HÖGSKOLAN
DALARNA

Dalarna Licentiate Theses 24

Gait and Movement Analysis in Neurodegenerative Disorders Using Machine Learning

MUSTAFA AL-HAMMADI

Microdata Analysis
School of Information and Engineering
Dalarna University, Borlänge, Sweden
2025

Licentiate thesis presented at Dalarna University to be publicly examined in lecture hall B101, campus Borlänge, Friday, 11 April 2025 at 09:00 for the Degree of Licentiate of Philosophy. The examination will be conducted in English. Opponent: Professor Mobyen Uddin Ahmed (Artificial Intelligence, Mälardalen University).

Abstract
Al-Hammadi, M. 2025. Gait and Movement Analysis in Neurodegenerative Disorders Using Machine Learning. *Dalarna Licentiate Theses 24*. Borlänge: Dalarna University. ISBN 978-91-88679-81-9.

Neurodegenerative disorders such as dementia and Parkinson's disease (PD) affect millions of individuals globally and are characterized by progressive cognitive decline and motor impairments. As life expectancy and the number of older people increases, the number of people with these disorders is expected to increase. Currently, neurodegenerative disorders have no cure, making early diagnosis crucial for effective management and timely intervention. Gait analysis offers a non-invasive, inexpensive, and useful method for neurodegenerative disorders detection. Gait abnormalities, particularly under dual-task (dt) conditions, are early cognitive and motor decline indicators.

This thesis aims to investigate the potential of movement analysis for the discrimination of neurodegenerative disorders compared to healthy control (HCs) persons, with a specific focus on dementia and PD. By employing machine learning techniques, the research evaluates the effectiveness of these methods in distinguishing between HCs and those with dementia or PD. This thesis utilized various traditional machine learning and deep learning models applied to the movement data. The models implemented across the studies are Support Vector Machines (SVM), Logistic Regression (LR), Random Forest (RF), Decision Trees (DT), Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and a hybrid CNN- LSTM architecture. Movement features extracted from the datasets were applied to those models.

For dementia, utilizing video-based data obtained from the Uppsala-Dalarna Dementia and Gait project (UDDGait™), the study performed pose estimation using YOLOV8, followed by feature engineering. In the current study, movement features, including velocity, acceleration, number of steps, cadence, stride length, total time, and joint angles (knee and hip) were computed and used in the machine learning algorithms to differentiate the groups. The dataset comprised 64 individuals with dementia and 67 HCs. The participants performed the Time- Up-and-Go tests (TUG) under single task and dt paradigms. Following the UDDGait study protocol, the test performance was documented with two synchronized video cameras. In the dt conditions, participants completed the TUG test while simultaneously performing a verbal/ cognitive task, which involved naming animals (TUGdt-NA) and reciting the months in reverse order (TUGdt-MB). For PD, gait features were extracted from a sensor-based dataset comprising 93 individuals with the disease and 73 HCs. The vertical ground reaction force (VGRF) was recorded for nearly two minutes using 16 sensors placed beneath each foot (8 per foot).

The results demonstrate that movement features extracted from video data, especially under dt conditions, are effective in distinguishing between HCs and those with dementia. The SVM algorithm achieved the highest accuracy of 88.5% and recall of 92.5% in dt animal naming (TUGdt-NA). For the PD study, the results demonstrate that RF obtained the highest accuracy and recall of 96%. The findings from these studies suggest that movement analysis using machine learning models offers a promising non-invasive, automated, and simple tool for the discrimination of dementia and PD compared to HCs. Future research could explore multimodal fusion approaches (i.e., speech and gait analysis) to enhance the accuracy and generalizability of these methods in clinical settings.

Keywords: Movement analysis, gait, neurodegenerative diseases, dementia, Parkinson’s disease, machine learning, deep learning

Mustafa Al-Hammadi, Microdata Analysis

ISBN 978-91-88679-81-9

urn:nbn:se:du-50292 (http://urn.kb.se/resolve?urn=urn:nbn:se:du-50292)

Acknowledgments

First and foremost, I would like to express my gratitude to my parents for their unconditional love, unwavering support, and endless care. To my dear mother, your belief in my abilities has been a constant source of strength, reminding me that no goal is out of reach. To my dear father, thank you for preparing me to face life's challenges by entrusting me with responsibilities and encouraging me to explore, experience, and learn from real-life situations. Your wisdom equipped me with the strength and understanding to navigate countless challenges. To my dear brothers and sisters, thank you for turning each achievement into a joyful celebration.

