal length, and V is the number of nodes. The ST-GCN [30] architecture consists of several ST-GCN blocks with graph and temporal convolution, batch normalization, and dropout. Edge importance weights were implemented as learnable parameters to increase performance. Hyperparameters such as window size and batch size were varied using grid search. The models were trained on training data and optimized for hyperparameters using validation data.

Intelligent Combination of modalities Different sensor data sources were integrated using a majority voting procedure based on the predictions of the best individual models for each modality. Pose landmark, smartphone, smartwatch, and audio data are each processed by the most powerful model for that modality. Sequences with specific window sizes were generated for each modality. The

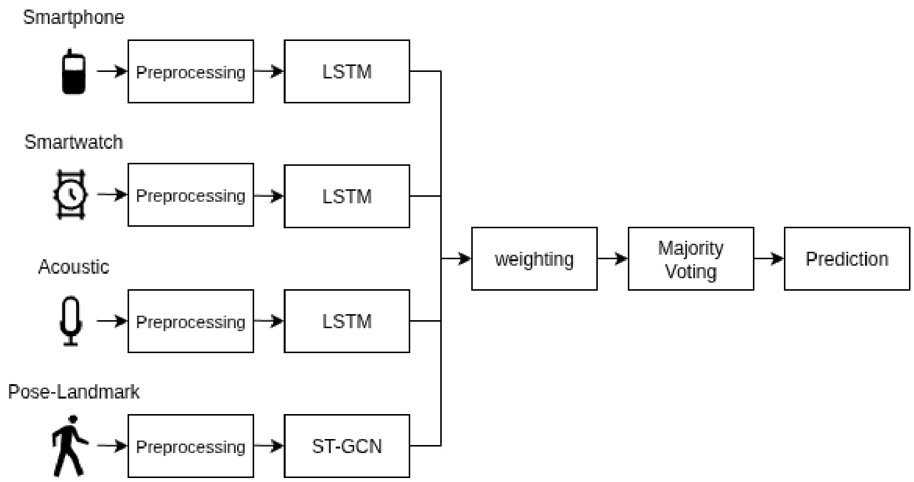

Fig. 3. shows a representation of our multimodal classification pipeline

window size differences between the modalities were then compensated for by weighting. The models provide class probabilities, which were converted into discrete predictions. The predictions were then combined and evaluated using a majority voting procedure to produce a classification.

Figure 3 shows the overall pipeline of our proposed system structure. All possible combinations of the four modalities were evaluated on the test data. This was done to check whether integrating multiple modalities increased classification accuracy. Moreover, it was determined which modalities deliver particularly good results in combination and which combination ultimately achieves the highest accuracy. For this purpose, the best individual models of each modality were used. The best combination on the test data to determine the final classification accuracy was ultimately reported in the ablation study. We also report on the percentage improvement in multimodal cases compared to lowest individual modality.

4 Evaluation

Here, the results of the models from Section 3.2 are presented and carefully analyzed. The aim is to systematically compare the performance of the individual approaches and to identify their strengths and weaknesses.

The evaluation focuses on the central question of the extent to which different data sources and model architectures are suitable for the automatic classification of simulated pathological gait patterns. For each modality, the most powerful model is highlighted based on accuracy. Additionally, confusion matrices are used to make misclassifications and confusion patterns between the classes visible. For multimodal models, it is further investigated whether consistent improvements over the unimodal baselines can be achieved by combining individual modalities.

For the valid recordings in our dataset, each gait pattern was annotated. The distribution is largely balanced across the five pathological classes, with normal gait being slightly overrepresented. An overview of the data distribution can be seen in Figure 4.

