Application of Deep neural networks and computer vision in rehabilitation robots

Capstone Report

Ibragim Gimalay

Nazarbayev University
Department of Electrical and Computer Engineering
School of Engineering and Digital Sciences

This project report was created on TexStudio editing platform using LaTeX. All the figures were drawn using draw.io online software tool.

Electrical and Computer Engineering
Nazarbayev University
http://www.nu.edu.kz

Title:
Application of Deep neural networks and computer vision in rehabilitation robots

Theme:
Robotics and Artificial Intelligence

Project Period:
Fall 2023

Project Group:
Robotics and Artificial Inteligence Lab (RAIL), Affordable Robot Assisted Gait Rehabilitation for Stroke Survivors

Participant(s):
Ibragim Gimalay

Supervisor(s):
Prashant Jamwal

Copies: 1

Page Numbers: 37

Date of Completion:
April 27, 2024

Abstract:

The objective of this research is to develop an automated system for detecting gait-related health issues using Deep Neural Networks (DNNs). The system processes video footage of patients to estimate their 3D body posture through a DNN-based method, then this 3D body posture gets classified using another DNN-based method. The analyzed 3D body pose data is classified into 3 categories: Healthy, Parkinson's disease and Post Stroke. This technology eliminates the need for bulky, complex equipment and extensive lab space, making it practical for use at home. It also doesn't require specialized knowledge for feature engineering, as it automatically extracts meaningful, high-level features from the data. The test results show classification accuracies ranging from 56% to 96% across different groups. The conclusion of this study indicates that this system is a promising tool for automatically classifying gait disorders and could be a foundational technology for future deep learning applications in clinical gait analysis. The significance of this system is underscored by its use of digital cameras as the sole required equipment, facilitating its use in patient homes and among the elderly for regular monitoring and early detection of gait changes.

Contents

Preface

Physical therapy constitutes a crucial facet of healthcare, aiding individuals in their recovery from injuries, disabilities, and various physical impediments. Nevertheless, conventional methods of physical therapy fall short in delivering specified treatment tailored to the unique needs of each patient, all while consuming substantial time and financial resources for both patients and medical institutions. The impetus behind this project lies in the exploration of the potential advantages derived from merging deep neural networks and computer vision within rehabilitation robots dedicated to physical therapy. Our objective is to create a model capable of scrutinizing patients' movements, generating comprehensive annotated data about their gait, and subsequently tailoring treatment based on the analysis of their walking patterns during procedures. This model also harnesses data from previous patients to enhance its efficacy. Our ultimate aim is to enhance the effectiveness of physical therapy, ultimately resulting in improved patient outcomes.

I want to extend my heartfelt appreciation to Professor Prashant Jamwal, my supervisor, who maintained close contact with me . I am immensely grateful to him for his unwavering support and understanding. His provision of robust computing and laboratory resources, as well as ensuring my continued access to them played a pivotal role in the quality and timely completion of my research. Without his invaluable assistance, my research would not have reached its level of excellence.

Additionally, I would like to express my gratitude to the NU Library for generously granting students free access to a vast repository of research papers and articles. Their contribution significantly enriched my research endeavors.

Nazarbayev University, April 27, 2024

Ibragim Gimalay

<Ibragim.Gimalay@nu.edu.kz>

Chapter 1

Introduction

1.1 State-of-the-art and related works

There are a lot of works discussing modelling human walking gait pattern and skeleton-muscular structure. Mostly, authors emphasize on collecting datasets of the normal walking healthy adults to standardize it and use image recognition approach to detect deviations in walking of the patients from the normal pattern.

In article [1], author uses smartphones to collect data using inertial sensors such as gyroscopes and accelerometers built in smartphones. This approach, unlike other traditional methods where subject is required to walk with average speed and/or walk along specified road, uses smartphone internal inertial sensors to collect data. This method is inexpensive and convenient as it is do not require special equipment to be worn and able to collect gait pattern everywhere and in any instance of the time. Author uses CNN+LSTM architecture it involves using Convolutional Neural Network (CNN) layers for feature extraction on input data combined with Long Term Short Memory (LSTM) to support sequence prediction. In the output is provides a textual description of an activity demonstrated in a sequence of images that is used for labelling gait pattern.

In [2] author uses Azure Kinect SDK integrated cameras to capture gait kinematics of the 7 subjects walking on the treadmill from the parietal perspective. In this study, a human skeleton detection and tracking software toolkit (Azure Kinect Body Tracking SDK) and Microsoft Azure Kinect SDK camera were used to capture 32 joint point angle readings and saved in the computer while the subject was being tracked, and the subject's motion was monitored in the depth picture field of view of the camera. Author used 4 types of the classifiers in this study. CNN classifier with 2D convolutional layers was used to extract data features from sequences of images, then spatial filter was applied to generate the output, which was rectified by a rectified linear unit (ReLU) and then exported into the pooling layer. From polling layer, output was put through a fully connected layer yielding the classification. SVM classifier was used to label data points based on their location in the dataspace. KNN classifier along with Bayesian optimizer predicts the category of a new data value according to the category of the K points closest to it. And finally, Bidirectional LSTM neural network process the input information both in forward and backward directions as the output of such data of the current time affected both from previous and future states of the object.

In manuscript [3] the authors assess the accuracy and efficiency of two markerless vision-based gait analysis approaches for person re-identification, with a focus on identifying and tracking persons by their gait patterns. The OpenPose algorithm and the DeepPoseKit algorithm are the two approaches that are being contrasted. The accuracy of each algorithm was assessed based on how effectively they could identify and match individuals across multiple camera angles and walking scenarios during the study, which used video footage of people walking to collect gait data. Results demonstrated that both algorithms were useful for evaluating gait patterns to re-identify people, with DeepPoseKit being a little more accurate and resilient under some circumstances. According to the study's findings, both approaches offer potential for use in surveillance, security, and other real-world settings that call for person tracking and identification.

In [4] author create system where sagittal and frontal recordings along with real time videos are analyzed using smartphone camera that utilizes OpenCV module, then being sent to Microsoft Azure REST Web Service to access the Machine Learning tools. In the cloud platform K-nearest neighbor (KNN) algorithm with Dynamic Time Warping (DTW) are being used for normal and abnormal classification of the gait pattern. System analyzes step length recording gait cycle starting from heel strike (HS) and till toe off (TO), along with cadence the individual walks. The system was tested and showed that it can differentiate normal walking pattern from abnormal. Future works shall be done to explicitly identify different abnormal walking patterns which include foot dragging and cycle period deviations.

In [5] article provides a literature survey on the major overground robotic gait rehabilitation approaches used for patients with paraplegia or hemiplegia due to spinal cord injury or stroke. Rehabilitation robotics has shown potential in aiding the rehabilitation process by increasing patient independence and reducing physical burden on therapists. The article compares the effectiveness of overground gait training with body weight supported treadmill training, recognizing the superiority of overground gait training due to its similarity to natural gait biomechanics. The survey provides an in-depth comparative study of the major overground robotic gait rehabilitation approaches in terms of their gait rehabilitation efficacy. The article suggests that overground gait training with robotic assistance can significantly increase the independence of patients during rehabilitation without compromising safety.

In [6] article focuses on developing robotic companion for human assistance and rehabilitation using 3D computer vision system. In article, author build cost efficient and non-contact robotic companion meaning that no wearable sensors required, that tracks relative position of the assistive device and the user to suppress the swaying of the upper body.

The article [7] presents a novel approach to classifying age groups based on dynamic gait outcomes using machine learning approaches. The authors collected gait data from 107 participants aged between 20 and 90 years old, and used three supervised machine learning approaches to classify participants into three age groups: young-middle age (20-59 years old), older (60-79 years old), and geriatric (80+ years old). The three machine learning approaches used in the study were support vector machines (SVMs), random forests (RFs), and gradient boosting machines (GBMs). All three approaches achieved high classification accuracies, with the best performing model (SVM) achieving an accuracy of 91.4 percent. The authors also investigated the gait characteristics that were most important for predicting age group. The authors also found that gait symmetry was an important feature for SVMs and RFs, but not for GBMs. Gait symmetry is a measure of the similarity between the left and right sides of the body during gait. Overall, the study demonstrates that machine learning approaches can be used to accurately classify age groups based on dynamic gait outcomes. The study also identifies step and stride regularity as the most important gait characteristics for predicting age group.

The article [8] proposes a novel image-based gait recognition approach using a convolutional neural network (CNN). The authors introduce a new two-dimensional representation of gait dynamics called Angle Embedded Gait Dynamic Image (AE-GDI), which is invariant to rotation and translation. The CNN's input image is encoded straightforwardly from the inertial sensor data sequences. The authors evaluate the proposed approach in gait authentication and gait labeling using two public datasets. Experimental results show that the proposed approach achieves competitive recognition accuracy over existing approaches and provides an effective parametric solution for identification among a large number of subjects by gait patterns. The proposed approach has several advantages over existing gait recognition methods. It is invariant to rotation and translation of the wearable device, which makes it more robust to changes in the device's orientation and position on the user's body. It is able to extract high-level features from the gait data without the need for manual feature engineering. This makes it more scalable to large datasets and less sensitive to noise in the data. It is able to achieve competitive recognition accuracy on public gait datasets, demonstrating its effectiveness in real-world scenarios. However, the proposed approach also has some limitations. It requires a large amount of training data to train the CNN model. This can be a challenge for applications where labeled gait data is scarce. The CNN model can be computationally expensive to train and deploy. This may limit its use in resource-constrained devices such as smartphones and smartwatches.