I would like to express my heartfelt gratitude to my dear supervisor, Dr. Hasan, for being an incredible mentor. Your generosity with time, unwavering support, and invaluable scientific guidance have been truly remarkable. Thanks for believing in my abilities and giving me the chance to work under your supervision. Working under his guidance has enriched my knowledge, developed my technical skills, and taught me the best practices in conducting research. His supervision has guided me through both personal and professional challenges. Beyond research, you have been supportive, generous with your time, and kind, ensuring I felt comfortable and supported. For this, I am deeply grateful.

A heartfelt thank you to my mentor, Ilias, for his support and time. Thanks for taking the time to share your knowledge and experience. I extend my gratitude to my advisors, Anna and Kjartan, for their valuable feedback and guidance.

I appreciate grants from the Swedish Research Council (2017-1259) that provided support for the collection and management of the dataset within the Uppsala-Dalarna Dementia and Gait Project (UDDGait, principal investigator AC Åberg), used for the current study.

I would also like to extend my gratitude to Vilmantas for his time and support in facilitating the data collection process.

I am deeply grateful to all the current and former Ph.D. students. Thank you for the moments of joy we shared, for the time spent together, and for the valuable experiences we've exchanged. Whether through academic discussions or collaboration, your presence has enriched my journey. The memories we've built together have made this experience truly memorable.

I am also grateful to the wonderful people and colleagues at Dalarna University for their collaboration, insightful discussions, constructive feedback during seminars, and generous knowledge sharing.

List of Papers included in this thesis

This thesis is based on the following papers, which are referred to in the text by their Roman numerals:

I. **Al-Hammadi,** M., Fleyeh, H., Åberg, A. C., Halvorsen, K., & Thomas, I. (2024). Machine Learning Approaches for Dementia Detection Through Speech and GAIT Analysis: A Systematic Literature Review. Journal of Alzheimer S Disease, 100(1), 1–27. https://doi.org/10.3233/jad-231459

II. **Al-Hammadi,** Åberg, A. C., Fleyeh, H., Halvorsen, K., & Thomas, I. Gait and Movement Analysis for Discrimination between People with Dementia and Healthy Control Persons Based on Pose Estimation and Machine Learning.

III. **Al-Hammadi,** M., Fazlali, M., & Fleyeh, H. (2024). Parkinson's Disease Classification through Gait Analysis: Comparative study of deep learning and machine learning algorithms. 2024 IEEE 19th Conference on Industrial Electronics and Applications (ICIEA), 1–5. https://doi.org/10.1109/iciea61579.2024.10665185

My contributions to the included papers are:

Paper I. Literature review, analysis, writing, and revising the manuscript.

Paper II. Research design, data processing, data analysis, writing, and revising the manuscript.

Paper III. Research design, data processing, data analysis, writing, and revising the manuscript.

Note: Since paper II is under submission, no digital version is appended to this thesis. The drafts of the paper can be requested by emailing Mustafa AL Hammadi at mum@du.se.

Acronyms

dt: dual-task

PD: Parkinson's Disease

TUG: Time-Up-and-Go test

SVM: Support vector machine

LR: Logistic regression

HCs: Healthy controls

DT: Decision trees

VGER: Vertical ground reaction force

LSTM: Long Short-Term Memory

TUGdt-NA: Time-Up-and-Go test animal naming

NDDs: Neurodegenerative diseases

CNN: Convolution Neural Networks

WHO: World Health Organisation

TUGdt-MB: Time-Up-and-Go test reciting months

RNN: Recurrent Neural Network

SCI: Subjective cognitive impairment

MCI: Mild cognitive impairment

UDDGait™: Uppsala-Dalarna Dementia and Gait project

TUGdt: Time-Up-and-Go test dual-task

Contents

1 Introduction

Neurodegenerative disorders (NDDs) are a major global health challenge, contributing significantly to disability and mortality [1]. In 2021, over three billion people globally were affected by a neurological condition [2]. NDDs such as dementia and Parkinson's disease (PD) are among the most pressing public health concerns globally. They are chronic diseases characterized by progressive nervous system degeneration, leading to impairments in cognitive and motor functions that severely impact daily life [3]. As the older population is increasing, the number of people affected by NDDs is expected to rise significantly. According to the World Health Organisation (WHO), over 55 million people live with dementia, and more than 8.5 million with PD worldwide [4,5].