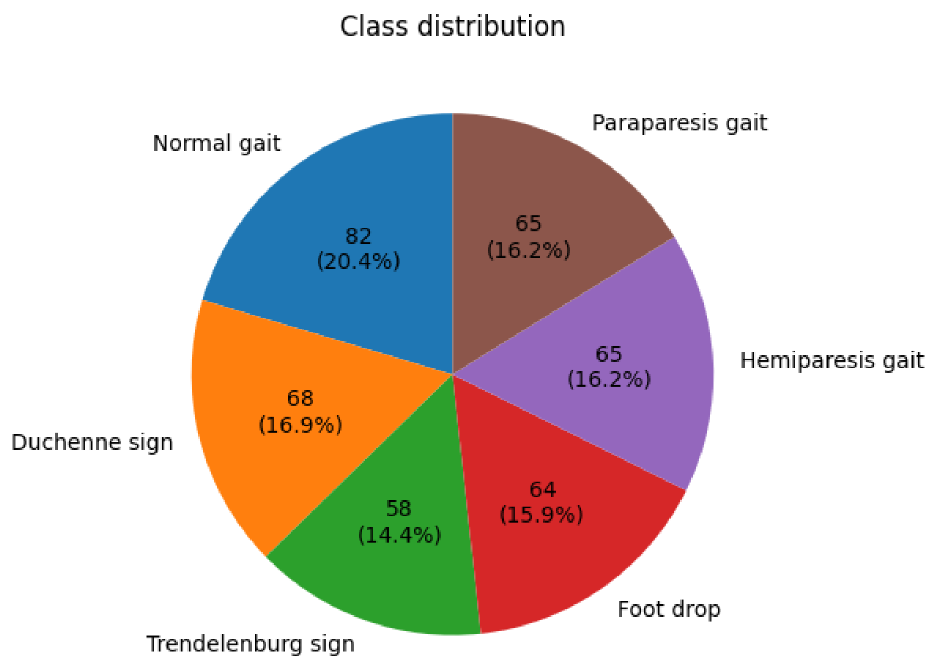

Fig. 4. Gait class distribution

4.1 Unimodal Classification

The unimodal classification examines the performance of each individual modality in detecting five simulated pathological gait patterns using our wearable setup. For each modality, various model variants are trained and compared in terms of their classification accuracy. Subsequently, the best model is analyzed in detail using a confusion matrix to make the strengths and weaknesses in the recognition of specific gait patterns visible.

Now, we first report the performance in terms of accuracy for each single sensor modality on the test set. The **smartwatch data without acoustic** achieves an accuracy of 65.2 % with an LSTM model using scaling, a window size of 2.0 s (50 samples at 25 Hz), batch size of 32, one layer, and 100 neurons per layer. Very reliably detected are the Duchenne sign (92 %) and the paraparetic gait (80 %). Average results are achieved for foot drop (73 %) and normal gait (69 %). Visibly weaker is the recognition performance on the Trendelenburg sign (44 %) and the Hemiparetic gait (30 %). Confusions occur particularly between Trendelenburg and foot drop, as well as between Hemiparetic and Paraparetic gait. Trendelenburg gait arises from hip abductor weakness. Some compensatory patterns can superficially mimic patterns seen in foot drop or other gait disorders [4]. Hemiparetic (unilateral weakness) and Paraparetic (bilateral weakness) gait share several global gait features. When classification relies on insufficient sensor setup, misclassification can occur.

The **smartphone in the right pant's pocket** achieves an accuracy of 68.2 % with an LSTM model using scaling, a window size of 1.0 s (60 samples at 60 Hz), batch size of 32, two layers, and 100 neurons per layer. Normal gait (92 %) and Paraparetic gait (81 %) are well recognized. Foot drop is identified with medium reliability (72 %). Weak recognition is seen for the Duchenne sign (41 %), the Trendelenburg sign (66 %), and Hemiparetic gait (50 %). The Duchenne sign is often confused with the Trendelenburg sign (50 %), while Hemiparetic gait is confused broadly across other pathological classes. By definition, both Duchenne and Trendelenburg gait patterns commonly originate from hip abductor weakness, which might lead to overlap in their observable characteristics. Misclassification involving hemiparetic gait again may be attributed to insufficient sensor coverage.