The article [9] proposes a new design and control framework for a lower limb rehabilitation robot based on human motion intention recognition with multi-source sensor information. The authors use a variety of sensors, including electromyography (EMG), inertial measurement units (IMUs), and force sensors, to collect data on the user's muscle activity, joint angles, and ground reaction forces. This data is then used to train a machine learning model to predict the user's intended motion. The predicted motion is then used to control the robot to assist the user in completing their desired movement.

The proposed approach has several advantages over traditional lower limb rehabilitation robots: it does not require the user to explicitly specify their desired motion, it can adjust its assistance level in real time based on the user's needs and it can tolerate noise and errors in the sensor data. The authors evaluate the proposed approach on a group of healthy subjects and subjects with post-stroke hemiparesis. The results show that the proposed approach is able to accurately predict the user's intended motion and provide effective assistance to users in completing their desired movements.

In article [10] proposed a low-cost gait recognition system for children using a pressure-sensor array and machine learning. The system was able to classify three pathological gaits (toe-in, toe-out, and flat) and normal gait with an accuracy of 94.7 percent. The system is portable and easy to use, and it has the potential to be used for early detection and diagnosis of pathological gait disorders.

The authors [11] propose a clinically interpretable computer-vision based method for quantifying gait in Parkinson's disease (PD). The method uses a single camera to capture videos of patients walking, and then extracts gait features such as stride length, cadence, and gait variability. These features are then used to train a machine learning model to predict the patient's PD severity.

The method was evaluated on a group of 50 PD patients and 50 healthy controls. The results showed that the method was able to accurately predict PD severity, with an accuracy of 94 percent. The method was also able to identify gait features that were associated with PD severity, such as reduced stride length and increased gait variability.

In the article [12] an ANN model for generating muscle activation patterns for human locomotion was proposed. The model is trained on EMG data from human subjects and can generate muscle activation patterns that are similar to those observed in human subjects. The model can also be used to generate muscle activation patterns for different walking speeds. The model has the potential to be used in a variety of applications, such as developing prosthetic devices, rehabilitation devices, and virtual reality applications.

This study [13] presents a proof-of-concept study for using video-based pose estimation to quantify gait parameters in stroke survivors during clinical assessments. The authors used the DeepLabCut framework to fine-tune a deep network to track five body keypoints (hip, knee, ankle, heel, and toe) in 82 below-waist videos of 8 patients with stroke performing overground walking during clinical assessments. The pose estimation model was then used to estimate five clinically relevant gait parameters (cadence, double support time, swing time, stance time, and walking speed).

The results showed that video-based pose estimation can be used to accurately quantify gait parameters in stroke survivors. The mean absolute error for the five gait parameters was less than 0.12 seconds, and the Pearson's correlation coefficients with the reference system were all greater than 0.93.

These findings suggest that video-based pose estimation has the potential to be used to develop a non-invasive and portable gait analysis system for stroke survivors. Such a system could be used to assess gait function in clinical settings and to track changes in gait over time.

This study [14] proposes a new, computer vision-based gait analysis method that is more cost-effective and less demanding for clinical populations. The method uses a deep learning-based model to track 17 points of interest (POIs) on the child's body in videos recorded using a single point-and-shoot camera. The POI data is then used to quantify various gait parameters, including gait synchrony and balance. The authors evaluated their method on a cohort of 15 children with a 16p11.2 mutation (a type of NDD) and their 12 typically developing (TD) siblings. They found that the children with 16p11.2 had significantly less whole-body gait synchrony and poorer balance compared to their TD siblings. The authors' work suggests that computer vision-based gait analysis has the potential to be a valuable tool for assessing motor function in children with NDDs. The method is cost-effective, less demanding for clinical populations, and can be used to collect data remotely. The study also introduced a new measure of whole-body gait synchrony that is more robust to handedness than previous measures. This measure could be used to enhance current motor assessment methods for children with NDDs.

Overall, this study provides promising evidence that computer vision-based gait analysis can be used to accurately and efficiently assess gait parameters in children with NDDs. The study also introduced a new measure of whole-body gait synchrony that could be used to enhance current motor assessment methods.

The article [15] provides a comprehensive review of the concepts, architectures, challenges, applications, and future directions of DL. The authors discuss the different types of DL models, the challenges associated with training and deploying DL models, and the many applications of DL in computer vision and other fields.The authors also highlight the potential of DL to revolutionize the field of healthcare. For example, DL can be used to develop new algorithms for medical image analysis, such as cancer detection and diagnosis. DL can also be used to develop new wearable devices that can track and monitor health data, such as heart rate and blood sugar levels.

Overall, the article provides a valuable resource for anyone who wants to learn more about DL. The article is well-written and informative, and it covers a wide range of topics in a comprehensive and engaging way.

Zhao and Zhou (2023) in article [16] propose a new gait recognition approach called DyGait, which exploits dynamic representations for high-performance gait recognition. DyGait uses a deep convolutional neural network (CNN) to learn dynamic representations of gait patterns from inertial measurement unit (IMU) data. The CNN is trained on a large dataset of IMU data collected from different subjects under different conditions. The authors evaluated DyGait on three public gait datasets and achieved state-of-the-art results on all datasets. DyGait outperformed other gait recognition methods, including traditional methods and deep learning-based methods. The authors also demonstrated that DyGait is robust to changes in walking speed, walking direction, and carrying conditions. This makes DyGait suitable for real-world applications, such as surveillance and security.

Overall, DyGait is a novel and effective gait recognition approach that exploits dynamic representations for high-performance gait recognition. It has the potential to be used in a variety of real-world applications.

Gait recognition is a biometric technology that identifies individuals based on their walking patterns. It is a challenging task due to the high intra- and inter-subject variability of gait patterns. Author in article [17] propose a novel gait recognition approach based on hierarchical spatio-temporal representation learning. The proposed approach uses a deep convolutional neural network (CNN) to learn hierarchical spatio-temporal representations of gait patterns from video data. The CNN is trained on a large dataset of video data collected from different subjects under different conditions. The proposed approach outperforms other gait recognition methods, including traditional methods and deep learning-based methods, on two public gait datasets. It is also robust to changes in walking speed, walking direction, and carrying conditions.

Hierarchical spatio-temporal representations of gait patterns are more informative than traditional static representations because they capture the spatial and temporal dynamics of gait. The proposed CNN architecture is able to learn hierarchical spatio-temporal representations of gait patterns at different levels of abstraction.The proposed approach has the potential to be used in a variety of real-world applications, such as surveillance and security. For example, it could be used to identify individuals from CCTV footage or to authenticate users in access control systems.

Existing gait recognition methods typically rely on 2D representations of gait patterns, such as silhouettes or skeletons. However, these representations are not robust to changes in viewpoint or occlusion. In article [18] author propose a novel gait recognition approach that uses dense 3D representations of gait patterns. The proposed approach takes advantage of recent advances in 3D human pose estimation to extract dense 3D representations of gait patterns from video data. The 3D representations are then used to train a deep convolutional neural network (CNN) for gait recognition.

The proposed approach is evaluated on a new gait recognition benchmark dataset, which is the first benchmark dataset for gait recognition in the wild. The dataset contains videos of people walking in a variety of real-world scenarios, including public streets, shopping malls, and train stations. The proposed approach outperforms other gait recognition methods, including traditional methods and deep learning-based methods, on the new benchmark dataset. It is also robust to changes in viewpoint and occlusion.

The authors of the paper [19] propose a novel deep learning method for gait recognition called MetaGait. MetaGait addresses the limitations of existing methods by learning to learn an omni-sample adaptive representation. An omni-sample adaptive representation is a representation that is robust to variations in clothing, footwear, carrying objects, and gait sequences. MetaGait achieves this by injecting meta-knowledge into the calibration network of the attention mechanism. Meta-knowledge is information about the learning process that can be used to improve the performance of a model.

MetaGait is evaluated on a number of gait recognition datasets, and it is shown to outperform existing state-of-the-art methods. For example, on the CASIA Gait Dataset, MetaGait achieves an accuracy of 95.1 percent, which is 2.3 percent higher than the previous state-of-the-art method.

The paper[20] introduces GaitContour, a robust gait recognition framework that combines pose keypoints with silhouette contour points in the novel Contour-Pose representation. This dual approach captures the intricacies of human form and movement more effectively. The system's architecture is two-pronged: firstly, it extracts local features from specific body regions using a Local Contour-Pose Transformer, then it amalgamates these local features into a comprehensive global representation via a Global Pose-Feature Transformer. This design not only ensures detailed feature extraction but also promotes the integration of these features into a holistic gait representation, significantly enhancing recognition accuracy. GaitContour's performance is rigorously validated through extensive evaluations on large-scale datasets, where it demonstrates marked superiority over existing point-based methods and robust competitiveness against advanced silhouette-based models, particularly in challenging scenarios involving distractors.

The paper [21] explores innovative gait recognition methods. It introduces the novel concept of 'skeleton maps,' which are detailed skeletal representations of human joints as heatmaps, providing a silhouette-like image. The skeleton map approach enhances the discrimination and robustness of gait recognition by focusing on structural features of the human body, excluding visual clues like clothing or carried items. The paper presents SkeletonGait and SkeletonGait++, two architectures leveraging skeleton maps. SkeletonGait++ integrates silhouette and skeleton data for improved recognition performance, demonstrating state-of-the-art results across several datasets and challenging conditions, like poor illumination or occlusions. The research showcases the potential of combining silhouette and structural features in gait recognition, setting a new benchmark in the field.