Among the NDDs, dementia and PD are the most common. Dementia is an umbrella term for several conditions that result in cognitive decline that interfere with activities of daily living. It is characterized by problems with memory, thinking, language, and perception [6]. PD is a progressive neurological disorder that primarily affects motor function, causing symptoms such as tremors, stiffness, slow movement, and balance problems [7]. Early diagnosis of these diseases is crucial for improving the quality of life of the people affected through early intervention, planning, and management.

Traditional diagnostic methods for NDDs, such as dementia and PD, typically include clinical assessments such as cerebrospinal fluid analysis, electroencephalography, and brain imaging. These methods are often costly and invasive, which may cause a burden on the people undertaking the test. In contrast, gait analysis might offer an alternative approach to the detection of NDDs [8]. Gait is the pattern of walking, which includes spatial and temporal parameters such as gait speed, stride length, cadence, and number of steps.

This method may involve analyzing gait patterns, which can reveal critical changes in gait that precede the onset of more apparent cognitive decline. For instance, slow gait speed was shown to precede cognitive decline by several years [9]. Gait analysis is a non-invasive method, making it suitable for the discrimination of NDDs from cognitively healthy people.

Gait abnormalities are early indicators of NDDs and vary based on the specific condition. In dementia, early gait signs include slower walking speed, reduced stride length, and increased gait variability [10]. In PD, early gait signs include increased stride variability, reduced arm swing and asymmetry, impaired postural control, bradykinesia, and motor task difficulties, especially under dual-task (dt) conditions [11]. These motor symptoms may manifest before other cognitive decline symptoms can be detected, making mobility analysis a potentially useful tool for NDDs detection. The ability to capture and analyze these changes in mobility patterns via video data or sensor technology, and employing machine learning algorithms can potentially offer an affordable, automated, and non-invasive complementary aid to traditional diagnostic methods.

Different methods exist to assess mobility and detect motor and cognitive impairment functions. The time-up-and-go test (TUG) assesses the mobility of individuals undertaking the test. The test involves standing up from a chair with armrests, walking three meters, turning around, walking back to the chair, and sitting down. [12]. The Time-Up-and-Go test dual-task (TUGdt) involves performing simultaneously cognitive tasks (e.g., reciting months or naming animals) while walking. This method demonstrates greater sensitivity in detecting early signs of dementia or prodromal PD [13,14]. The dt approach enhances the ability to detect subtle impairments that might not be evident in simpler tests.

Advances in movement analysis using machine learning through non-invasive methods present a promising approach to discriminate NDDs, including dementia and PD. By extracting movement features and applying machine learning, it offers an automated method for identifying movement alterations that help differentiate between the conditions.

1.1 Context and Research Gaps

In recent years, various studies have introduced different methods for movement analysis. Traditionally, movement analysis depends on methods such as visual observation or timing devices. Modern techniques utilize vision-based systems, wearable sensors, and motion capture systems. These advancements allow for precise measurements of different spatio-temporal, kinetic, and kinematic parameters, which are important for detecting abnormalities in gait [15]. Gait analysis, using machine learning, enabled the identification of gait activities and events [16].

NDDs manifest gait disturbances such as shorter stride length and slower walking speed [17]. Research has explored the potential of machine learning techniques to discriminate between Amyotrophic Lateral Sclerosis, PD, and Huntington's Disease through gait analysis [18, 19, 20].

Automated gait analysis using pose estimation and machine learning from video data has been utilized in the discrimination of NDDs, such as PD, through methods such as AlphaPose and MediaPipe [21, 22]. However, its application to dementia detection remains underexplored [10].