The **acoustic data** performs the worst. The best model is an LSTM using scaling, a window size of 2.0 s (86 samples at 43 Hz), batch size of 32, one layer, and 100 neurons per layer, achieving an accuracy of 56.1 % . Good results are achieved for normal gait (85 %), Paraparetic gait (80 %), and Hemiparetic gait (70 %). Average performance is shown in the Trendelenburg sign (56 %) and in foot drop (36 %). The Duchenne sign is recognized very poorly, with only 8 %, and in 50 % of cases, it is incorrectly classified as the Trendelenburg sign.

The **pose landmark data** from the visual feed delivers the best results. An ST-GCN architecture with edge importance, a window size of 3.0 s (75 samples at 25 Hz), and a batch size of 64 achieves an accuracy of 76.9 %. Excellent recognition is achieved for the normal gait (100 %), the Paraparetic gait (90 %), and the Hemiparetic gait (90 %). The Duchenne sign is recognized moderately well at 75 %. Weaker results are observed with the Trendelenburg sign (44 %) and foot drop (54 %), which are often confused with a normal gait, reducing the accuracy of the normal gait.

Overall, Figure 5 shows that the pose landmark data provides the highest classification accuracy with 77 %. Smartphone (68 %) and smartwatch data (65 %) achieve similar results, both about four times better than random. The acoustic data, with 56 %, come in last, but are still about three times more reliable than random guesses.

Figure 6 shows the trend of classification accuracy as a function of window length for all four investigated modalities. It demonstrates that shorter window sizes generally yield the best results, while the classification performance of all

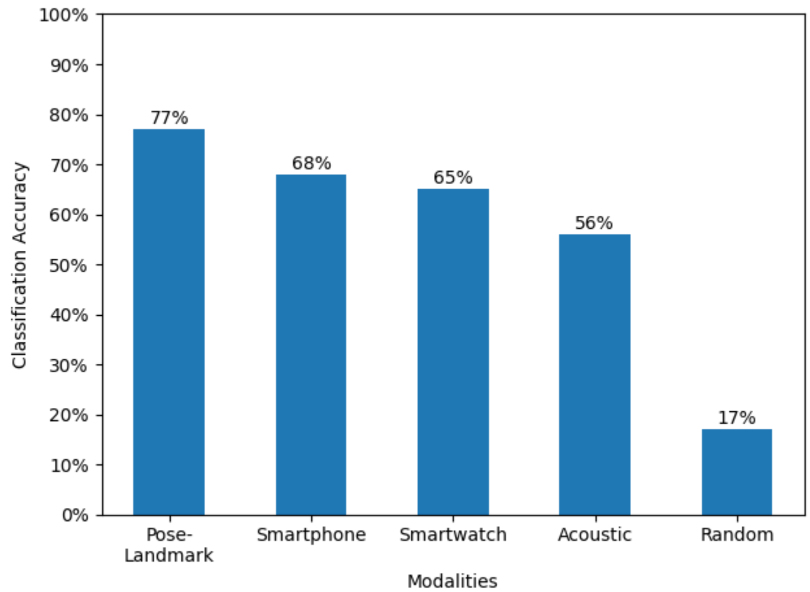

Fig. 5. shows the unimodal classification accuracy on our test dataset.

modalities continuously decreases with increasing window length. The pose landmark model achieves the highest accuracy of over 75 % with a window size of 2–3 seconds and then gradually drops below 50 % at 16 seconds. The smartwatch achieves its maximum of around 65 % with a window size of 2–4 seconds, but then drops sharply and falls below 30 % starting from 10 seconds. The body-worn smartphone data shows a similar pattern: good results with short windows (around 65 % at 2 seconds), followed by a continuous decline to under 25 %. The acoustic modality behaves similarly: it reaches around 55 % with short window size, then drops to about 21 % at 8 seconds, and further declines to 19 % at 16 seconds.