The paper [22] addresses the challenge of diagnosing musculoskeletal and cognitive impairments through gait analysis. Traditionally reliant on costly motion capture systems, the paper proposes a cost-effective, Transformer-based network to estimate critical gait parameters from single-view RGB videos. This novel approach significantly outperforms current methods in predicting parameters like Walking Speed, Gait Deviation Index (GDI), and Knee Flexion Angle at Maximum Extension, with reduced complexity and manual intervention. The model, utilizing spatial and temporal attention blocks, captures intricate gait dynamics, evident through extensive testing on a cerebral palsy patient dataset. The results demonstrate its superior performance and reduced computational load compared to the state-of-the-art models, particularly in terms of correlation and mean absolute error (MAE) for GDI and Knee Flexion Angle. However, challenges remain in accurately predicting the cadence, attributed to the model's reliance on frame-level temporal dependencies.

The paper [23] presents a novel approach for accurate gait analysis in lower-limb amputees. The study introduces a zero-shot method that leverages image generative diffusion models and existing pose estimation frameworks. This approach significantly improves the accuracy of keypoint detection on prosthetic limbs, thereby enhancing the kinematic analysis of lower-limb amputees throughout the gait cycle. The method's effectiveness is demonstrated through substantial improvements in pose estimation accuracy, particularly in challenging cases of transfemoral amputees. The study's results offer a promising advancement in rehabilitation and gait analysis for lower-limb prosthetic users, suggesting potential for more personalized and effective therapeutic strategies. Limitations, such as the method's processing speed and potential for keypoint swapping, are acknowledged, with suggestions for future research to refine and expedite the process.

The paper [24] "QAGait" focuses on enhancing gait recognition by addressing quality variances in silhouettes, offering innovative strategies for background noise removal, posture alignment, and quality-centric feature optimization. The approach, tested on multiple datasets, shows notable improvements in recognition accuracy, particularly by refining data quality and integrating quality-aware loss functions, showcasing its effectiveness in real-world, challenging environments.

Also, for research purposes, code development and implementation Awesome Gait Recognition Github repository was used[25].

1.2 Motivation

Physical therapy is an essential part of healthcare, helping individuals recover from injuries, disabilities, and other physical impairments. However, traditional physical therapy methods cannot provide personalized treatment to meet the specific needs of each patient, while being time-consuming and expensive for both patients and medical institution. The integration of deep neural networks and computer vision in rehabilitation robots has the potential to revolutionize physical therapy by providing more accurate and efficient personalized treatment for patients. The motivation behind this project is to explore the potential benefits of integrating deep neural networks and computer vision in rehabilitation robots for physical therapy. By developing a model that can analyze patient movements to provide extensive labelled status data about the patient gait and provide personalized treatment based on data collected while analyzing his/her manner of walking during procedure as well as data already collected from other patients, we hope to improve the effectiveness of physical therapy and ultimately lead to better outcomes for patients. The project will involve an analysis of the literature review and already existing methods in this area to implement them into the prototype. The results of the work can be viewed as significant implications for the future of the automated human gait pattern analysis research area.

1.3 Beneficiaries

The project will be beneficial to several groups of individuals. Such as research society, patients, healthcare industry and especially physical therapists. For research groups, this project can provide insights into the effectiveness of these technologies in improving patient outcomes and contribute to the overall advancement of medical technologies. For patients, the robots can adapt to the specific needs and progress of individual patients, providing more personalized and effective treatment. This can lead to faster recovery times and improved outcomes for patients. For medical industry, these methods can lead to cost savings in the healthcare industry. By providing more effective treatment and reducing the workload of physical therapists, the robots can lead to shorter hospital stays and reduced need for follow-up appointments. Physical therapists can also benefit, the robots can provide more accurate and objective data on patient progress, allowing therapists to make more informed decisions about the course of treatment. Additionally, the robots can assist therapists in providing more personalized treatment, reducing their workload, and improving patient outcomes.

Chapter 2

Background

2.1 Preliminary results

2.1.1 Results obtained

In order to facilitate my knowledge about the theme and deeply understand the process, the experiment from article[26] was simulated in my own environment using the same metrics, methodology and procedure.

As suggested by the authors publicly avaliable dataset was utilized which contains synchronized digital videos and three-dimensional motion capture gait data [27]. Then, those captured video sequences was analyzed in OpenPose[28] to detect keypoints in videos of healthy adults walking overground. Later, extracted gait features was processed and calculated in MATLAB to identify a variety of spatiotemporal and kinematic gait parameters from the OpenPose outputs. Subsequent post-processing in MATLAB was done according to this guidelines[29].

Gait event timings (including heel-strikes and toe-offs), along with spatial and temporal parameters of gait (such as the duration of steps, stance and swing phases, time spent in double support, step length, and walking speed), as well as sagittal joint angles of the lower extremities at the hip, knee, and ankle, were separately determined using motion capture and OpenPose's left and right perspective views.

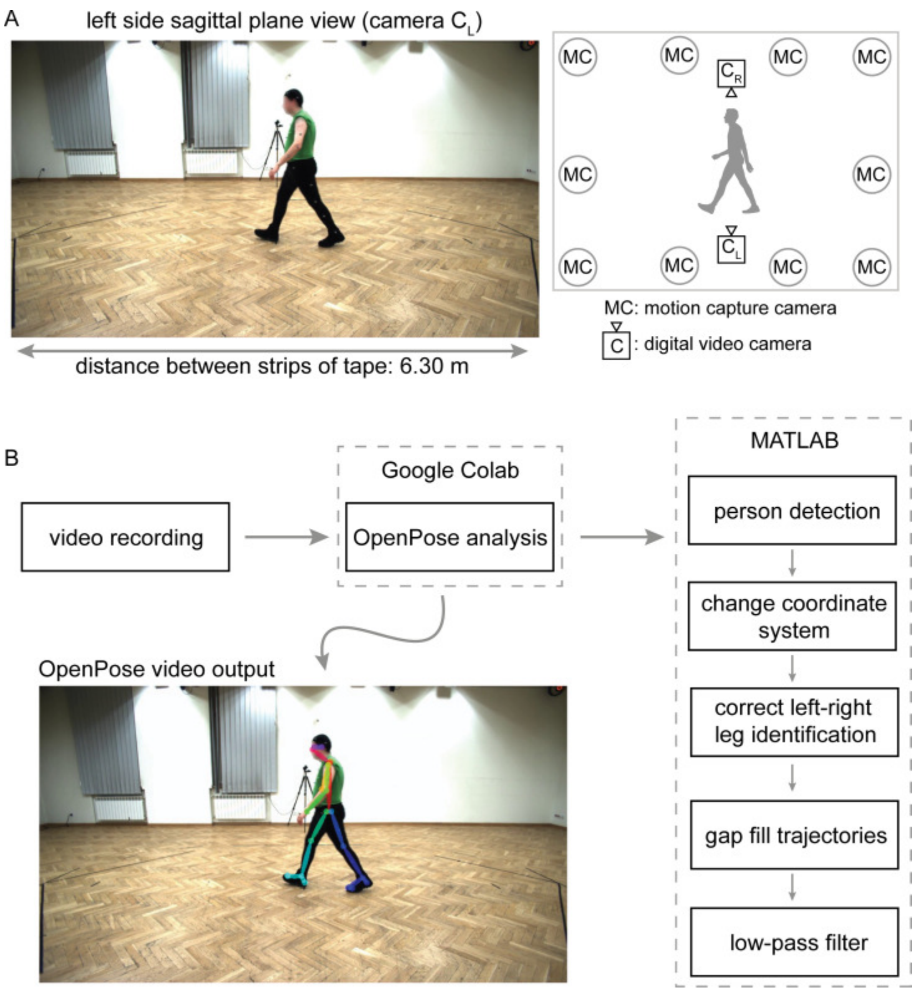

Figure 2.1: A. Representative image frame of original video recording from left-side sagittal plane view with diagram of motion capture and video cameras. B. Workflow of video recordings analyzed video with OpenPose using Google Colaboratory and next processing done using MATLAB scripts.

From the procedure performed above event timings and spatiotemporal gait parameters were retrived.

First, the effectiveness of OpenPose in identifying common gait events (namely, heel-strikes and toe-offs) was examined by comparing it to event times identified in motion capture data (refer to Figure 2.2, the leftmost and middle columns in each section). The group mean difference (indicating bias between measurement systems) in detecting heel-strikes between motion capture and OpenPose left (CL) or right (CR) side views was found to be up to one motion capture frame (10 ms, with the sampling frequency of motion capture data being 100 Hz). Meanwhile, the group mean absolute difference (representing error between measurement systems) was up to two motion capture frames (20 ms), and the largest discrepancy observed in any individual heel-strike detection was six motion capture frames (60 ms). In the case of toe-off detection, the group mean difference between motion capture and OpenPose left- or right-side views was up to two motion capture frames (20 ms), the mean absolute difference was up to three motion capture frames (30 ms), and the most significant discrepancy in any individual toe-off detection was 11 motion capture frames (110 ms).

Differences in event times for all steps.

MC: motion capture; C_L: OpenPose left-side view; C_R: OpenPose right-side view. Asterisks (*) denote P < 0.05.

	N	mean±SD			mean±SD			range		
		$MC-C_L$	$MC-C_R$	C_L-C_R	$\lvert MC-C_L\rvert$	$\lvert MC-C_R\rvert$	$\lvert C_L-C_R\rvert$	$MC-C_L$	$MC-C_R$	C_L-C_R
Left Heel-Strike Time (s)	107	−0.01±0.02	0.00±0.02	0.00±0.02	0.02±0.01	0.01±0.01	0.01±0.02	[−0.05, 0.06]	[−0.05, 0.04]	[−0.08, 0.04]
Right Heel-Strike Time (s)	109	−0.01±0.02	−0.01±0.02	0.00±0.02	0.01±0.01	0.01±0.01	0.01±0.02	[−0.04, 0.04]	[−0.05, 0.03]	[−0.04, 0.04]
Left Toe-Off Time (s)	109	−0.01±0.03	−0.01±0.02	0.00±0.04	0.03±0.02	0.02±0.02	0.02±0.03	[−0.09, 0.06]	[−0.08, 0.04]	[−0.12, 0.08]
Right Toe-Off Time (s)	107	−0.02±0.02	−0.01±0.03	0.01±0.03	0.02±0.02	0.02±0.02	0.02±0.02	[−0.11, 0.05]	[−0.07, 0.07]	[−0.04, 0.12]

Figure 2.2: Differences in event times for all steps. MC: motion capture; CL: OpenPose left-side view; CR: OpenPose right-side view. Asterisks (*) denote P < 0.05.

Temporal parameters–all steps. The group mean difference in temporal gait parameters (step time, stance time, swing time and double support time; compared for all individual steps in the walking bouts) between motion capture and OpenPose left- or right-side views was up to one motion capture frame (10 ms), the mean absolute difference in temporal gait parameters was two motion capture frames (20 ms) and the greatest difference was 10 motion capture frames (100 ms)

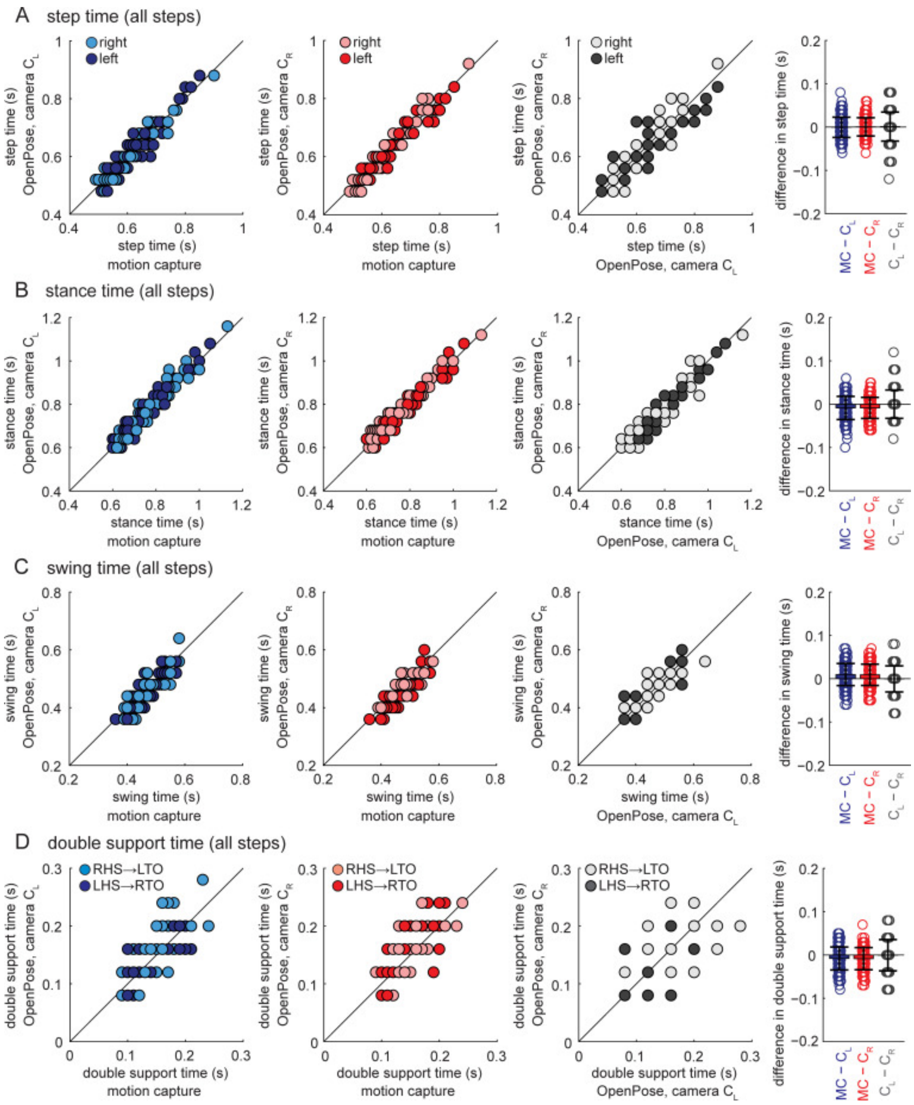

Figure 2.3: Temporal gait parameters for all individual steps shown for all participants and measurement systems. Panels show A) step time, B) stance time, C) swing time, and D) double support time. Data shown in blue represent comparisons between motion capture and OpenPose left-side (CL) views, data shown in red represent comparisons between motion capture and OpenPose right-side (CR) views, and data shown in gray represent comparisons between the two OpenPose views (CL and CR). Dark circles represent left leg data, light circles represent right leg data. Bar plots on the far right show individual data, group means, and SD to visualize the distribution of the differences observed between the measurement systems.

2.1.2 Results of the simulation of state-of-the-art approach

In the process of literature review, manuscript [30] was analyzed and reviewed. According to the author's conclusion and results among multiple pre-trained CNN based pose estimation algorithms (OpenPose[28], AlphaPose[31] and DeepLabCut[32]) to accurately reconstruct the location of joint centers against high quality ground truth data (biomechanics marker-based model).

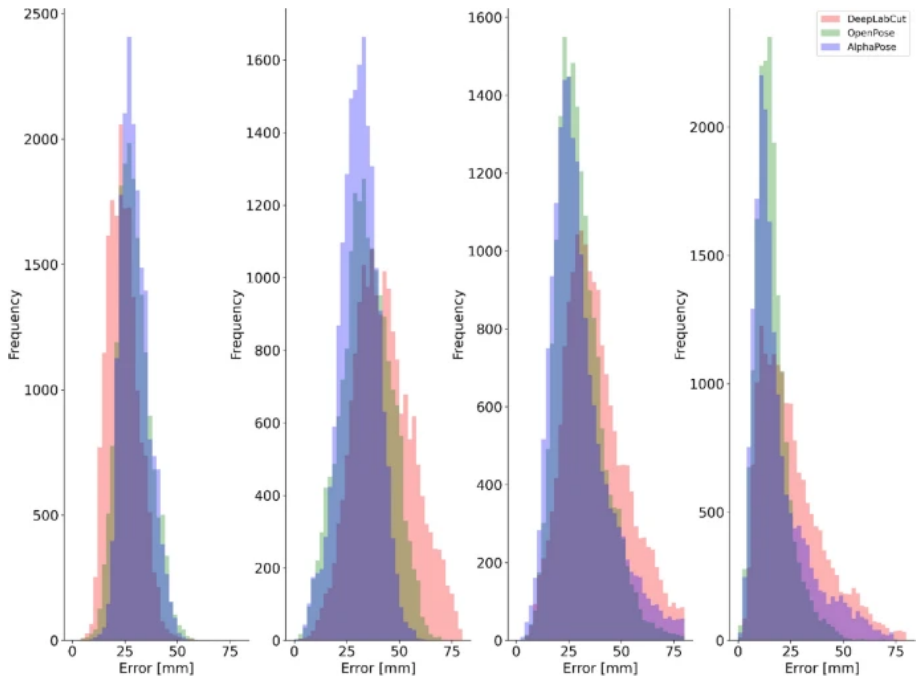

Figure 2.4: Difference distributions during walking for each pose estimation method at the right shoulder (left), hip (centre-left), knee (centre-right) and ankle (right) joint centres.

In addition to the literature review, I performed model development using CNN and LSTM models. This model can be used to assess, record, and make any necessary adjustments for a smooth gait. There are five types of silhouette images in the dataset: normal, fast-walk, slow-walk, with-bag, and with-coat. In image classification, the hybrid architecture of CNN and LSTM can be used: LSTM is used as a classifier, and CNN is used to extract complex properties from images. Five layers of Time-Distributed Convolutional (TDC) layers, along with an LSTM layer make up the developed CNN-LSTM architecture. This model gave a 87.25 percent accuracy at a 100 epochs, 6 convolution layers and 7 dropout layers. Some aspects of the code were retrived from [33].