4.2 Multimodal Classification

Multimodal classification combines the best individual model of the four modalities (i.e. pose landmarks, smartphone, smartwatch, and/or acoustics data) to improve recognition performance. In the following, we report the accuracy of these multimodal setups with results listed in Table 1.

Bimodal combinations achieve accuracies between 71 % and 89 %. Test data shows particularly strong performance on data with pose landmarks + smartwatch (89 %) and pose landmarks + acoustics (85 %). Combinations without pose landmarks also improve performance, such as smartphone + smartwatch (67 % and 79 %).

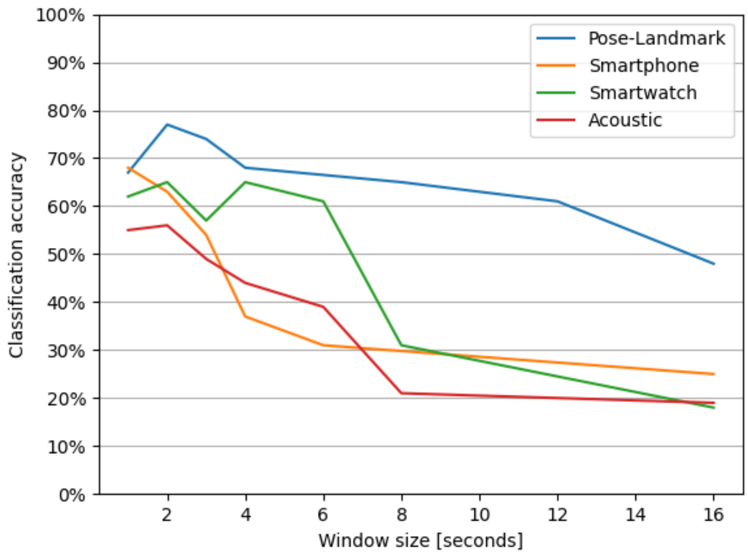

Fig. 6. visualizes the trend of classification accuracy in relation to window size. The classification accuracy of the individual modality drops with increasing window length.

Trimodal combinations achieve accuracies of 83 % to 89 %. Test data indicates that every combination is better than the best individual model. Particularly strong is the combination of Pose-Landmarks + Smartwatch + Acoustics (80 % and 89 %). The greatest relative improvement occurs with smartphone + smartwatch + acoustics (up to +12 %).

Quadrimodal combinations yield the best results. With equal weighting, an accuracy of 86.4 % (+7.6 %) on test data. By weighting according to individual model performance, the accuracy increases to 87.9 % (test data). A double weighting of individual modalities, particularly pose landmarks in combination with smartwatch or acoustics, further boosts the performance to up to 89.4 %.

As a practical recommendation for clinical settings, our findings indicate that a minimal yet informative multimodal configuration can be achieved using a bimodal setup that combines a single RGB camera for markerless pose estimation with a wrist-worn smartwatch to capture on-body dynamics. This combination provides complementary spatial and inertial information while minimizing hardware complexity, patient burden, and setup time.

Overall, the quadrimodal classification yields the most reliable results. On test data, the Figure 7 shows the confusion matrix on the individual gait classes considered in this work: normal gait, paraparetic gait, and Duchenne gait achieve partially 100 % classification performance, but weaknesses still exist with the Trendelenburg sign, which, despite multimodal fusion, shows recall values below

Pose-Landmark	Smartphone	Smartwatch	Acoustic	Accuracy	Improvment
✓	✓	✗	✗	78,8 %	0 %
✓	✗	✓	✗	**89,4** %	10,6 %
✓	✗	✗	✓	84,8 %	6 %
✗	✓	✓	✗	78,8 %	7,6 %
✗	✓	✗	✓	71,2 %	3 %
✗	✗	✓	✓	71,2 %	0 %
✓	✓	✓	✗	84,8 %	6 %
✓	✓	✗	✓	83,3 %	4,5 %
✓	✗	✓	✓	**89,4** %	10,6 %
✗	✓	✓	✓	83,3 %	12,1 %
✓	✓	✓	✓	86.4	7,6 %
2	1	1	1	87,9 %	9,1 %
2	1	2	1	**89,4** %	10,6 %
2	1	1	2	89,4 %	10,6 %
2	2	2	1	86,4 %	7,6 %
2	1	2	2	87,9 %	9,1 %
1	2	2	2	86,4 %	7,6 %