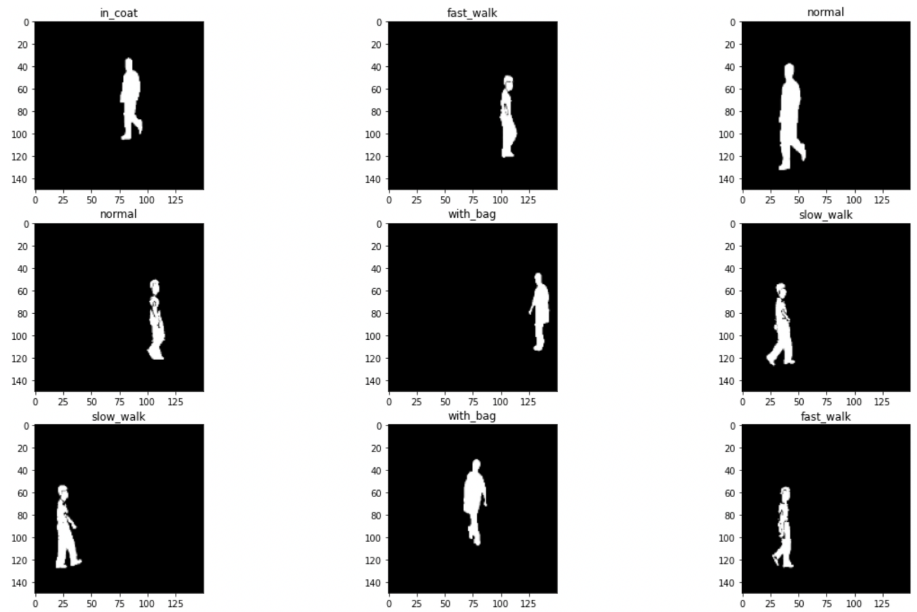

Figure 2.5: Silhouette images in the dataset: normal, fast-walk, slow-walk, with-bag, and with-coat.

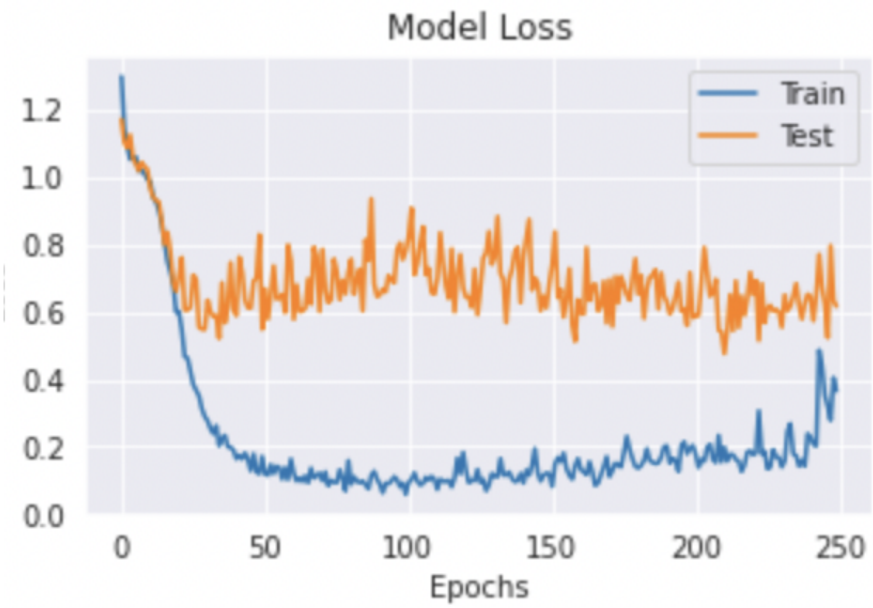

Figure 2.6: Model losses for train and test data.

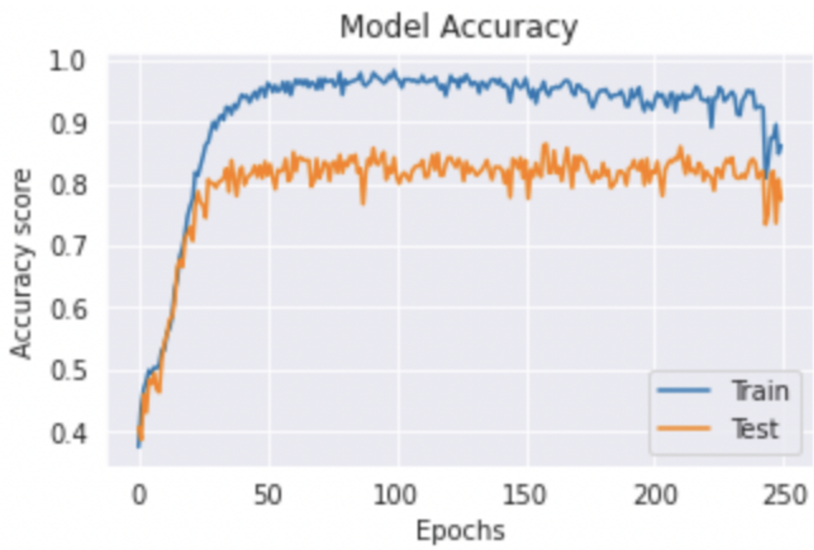

Figure 2.7: Model accuracy for train and test data.

2.2 Project implementation and plan

2.2.1 Expected project results

Capstone Project I and Capstone Project II	The expected results of this project is to develop DNN based CV system that will be analyzing patient gait pattern during physical procedure on the treadmill and provide labelled information about patient rehabilitation status. The proposed model will be validated by the simulation results tested on the prerecorded videos, comparison of the achieved data with experimental data available in the literature and real patients if accessible. Main inspiration of the system was this paper[34]

2.2.2 Project activities and work plan

Capstone Project I	**Literature Review** Conduct a comprehensive review of existing articles, papers and conferences and update the list with the progress of the project. **Dataset collection and parameter analysis** Analyze datasets for their compatibility with my system. Also identify parameters that shall be tracked during real-time and while pre capturing pattern. **Development of the model** Development of the working simplified model that can analyze walking pattern. Differentiate normal/abnormal walking patterns, identify deviations from standard pattern.. We can use MATLAB or Python. **Performance evaluation of the model** Evaluate the system performance on the volunteers.
Capstone Project II	**Further development of the model** Implement model to the rehabilitation robot and adapt it. **System Performance Analysis** Analyze system performance on the real patients. **Final Report** Discuss and evaluate achieved results, and future directions of the work.

2.3 SWOT analysis

Strengths
The strength of the project is that it can contribute to the medical industry and research community providing inexpensive and personalized solution for rehabilitation of the patients especially for those who are facing long-covid symptoms.

Weaknesses
The main limitation is implementation of the model is the equipment with tracking devices and requirement of cloud computing service in case of insufficient computational power, as not all medical institutions can fund such equipment in the

current period of time.

Opportunities
The proposed system can be changed slightly to create other applications in this field, such as human identification and flat fall prediction. Also, this project may lead to the future investigation and development CV based systems in the medical therapy.

Threats
Due to the novelty of the research, proposed emerging methods and algorithms are not open-source repositories, some algorithms and marker parameters shall be done manually.

2.4 Skills and background

During this course I expanded my background knowledge on the topic of the research project. I finished university courses required for this work such as: ELCE308 communication systems, ELCE305 data structures and algorithms, and ELCE203 signal and systems. Also, during this semester I Built several CV projects according to the guidelines found on the GitHub.

Technical skills and background that I want to achieve:

- Advanced programming skills: By the end of the project I want to be able to create CV projects only relying on NumPy instead of using TensorFlow or OpenCV.
- Research and analysis skills: Develop comprehensive view on how to critically and efficiently analyze research papers.
- Data analysis : Learn how to analyse and present data visually for the further usage.
- Academic writing: Provided good quality of the reports and presentations.

Table 2 – Risks and contingency plan

Risks	**Risk-mitigating solutions**
Ethical concerns	Informed consent forms and use of prerecorded testing videos for simulation.
Inefficient processing	Development of better algorithms for gait classification..
Limited computational power	Inquiring access to the lab machines.
Poor mage and motion capturing quality	obtaining special equipment such as Azure Kinect.

Chapter 3

Methodology

3.1 Data Collection

3.1.1 Data Acquisition for Pose Estimator Network

The neural network developed for pose estimation was trained, tested, and validated utilizing the Human3.6M dataset[35].

3.1.2 Data Acquisition for Classifier Network

Dataset utilized for training, testing and validation of the classifier network comprises of gait pattern recordings from patients with Parkinson's disease[36], post-stroke patients[37], and healthy control participants[38] 25 subjects in each group. Each subject walked on a treadmill for approximately one minute while being recorded by one or two digital cameras positioned on either side of the treadmill (sagittal plane), each with a resolution of around 480x640 pixels. Their movements were also directly captured by a synchronized motion capture system. Eight reflective markers were placed on specific body parts—neck, chest, hips, knees, and ankles—on both sides. These markers were tracked by the motion capture system, which recorded data at a sampling rate of 128 Hz.

3.2 Methods and Procedures

The objective of the system is to identify gait-related health issues using video analysis of gait. Figure 3.1 provides a schematic of the system, which incorporates two deep neural networks (DNNs). The initial DNN, Pose Estimator, processes video input to determine the 3D body pose for each video frame, generating 24 time series (corresponding to 3 directions and 8 joints). Each time series captures the position of a joint across the x, y, and z axes. The subsequent DNN, Classifier, uses these time series as input to categorize them into one of 3 predetermined groups.

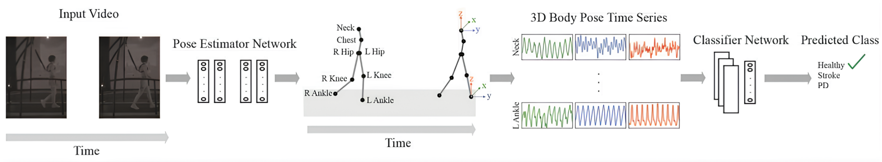

Figure 3.1: Overview of the proposed system. The input of the system is a video of the subject recorded from sagittal plane. Pose Estimator network estimates 3D body pose for each frame of the video and constructs corresponding time series. Classifier network, on the other hand, takes estimated time series as the inputand classifies it into one of the 3 pre-defined groups.