Table 1. presents the performance for the multimodal late-fusion. We start with an ablation study, where each modality is alternately enabled, and then evaluate the impact of individually weighting each modality. The best 3 performing combinations are in bold.

70 %. This could be, in mild or intermittent forms, its spatiotemporal deviations are less distinct and can fall within the variability range of normal gait. The pelvic motion can be partially masked by clothing, camera viewpoint, or limited pose-estimation accuracy at the pelvis and hip landmarks, especially when using a single RGB camera.

In this section, we first conducted an ablation study by alternately enabling and disabling individual modalities to assess their isolated contributions. Next, we examine the effect of applying different weights to each modality on the aggregated outcome, reporting performance metrics for each configuration and discussing which modalities and weightings yield the best overall results. The results clearly show that multimodal approaches compensate for the weaknesses of individual modalities and strongly enhance the performance and robustness of classification. However, these results in this section should be considered exploratory rather than a fully independent estimate of generalization performance.

5 Discussion

Our evaluation shows strong differences between the modalities: Pose landmarks with ST-GCN achieved an accuracy of 77 %, while acoustic data alone only reached 56 %. The multimodal late fusion achieved an accuracy of 89 %, clearly surpassing the unimodal approaches. Compared to a fully wearable setup by Gao

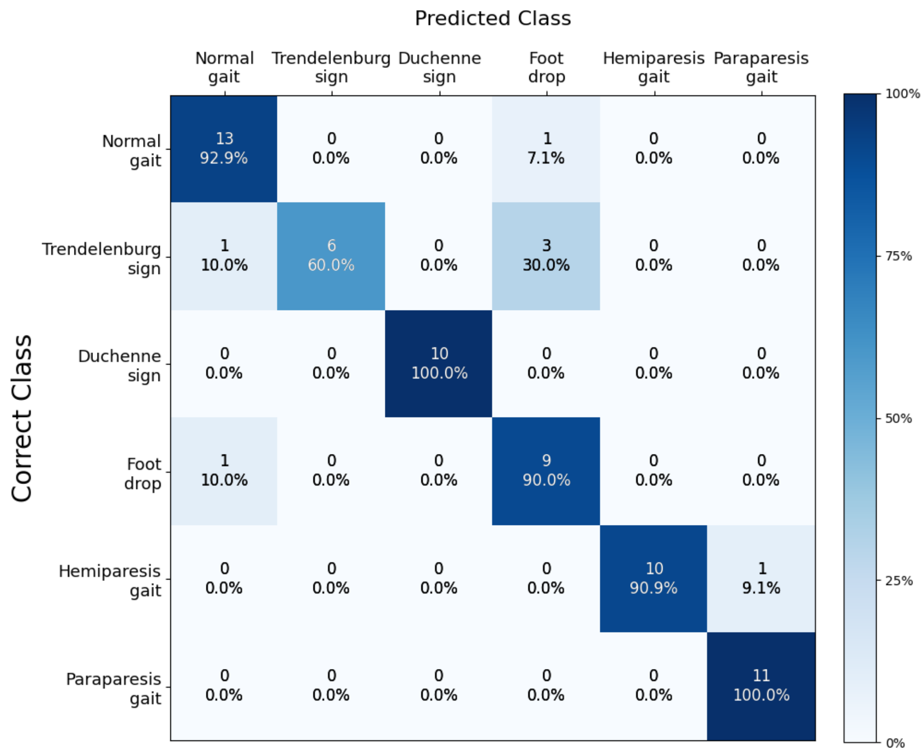

Fig. 7. Confusion matrix of the best quadrimodal combination

et al. [9], our accuracy is at a comparable level despite adding more pathological gait classes and with a more practical sensor placement (non-symmetric).

Certain gait patterns, such as the normal and Paraparetic gait, were reliably recognized, while the Trendelenburg sign posed major difficulties. This could be explained by similarities to the Duchenne sign. Modalities with weaker individual performance can achieve strong improvements in combination, provided that the underlying modalities complement each other. Therefore, a sufficient sensor coverage is the key for a good performance.

Our analysis of the window size shows that smaller windows in the sliding-window approach yield better results than processing entire recordings. This deems promising for real-time processing. For pose data, a graph-based architecture proved to be adequate. More complex architectures like ST-GCN, combining both spatial and temporal dynamics from pose data, strongly improve results compared to the LSTM only architecture.

The combination of all four modalities achieved the highest accuracy of 89 %. Partial combinations, such as those of smartphone, smartwatch, and acoustic data, achieved 83 % and also delivered robust results. Multimodal approaches proved to be more superior to unimodal ones, as they compensate for the weaknesses of individual modalities. Particularly effective were specialized architectures like ST-GCN for pose landmarks and sliding-window with smaller window size, which consistently showed better performance than processing entire recordings.

Although the recordings were obtained from trained trainees rather than clinical patients, the dataset remains a valuable proof-of-concept for pathological gait classification using a minimal, consumer-grade sensor setup. Data were collected under supervision and based on standardized demonstrations, which ensures consistent, reproducible and realistic execution of gait patterns. A follow-up interview with 28 physiotherapy trainees and specialists corroborated the study's practical relevance and judged the reduced sensor configuration as useful for training and preliminary assessment, even if it is not directly equivalent to measurements from a professional laboratory.

The recorded gaits exhibit clinically relevant characteristics but may be partially exaggerated due to conscious enactment and controlled study conditions, which may limit their ability to capture the variability of real-world pathological gait. The dataset was collected from a relatively homogeneous population, and since gait characteristics vary with age, sex, and physical constitution, generalization to older or underrepresented groups should be approached with caution and validated on more diverse cohorts. While off-the-shelf devices were employed, variations in sensor quality, sampling rates, placement, and signal processing across hardware platforms may affect performance, necessitating calibration, domain adaptation, or device-specific retraining for robust generalization.

6 CONCLUSIONS

We developed and evaluated multimodal methods for classifying five simulated pathological gait patterns using off-the-shelf smartphones and smartwatches. By fusing inertial (smartwatch) and pose-landmark features, we present an ecologically valid alternative to traditional clinical gait analysis. While unimodal setups suffer from limited sensor coverage, appropriately configured multimodal fusion substantially improves accuracy, underscoring the value of heterogeneous data integration.

We introduce a novel dataset collected with three physiotherapy training centers, comprising recordings of six gait types performed by recruited trainees. This dataset underpins our experiments and addresses a gap in publicly available pathological gait data acquired with consumer-grade devices, while its realistic yet controlled acquisition enhances practical relevance and reproducibility.

Although clinical transferability requires further validation, the dataset supports reproducible method development, enables systematic evaluation of multimodal fusion strategies with consumer-grade devices, and provides a practical resource for therapists and researchers exploring scalable gait-assessment workflows. Our results demonstrate that reliable pathological gait classification for simulated gait classes is feasible with consumer-grade sensors, providing evidence of technical feasibility under controlled conditions. Future work should include patient data, more diverse cohorts, uncontrolled environments, and real-time implementations.