3.2.1 Pose estimator network

The 3D body pose is derived from videos using our innovative DNN-based approach. Figure 3.2 displays the network architecture. Initially, 3D body poses are estimated in the camera's coordinate system for each view independently. These poses are then transformed into global coordinates and combined across different views to enhance the accuracy of the outcomes.

In the initial phase, videos are segmented into frames, and 2D joint coordinates are determined using the Hourglass Network [39], which has achieved leading results in large-scale datasets for 2D human pose estimation. This network consists of an encoder and a decoder; the encoder utilizes convolution and pooling layers to produce low-resolution feature maps, while the decoder uses upsampling and convolution layers to create high-resolution heatmaps for each joint.

The highest probability values in these heatmaps indicate the estimated 2D joint locations. These coordinates are then processed through blocks containing fully-connected layers, ReLU activation [40], batch normalization [41] was applied to reduce internal covariate shift, dropout [42] was used to prevent overfitting, and Residual connections [43] to derive 3D joint coordinates as shown in Figure 3.2. The block architecture is adapted from Martinez et al. [44] for estimating 3D human poses from single images. We further elaborate on our method to adapt their design for multi-view analysis in the following section.

3.2.2 Multi-view fusion

The objective for multi-view fusion is to improve the accuracy of estimated 3D body pose. As mentioned earlier, the output of the Pose Estimator network is the 3D joints position in the camera coordinates. Given the location of the cameras(rotation and translation matrix), the estimated 3D jointsposition can be transferred into the global coordinates as follow:

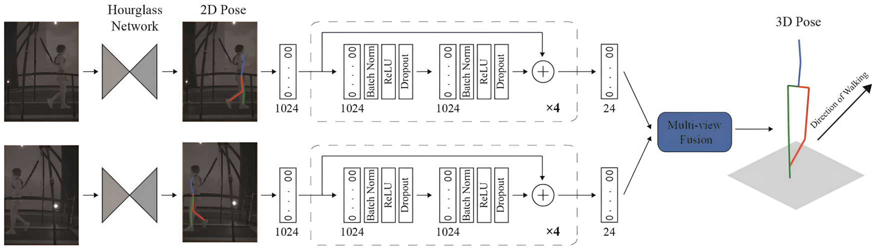

Figure 3.2: Network architecture of the "Pose Estimator" network. It starts with Hourglass Network, which estimates 2D body pose from the input image and continues by a series of blocks comprised of fully-connected layers, ReLU activation function, batch normalization, dropout, and Residual connection. The blocks are repeated four times. Numbers under each fully-connected layer illustrate the number of neurons. DNNs for each view share the same architecture and parameters, and are then fused together to estimate 3D body joint locations in the global coordinates.

$$P_i^g = R_i^{-1} P_i + T_i \tag{3.1}$$

where R_i and T_i are the rotation and translation matrices of camera i, respectively. P_i and P_i^g represent the estimated 3D body pose in camera coordinates i and global coordinates, respectively. Let $x_{i,j}, y_{i,j}, z_{i,j}$ denote the x, y, and z coordinates of joint j in view i, and $x_{i,j}^g, y_{i,j}^g, z_{i,j}^g$ denote the x, y, and z coordinates of joint j in global coordinates calculated from view i. Then P_i and P_i^g are vectors with size $3 \times J$, where J is the total number of joints (8 in this study):

$$\begin{aligned} P_i &= \left[x_{i,1}, y_{i,1}, z_{i,1}, \ldots, x_{i,j}, y_{i,j}, z_{i,J}\right] \\ P_i^g &= \left[x_{i,1}^g, y_{i,1}^g, z_{i,1}^g, \ldots, x_{i,J}^g, y_{i,j}^g, z_{i,J}^g\right] \end{aligned} \tag{3.2}$$

The ideal situation is when the estimated 3D body pose in global coordinates are exactly the same for all the views i.e., $P_1^g = P_2^g = \cdots = P_n^g$, where n is the number of cameras. However, due to the error associated with the estimated 3D joints position, it does not usually happen. The most straightforward technique to fuse the views in order to get the final 3D body pose in the global coordinates $P^g = [x_1^g, y_1^g, z_1^g, \ldots, x_J^g, y_J^g, z_J^g]$ is by taking the average of P_i^g across views. But in this work, we propose a weighted average technique, which takes into account the accuracy of the estimated 2D pose. In other words, we calculate P^g as follows:

$$\begin{aligned} \left[x_j^g, y_j^g, z_j^g\right] &= 1/n \sum_{i=1}^{n} w_{i,j} \times \left[x_{i,j}^g, y_{i,j}^g, z_{i,j}^g\right] \\ \sum_{i=1}^{n} w_{i,j} &= 1, \text{ for } j = 1, \ldots, J \end{aligned} \tag{3.3}$$

where $w_{i,j}$ is equal to the confidence probability of the estimated joint j in 2D space obtained from the heatmaps of view i. In other words, for each joint, we assign more weights to the view that estimates the 2D pose with higher confidence.

3.2.3 Classifier network

After obtaining the 3D body pose time series, the final step involves using these time series to identify health issues. Instead of engaging in extensive data preprocessing and feature engineering, we input the unprocessed time series directly into the Classifier network. This allows the network to autonomously learn complex feature representations. The design of our network, depicted in Figure 3.3, draws inspiration from the work of Wang et al. [45]. It features fully convolutional blocks that serve as feature extractors, incorporating a convolutional layer, followed by batch normalization and a ReLU activation function. The convolution process is completed by a fully connected layer and concludes with a Softmax layer that assigns the final classification label.

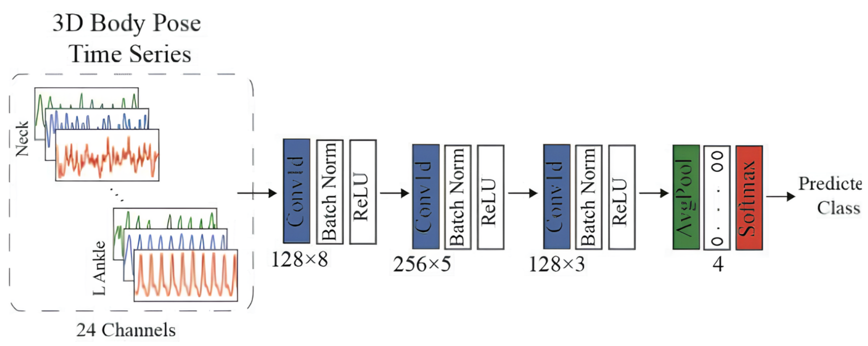

Figure 3.3: Network architecture of the "Classifier" network. It starts with a series of fully convolutional blocks comprised of 1D convolutional layer, batch normalization, and a ReLU activation function, and ends with a fully connected layer and a Softmax layer to produce final label. Numbers under each layer illustrate the corresponding size of the feature maps (number of channels × resolution) for convolutional layers and the number of neurons for fully connected layers.

Due to the varying lengths of video sequences, the associated time series differ in temporal length. Proposed DNN model, however, requires inputs of a consistent size. To overcome this, the "Window Slicing" technique was implemented. Consider a time series $TS = [ts_1, \ldots, ts_L]$ of length L. A slice is a segment of the original time series of predefined length l (where $l < L$) starting at a randomly chosen point i:

$$S_i = [ts_i, ts_{i+1}, \ldots, ts_{i+l-1}].$$

Slicing is repeated 50 times, and each time series is converted into 50 subsequences of a fixed length, which may overlap. Each subsequence is classified independently, and to determine the final label across the entire time series, the "Majority Voting" technique is applied. Another benefit of slicing is data augmentation. Through slicing, the dataset is increased by a factor of 50, which aids in avoiding overfitting and enhancing the generalization capability. The length of each slice, set to 100 frames (approximately two seconds, covering roughly one gait cycle), is subsequently down-sampled to 20 frames.

3.2.4 Implementation details

In this study, the deep learning platform employed was Pytorch, and both training and testing were conducted on a machine equipped with an M1 MAX processor and 32 GB RAM. The network training was carried out in a fully-supervised manner using an L2 loss function and Adaptive Moment Estimation (Adam) as the optimization method (b1 = 0.9, b2 = 0.999). The performance of the proposed system was evaluated using a 5-fold cross-validation method, which allocated 80% of the data for training and 20% for testing, and this process was repeated five times to assess performance across the entire dataset.

A pre-trained Hourglass model was fine-tuned on dataset using a learning rate of 0.00025 and a mini-batch size of 6 over 20,000 iterations. Subsequently, the pose estimator network was trained from scratch.

A two-stage training strategy was proposed, where initially, a network with only two blocks was trained for a single view input using a starting learning rate of 0.001 and exponential decay over 200 epochs. In the second stage, the network, expanded to four blocks as shown in Figure 3.2, was further trained for multi-view input at a reduced learning rate of 0.0001 for 5 epochs. The classifier network was also trained from scratch, employing a learning rate of 0.01 for 5 epochs.

3.3 Ethical Issues

All datasets and video recording used for training, testing and validation of the models were official and publicly avaliable on the web. Participants captured on the video recordings signed approval forms that can be found on the publishers websites.