References

1. Al-Zahrani, K.S., Bakheit, M.O.: A historical review of gait analysis. Neurosciences Journal **13**(2), 105–108 (2008)
2. Albuquerque, P., Machado, J.P., Verlekar, T.T., Correia, P.L., Soares, L.D.: Remote gait type classification system using markerless 2d video. Diagnostics **11**(10), 1824 (2021)
3. Altilio, R., Rossetti, A., Fang, Q., Gu, X., Panella, M.: A comparison of machine learning classifiers for smartphone-based gait analysis. Medical & Biological Engineering & Computing **59**(3), 535–546 (2021)
4. Anable, N.R., Luginsland, L.A., Carlos Jr, C., Stevens Jr, W.R., Loewen, A.M., Jeans, K.A., Sucato, D.J.: Investigating pelvic drop gait abnormality in adolescent hip pathology patients. Gait & Posture **110**, 65–70 (2024)
5. Anderson, A.J., Eguren, D., Gonzalez, M.A., Khan, N., Watkinson, S., Caiola, M., Hirczy, S.S., Zabetian, C.P., Mills, K., Moukheiber, E., et al.: Weargait-pd: An open-access wearables dataset for gait in parkinson's disease and age-matched controls. medRxiv pp. 2024–09 (2024)
6. Anderson, J., et al.: Validation of openpose-based gait analysis in children. Frontiers in Digital Health **2**, 1542012 (2025). `https://doi.org/10.3389/fdgth.2025.1542012`
7. Dozat, T.: Incorporating nesterov momentum into adam (2016)
8. Ferreira, L.C.V., Rabelo, A.G., Vieira, M.F., Pereira, A.A., Andrade, A.d.O.: Gait variability and symmetry assessment with inertial sensors for quantitative discrimination of trendelenburg sign in total hip arthroplasty patients: a pilot study based on convenience sampling. Research on Biomedical Engineering **34**, 65–72 (2018)
9. Gao, J., Gu, P., Ren, Q., Zhang, J., Song, X.: Abnormal gait recognition algorithm based on lstm-cnn fusion network. IEEE Access **7**, 163180–163190 (2019). `https://doi.org/10.1109/ACCESS.2019.2950254`
10. Google AI Edge – MediaPipe Team: Pose landmarker – mediapipe solutions (vision). `https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=de` (2025), zugriff: 20. August 2025
11. Hii, C.S.T., Gan, K.B., Zainal, N., Mohamed Ibrahim, N., Azmin, S., Mat Desa, S.H., van de Warrenburg, B., You, H.W.: Automated gait analysis based on a marker-free pose estimation model. Sensors **23**(14) (2023). `https://doi.org/10.3390/s23146489`, `https://www.mdpi.com/1424-8220/23/14/6489`
12. Hulleck, A.A., Menoth Mohan, D., Abdallah, N., El Rich, M., Khalaf, K.: Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. Frontiers in Medical Technology

4, 901331 (2022). https://doi.org/10.3389/fmedt.2022.901331, https://doi.org/10.3389/fmedt.2022.901331
13. Jun, K., Lee, Y., Lee, S., Lee, D.W., Kim, M.S.: Pathological gait classification using kinect v2 and gated recurrent neural networks. Ieee Access **8**, 139881–139891 (2020)
14. Khokhlova, M., Migniot, C., Morozov, A., Sushkova, O., Dipanda, A.: Normal and pathological gait classification lstm model. Artificial intelligence in medicine **94**, 54–66 (2019)
15. Lee, J., Park, S., Shin, H.: Detection of hemiplegic walking using a wearable inertia sensing device. Sensors **18**(6), 1736 (2018)
16. Liashchynskyi, P., Liashchynskyi, P.: Grid search, random search, genetic algorithm: a big comparison for nas. arXiv preprint arXiv:1912.06059 (2019)
17. Maggio, M., Ceda, G.P., Ticinesi, A., De Vita, F., Gelmini, G., Costantino, C., Meschi, T., Kressig, R.W., Cesari, M., Fabi, M., et al.: Instrumental and non-instrumental evaluation of 4-meter walking speed in older individuals. PloS one **11**(4), e0153583 (2016)
18. Nieto-Hidalgo, M., Ferrández-Pastor, F.J., Valdivieso-Sarabia, R.J., Mora-Pascual, J., García-Chamizo, J.M.: A vision based proposal for classification of normal and abnormal gait using rgb camera. Journal of biomedical informatics **63**, 82–89 (2016)
19. Ordóñez, F.J., Roggen, D.: Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors **16**(1), 115 (2016). https://doi.org/10.3390/s16010115, https://www.mdpi.com/1424-8220/16/1/115
20. Perry, J., Burnfield, J.M.: Gait Analysis: Normal and Pathological Function. SLACK Incorporated, Thorofare, New Jersey, 2nd edn. (2010). https://doi.org/10.1201/9781003525592
21. Pienaar, S., Malekian, R.: Human activity recognition using lstm-rnn deep neural network architecture. arXiv preprint arXiv:1905.00599 (2019), https://arxiv.org/abs/1905.00599
22. Prasanth, H., Caban, M., Keller, U., Courtine, G., Ijspeert, A., Vallery, H., Von Zitzewitz, J.: Wearable sensor-based real-time gait detection: A systematic review. Sensors **21**(8), 2727 (2021)
23. Rabiner, L., Schafer, R.: Theory and applications of digital speech processing. Prentice Hall Press (2010)
24. Ranjan, R., Ahmedt-Aristizabal, D., Armin, M.A., Kim, J.: Computer vision for clinical gait analysis: A gait abnormality video dataset. IEEE Access **13**, 45321–45339 (2025)
25. Reimers, N., Gurevych, I.: Optimal hyperparameters for deep lstm-networks for sequence labeling tasks (2017), https://arxiv.org/abs/1707.06799
26. Sarshar, M., Potluri, S., Schega, L.: Gait phase estimation by using lstm in imu-based gait analysis—proof of concept. Sensors **21**, 5749 (08 2021). https://doi.org/10.3390/s21175749
27. Singh, S.P., Sharma, M.K., Lay-Ekuakille, A., Gangwar, D., Gupta, S.: Deep convlstm with self-attention for human activity decoding using wearable sensors. IEEE Sensors Journal **21**(6), 8575–8582 (Mar 2021). https://doi.org/10.1109/jsen.2020.3045135, http://dx.doi.org/10.1109/JSEN.2020.3045135
28. Spilz, A., Oppel, H., Werner, J., Stucke-Straub, K., Capanni, F., Munz, M.: Gaitex: Human motion dataset of impaired gait and rehabilitation exercises using inertial and optical sensors. arXiv preprint arXiv:2507.21069 (2025)
29. Voisard, C., Barrois, R., l'Escalopier, N.d., Vayatis, N., Vidal, P.P., Yelnik, A., Ricard, D., Oudre, L.: A dataset of clinical gait signals with wearable sensors from healthy, neurological, and orthopedic cohorts. Scientific Data **12**(1), 1674 (2025)

30. Wang, Q., Zhang, K., Asghar, M.A.: Skeleton-based st-gcn for human action recognition with extended skeleton graph and partitioning strategy. IEEE Access **10**, 41403–41410 (2022)
31. Wren, T.A., Tucker, C.A., Rethlefsen, S.A., Gorton III, G.E., Õunpuu, S.: Clinical efficacy of instrumented gait analysis: Systematic review 2020 update. Gait & posture **80**, 274–279 (2020)
32. Yamamoto, M., Shimatani, K., Ishige, Y., Takemura, H.: Verification of gait analysis method fusing camera-based pose estimation and an imu sensor in various gait conditions. Scientific reports **12**(1), 17719 (2022)
33. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
34. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition (2018), `https://arxiv.org/abs/1801.07455`
35. Zhang, W., Zhang, H., Jiang, Z., Wang, J., Servati, A., Servati, P.: Gaitmotion: A multitask dataset for pathological gait forecasting. arXiv preprint arXiv:2405.09569 (2024)