Chapter 4

Results and Discussions

4.1 Results

4.1.1 Pose estimation accuracy

The accuracy of the Pose Estimator network is assessed by comparing the obtained results with those from a marker-based motion capture system (ground-truth), in terms of 3D pose error. The 3D pose error is determined by calculating the average Euclidean distance between the estimated 3D joint coordinates and the corresponding ground-truth data across all joints. The averaged 3D pose errors across the entire dataset are reported as 35.21±10.86 mm. Table 4.1 presents the 3D pose error for each subject and group separately. On average, the Healthy group exhibits the lowest 3D pose error, while the Post Stroke group shows the highest.

Table 4.1: AVERAGE 3D POSE ERROR (MM) FOR EACH SUBJECT AND GROUP SEPARATELY, WHERE H = HEALTHY, P = PARKINSON'S DISEASE AND S = POST STROKE.

Subjects	**H**	**P**	**S**
1.0	27.700	30.500	35.700
2.0	27.200	33.200	36.300
3.0	28.900	41.900	38.100
4.0	27.700	37.300	37.800
5.0	28.300	22.400	29.900
6.0	28.600	36.300	35.400
7.0	36.500	27.800	31.400
8.0	25.800	38.000	33.800
9.0	25.300	36.300	40.400
10.0	38.300	50.700	59.000
11.0	26.400	32.600	36.600
12.0	28.200	38.300	52.900
13.0	27.400	25.300	44.900
14.0	31.900	38.100	38.900
15.0	35.400	30.800	27.700
16.0	26.100	20.800	88.600
17.0	36.400	37.100	47.600
18.0	32.500	38.000	70.000
19.0	32.600	37.300	46.700
20.0	28.400	30.400	49.700
21.0	25.100	44.800	50.500
22.0	23.200	41.100	43.700
23.0	27.400	20.000	27.700
24.0	25.000	37.800	27.500
25.0	30.000	27.800	25.300
Average	29.212	34.184	42.244

4.1.2 Classification accuracy

In this section, results are presented for the automatic detection of health problems from the estimated 3D body pose time series. The confusion matrix and recall (sensitivity) values for each class are displayed in Table 4.2. It can be observed that the classification of healthy subjects was accomplished with greater ease by the proposed system, and the recall for this class is markedly higher in comparison to the other classes.

In the recorded instances, only 5 False Positive and 1 False Negative were observed: a patients with Parkinson's disease and Stroke patients were erroneously classified as healthy, while a healthy subject was incorrectly identified as a Post

Stroke patient. The infrequency of such misclassifications underscores the high reliability of the proposed automated system for gait monitoring in a home setting for patients and the elderly. Furthermore, the system's precision in identifying the type of health issue presented an accuracy of 77.3%, with 17 misclassifications occurring out of 75 patients.

Table 4.2: Confusion matrix of the proposed classifier network from the estimated 3D body pose time series

	Classification Output			
Actual Class	**H**	**P**	**S**	**Recall**
H	24.00	0.00	1.00	0.96
P	3.00	19.00	3.00	0.76
S	2.00	8.00	15.00	0.60

4.1.3 Impact of Estimated 3D Pose on Classification Accuracy

To assess the influence of 3D pose estimation accuracy on the classification of health issues, the ground-truth 3D body pose time series were utilized in place of the estimated 3D body pose time series to feed the Classifier network. The outcomes are presented in Table 4.3. In comparison to Table 4.2, which utilizes the estimated 3D body pose series as input, False Positive cases increase, while False negative cases decreased. Additionally, there was a reduction in misclassifications among the health problem groups. Notably, when employing the ground-truth 3D body pose time series, the accuracy of identifying the type of health problem rose to 81.3% (with 14 misclassifications out of 75 patients), marking an 5% improvement compared to the results obtained using the estimated 3D body pose time series as input.

Table 4.3: Confusion matrix of the classifier network from the groundtruth 3D body pose time series

	Classification Output			
Actual Class	**H**	**P**	**S**	**Recall**
H	24.00	1.00	0.00	0.96
P	5.00	18.00	2.00	0.72
S	1.00	5.00	19.00	0.76

4.1.4 Comparison with SVM

The capability of the DNN to classify gait-related health issues was examined by comparing its classification outcomes with those achieved by an SVM, a widely employed machine learning technique in gait classification. The SVM, a feature-based classifier, delineates hyperplane boundaries to maximize the margin between different classes. The same experiment was replicated using the concatenation of estimated 3D body pose time series as input features for the SVM model. The findings, displayed in Table 4.4, reveal that both False Positive and False Negative incidents escalated. Additionally, there was a significant rise in misclassifications across health problem groups and in general, with only 47 out of 75 patients being accurately classified, yielding an accuracy rate of 62.6%. These results underscore the superior classification capabilities of the proposed DNN, attributable to the network's proficiency in autonomously learning semantic and high-level features from the input time series, without the need for additional feature engineering.

Table 4.4: Confusion matrix of the SVM classifier from the estimated 3D body pose time series

	Classification Output			
Actual Class	**H**	**P**	**S**	**Recall**
H	18.00	4.00	3.00	0.72
P	5.00	14.00	6.00	0.56
S	3.00	7.00	15.00	0.60

4.2 Discussions

In the conducted study, an automatic system was developed for the detection of gait-related health issues, intended for continuous patient monitoring within their usual living environments. Videos were utilized to estimate the user's 3D body pose through a DNN-based method. The estimated 3D body pose time series then underwent analysis, with semantic and high-level features being extracted to identify specific health problems.

The necessity for complex equipment and large laboratory space was eliminated by the proposed system, and there was no requirement for domain-specific medical knowledge in the feature engineering process. The outcomes indicated that the system can detect health issues with a high degree of confidence and safety, characterized by infrequent False Positive and False Negative instances, using two digital cameras. This underscores the system's potential for in-home gait monitoring for patients and the elderly.

Although a vast body of literature exists on binary gait classification for clinical purposes (healthy vs. not-healthy) [13] [46], the area of multi-class gait classification has not been thoroughly explored. Multi-class gait classification presents more of a challenge than binary classification because it necessitates not only the recognition of abnormal gait patterns but also the differentiation between various abnormalities. This task is complex because neurological conditions affecting gait often share similar disturbances, such as short steps, leg rigidity, and compromised posture. A handful of studies have ventured into machine learning to propose multi-class gait classification methods that discern health issues from gait patterns. However, these methods typically rely on high-end equipment, such as optical motion capture systems[8] and IMU sensors [36], which are not feasible for in-home use.

In this research, the DNN was able to achieve precise 3D body pose estimations without the need for complex setups and data processing procedures, relying solely on digital cameras. Compared to other leading 3D body pose estimation techniques, our proposed multi-view fusion technique reported accuracy that was comparable or superior on the testing datasets. Additionally, tests on the gait dataset demonstrated that the proposed method could estimate 3D body pose with high accuracy, fitting for clinical application. The average 3D body pose error ranged from 29.2 mm to 42.2 mm across different groups, with the lowest error occurring in the Healthy group. This was somewhat anticipated, as the abnormal body postures and greater intra-subject variability in patients render the network's estimation of their body pose more challenging.

4.3 Progress

The Gantt chart in Figure 4.1 illustrates the successful completion of all stages of the project, including the system performance analysis and final report phase which has been concluded as planned.

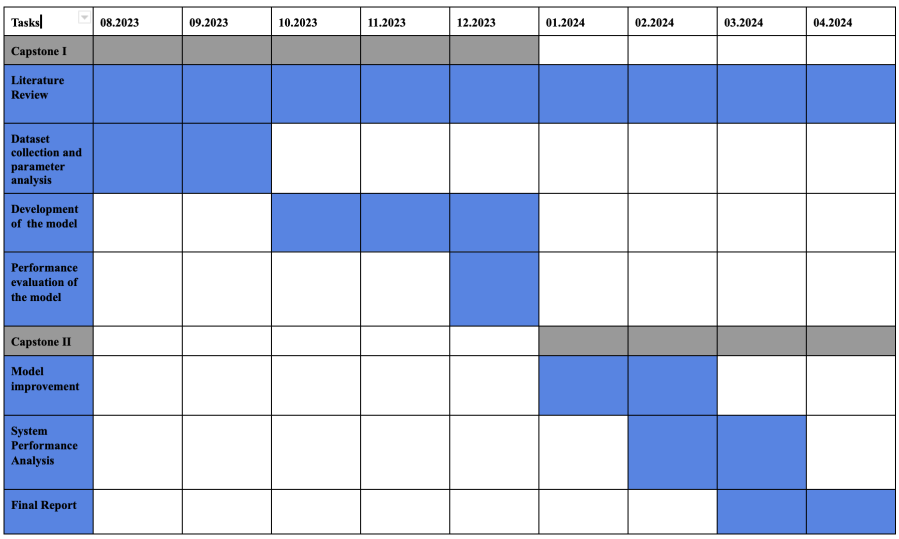

Figure 4.1: Gantt chart plan for the ongoing project.

Chapter 5

Conclusion

To sum up, this research, introduced a system for the automated detection of gait-related health issues, utilizing the capabilities of deep learning. This approach eliminates the need for advanced equipment and large lab spaces, making the system appropriate for home use. The classification accuracy achieved was 74%. Most misclassifications occurred among the pathological groups, with rare instances of a False Positive and a False Negative cases. The ultimate objective of this study is to develop a tool for ambient-assisted living that enables gait monitoring for patients and the elderly within their own homes, leveraging deep neural networks. This study serves as an initial step in this research area and provides a foundation for future deep learning applications in clinical gait analysis and pathological gait diagnosis. Future efforts will aim to broaden the study to include additional pathological groups and enhance classification accuracy by incorporating joint kinetics time series data into the classifier input.

Bibliography

[1] Qin Zou et al. "Deep Learning-Based Gait Recognition Using Smartphones in the Wild". In: *IEEE Transactions on Information Forensics and Security* 15 (2020), pp. 3197–3212. DOI: 10.1109/TIFS.2020.2985628.

[2] Biao Chen et al. "Computer Vision and Machine Learning-Based Gait Pattern Recognition for Flat Fall Prediction". In: *Sensors* 22.20 (2022). ISSN: 1424-8220. DOI: 10.3390/s22207960. URL: https://www.mdpi.com/1424-8220/22/20/7960.

[3] Jaerock Kwon, Yunju Lee, and Jehyung Lee. "Comparative study of markerless vision-based gait analyses for person re-identification". In: *Sensors* 21.24 (2021), p. 8208. DOI: 10.3390/s21248208.

[4] Mario Nieto-Hidalgo et al. "Gait analysis using computer vision based on Cloud Platform and mobile device". In: *Mobile Information Systems* 2018 (2018), 1–10. DOI: 10.1155/2018/7381264.

[5] Nor Akmal Alias et al. "The efficacy of state of the art overground Gait Rehabilitation Robotics: A bird's eye view". In: *Procedia Computer Science* 105 (2017), 365–370. DOI: 10.1016/j.procs.2017.01.235.

[6] Tao Shen et al. "A 3D computer vision-guided robotic companion for non-contact human assistance and rehabilitation". In: *Journal of Intelligent & Robotic Systems* 100.3-4 (2020), 911–923. DOI: 10.1007/s10846-020-01258-1.

[7] Yuhan Zhou et al. "The detection of age groups by dynamic gait outcomes using machine learning approaches". In: *Scientific Reports* 10.1 (2020). DOI: 10.1038/s41598-020-61423-2.

[8] Yongjia Zhao and Suiping Zhou. "Wearable device-based gait recognition using angle embedded gait dynamic images and a convolutional neural network". In: *Sensors* 17.3 (2017), p. 478. DOI: 10.3390/s17030478.

[9] Pengfei Zhang et al. "Design and control of a lower limb rehabilitation robot based on human motion intention recognition with Multi-Source Sensor Information". In: *Machines* 10.12 (2022), p. 1125. DOI: 10.3390/machines10121125.

[10] Linghui Xu et al. "Machine-learning-based children's pathological gait classification with low-cost gait-recognition system". In: *BioMedical Engineering OnLine* 20.1 (2021). DOI: 10.1186/s12938-021-00898-0.

[11] Samuel Rupprechter et al. "A clinically interpretable computer-vision based method for quantifying gait in parkinson's disease". In: *Sensors* 21 (2021), p. 5437. DOI: 10.3390/s21165437.

[12] S.D. Prentice, A.E. Patla, and D.A. Stacey. "Artificial neural network model for the generation of muscle activation patterns for human locomotion". In: *Journal of Electromyography and Kinesiology* 11.1 (2001), 19–30. DOI: 10.1016/s1050-6411(00)00038-9.

[13] Luca Lonini et al. "Video-based pose estimation for gait analysis in stroke survivors during clinical assessments: A proof-of-concept study". In: *Digital Biomarkers* 6.1 (2022), 9–18. DOI: 10.1159/000520732.

[14] Adel Ardalan et al. "Analysis of gait synchrony and balance in neurodevelopmental disorders using computer vision techniques". In: *Health Informatics Journal* 27.4 (2021). DOI: 10.1177/14604582211055650.

[15] Laith Alzubaidi et al. "Review of Deep Learning: Concepts, CNN Architectures, challenges, applications, Future Directions". In: *Journal of Big Data* 8.1 (2021). DOI: 10.1186/s40537-021-00444-8.

[16] Ming Wang et al. "DyGait: Exploiting Dynamic Representations for High-performance Gait Recognition". In: (2023). DOI: 10.48550/arXiv.2303.14953. arXiv: 2303.14953 [cs.CV].

[17] Lei Wang et al. "Hierarchical Spatio-Temporal Representation Learning for Gait Recognition". In: (2023). DOI: 10.48550/arXiv.2307.09856. arXiv: 2307.09856 [cs.CV].

[18] Jinkai Zheng et al. "Gait Recognition in the Wild with Dense 3D Representations and A Benchmark". In: (2022). DOI: 10.48550/arXiv.2204.02569. arXiv: 2204.02569 [cs.CV].

[19] Huanzhang Dou et al. "MetaGait: Learning to Learn an Omni Sample Adaptive Representation for Gait Recognition". In: (2022), pp. 357–374. DOI: doi.org/10.1007/978-3-031-20065-6_21.

[20] Yuxiang Guo et al. *GaitContour: Efficient Gait Recognition based on a Contour-Pose Representation*. 2023. arXiv: 2311.16497 [cs.CV].

[21] Chao Fan et al. *SkeletonGait: Gait Recognition Using Skeleton Maps*. 2023. arXiv: 2311.13444 [cs.CV].

[22] Quoc Hung T. Le and Hieu H. Pham. *Learning to Estimate Critical Gait Parameters from Single-View RGB Videos with Transformer-Based Attention Network*. 2023. arXiv: 2312.00398 [cs.CV].

[23] Tianxun Zhou, Muhammad Nur Shahril Iskandar, and Keng-Hwee Chiam. *Diffusion Models Enable Zero-Shot Pose Estimation for Lower-Limb Prosthetic Users*. 2023. arXiv: 2312.07854 [cs.CV].

[24] Zengbin Wang et al. *QAGait: Revisit Gait Recognition from a Quality Perspective*. 2024. arXiv: 2401.13531 [cs.CV].

[25] Yang Fu. *BNU-IVC/awesome-gait-recognition: A curated list of gait recognition and Related Area Resource*. https://github.com/BNU-IVC/Awesome-Gait-Recognition.

[26] Jan Stenum, Cristina Rossi, and Ryan T Roemmich. *Two-dimensional video-based analysis of human gait using pose estimation*. 2021. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8099131/.

[27] Bogdan Kwolek et al. "Calibrated and synchronized multi-view video and motion capture dataset for evaluation of Gait Recognition". In: *Multimedia Tools and Applications* 78.22 (2019), 32437–32465. DOI: 10.1007/s11042-019-07945-y.

[28] Zhe Cao et al. "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.1 (2021), 172–186. DOI: 10.1109/tpami.2019.2929257.

[29] Janstenum. *Janstenum/gaitanalysis-poseestimation: Gait analysis using marker-less pose estimation with openpose*. URL: https://github.com/janstenum/GaitAnalysis-PoseEstimation.

[30] Laurie Needham et al. "The accuracy of several pose estimation methods for 3D Joint Centre localisation". In: *Scientific Reports* 11.1 (2021). DOI: 10.1038/s41598-021-00212-x.

[31] Hao-Shu Fang et al. "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[32] Alexander Mathis et al. "Deeplabcut: Markerless pose estimation of user-defined body parts with deep learning". In: *Nature Neuroscience* 21.9 (2018), 1281–1289. DOI: 10.1038/s41593-018-0209-y.

[33] PaviVasudev. *PaviVasudev/GAITANALYSIS CNNLSTM: Analysis of Gait Data using a CNN LSTM model*. URL: https://github.com/PaviVasudev/GaitAnalysis_CNNLSTM/tree/main.

[34] Rahil Mehrizi et al. *Automatic Health Problem Detection from Gait Videos Using Deep Neural Networks*. 2020. arXiv: 1906.01480 [cs.CV].

[35] URL: http://vision.imar.ro/human3.6m/.

[36] Caroline Ribeiro De Souza et al. *A public data set of videos, inertial measurement unit, and clinical scales of freezing of gait in individuals with parkinson's disease during a turning-in-place task*. 2022. URL: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2022.832463/full.

[37] Rebecca L. Routson, Steven A. Kautz, and Richard R. Neptune. "Modular organization across changing task demands in healthy and poststroke gait". In: *Physiological Reports* 2.6 (2014). DOI: 10.14814/phy2.12055.

[38] URL: http://bytom.pja.edu.pl/projekty/hm-gpjatk/.

[39] Alejandro Newell, Kaiyu Yang, and Jia Deng. *Stacked Hourglass Networks for Human Pose Estimation*. 2016. arXiv: 1603.06937 [cs.CV].

[40] Vinod Nair and Geoffrey E Hinton. "Rectified linear units improve restricted boltzmann machines". In: *ICML 2010*. 2010, pp. 807–814.

[41] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG].

[42] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[43] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV].

[44] Julieta Martinez et al. *A simple yet effective baseline for 3d human pose estimation*. 2017. arXiv: 1705.03098 [cs.CV].

[45] Zhiguang Wang, Weizhong Yan, and Tim Oates. *Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline*. 2016. arXiv: 1611.06455 [cs.LG].

[46] Chang Soon Tony Hii et al. "Automated Gait Analysis Based on a Marker-Free Pose Estimation Model". In: *Sensors* 23.14 (2023). ISSN: 1424-8220. DOI: 10.3390/s23146489. URL: https://www.mdpi.com/1424-8220/23/14/6489.

Appendix A

Appendix A name

In this section, appendix information will be included.