However, in the Uppsala-Dalarna Dementia and Gait project (UDDGait™), the TUGdt test development and evaluation have been carried out within longitudinal studies, in which video-recorded TUG and TUGdt tests were investigated and complemented by clinical cognitive and motor function tests [23]. Test set-up for video-recorded TUG testing was developed and performed in a clinical environment, where all study participants completed the TUG tests in the following order: single-task TUG and two TUGdt, i.e., TUGdt-NA (name animals) and TUGdt-MB (Months back, i.e. reciting months in reverse order). To capture both mobility and verbal performance of the TUGdt tests, results have been based on the quantification of correctly mentioned words per 10s (i.e., words/time). Other parameters analysed by Åberg et al. [24] in UDDGait and published to date are: time scores, number of correct words, dual-task cost (calculated as 100*(TUGdt time score - single-task TUG time score)/single-task TUG time score), and step parameters extracted from video-recorded tests.

Results indicate an association between TUGdt words/time performance and neurodegeneration [25]. Additionally, they show that TUGdt words/time can discriminate well between HCs, subjective cognitive impairment (SCI), mild cognitive impairment (MCI), and dementia [26] and that TUGdt can predict conversion to dementia from SCI and MCI over a period of 2.5 years. Other novel findings [27] are that the TUGdt parameter words/time can predict conversion to dementia over a period of up to five years in participants with SCI or MCI. Among the TUG-related parameters investigated, word/time shows the best predictive capacity for dementia, while time scores of TUG and TUGdt, as well as TUGdt cost, are not statistically significant predictors. Results further show that the step parameter step length during TUG single task can predict conversion to dementia before adjustment for age, gender, and education. Moreover, the results also comprise optimal TUGdt cutoffs for predicting dementia at 2- and 4-year post-baseline. The sensitivity of the TUGdt cutoffs is high at 2-year follow-up and slightly reduced after 4 years: TUGdt-NA words/time, 0.79 and 0.64; TUGdt-MB words/time, 0.71 and 0.65. Recent UDDGait results also show that extracted step parameters during TUG can discriminate between groups with different levels of cognitive ability i.e. groups with dementia, MCI, SCI, and HCs, based on gait parameters extracted from the TUG test using the OpenPose model from video data [28], using odds ratios to identify significant differences in step characteristics.

Based on the systematic literature review (Paper I) conducted, most existing studies on dementia have relied on Kinect sensors for gait data collection, which shows their promise in classification, however, these approaches have not extended to the use of video camera-based systems, which represent a potentially non-invasive alternative. This gap highlights the need for further research to explore the potential of machine learning utilizing video camera-based pose estimation for movement analysis in dementia discrimination

Seifallahi et al. [29] utilized a Kinect V.2 camera to analyze the TUG test. The study identified 12 significant gait features using the TUG test to distinguish people with Alzheimer's disease (AD) from HCs. The study achieved a classification accuracy of 98.68% using SVM. Similarly, You et al. [30] explored gait analysis using multiple Kinect cameras while participants performed straight-line walking tests for 10 meters. The study employed the LSTM model to analyze skeleton data for distinguishing AD and mild cognitive impairment (MCI) from HCs and achieved a classification accuracy of 90.48% and a sensitivity of 92%.

Zhang et al. [31] employed a Kinect 2.0 camera to analyze single-task (walking for 4-5 straight line meters) and dt gaits (counting down from 20 in increments of three) using a CNN-based classifier. The study achieved an accuracy of 58.75% and a sensitivity of 74.10 %.

Kondragunta et al. [32] utilized the Kinect V2 for gait data collection, analyzing participants' walking patterns during a 6-meter walk under both single-task and dt conditions. The dt included counting backward and reciting words with special letters, alongside regular walking. The study employed SVM for classification, achieving an accuracy of 87.3% in single-class classification. Aoki et al. [33] utilized a Kinect sensor to collect gait data for people walking for 1 minute. The study demonstrated that dt gait features (counting down from 100 by ones) outperformed single-task features. The study achieved an ROC of 74.7% using SVM.

Building on the findings of Paper I, which showed the potential for gait analysis for dementia detection using gait analysis, and Paper II, which analyzed gait features for dementia discrimination from video data, Paper III extended the investigation by exploring different machine learning approaches for PD classification using gait analysis. In the context of PD, the literature shows that various studies have applied deep learning models, such as CNN and LSTM, for PD classification. However, there is a gap in comprehensive investigations comparing the performance of these deep learning approaches with traditional machine learning methods in terms of accuracy and time complexity, which will allow for the practical evaluation of selecting the suitable model for PD detection Paper III addresses this gap by assessing various models and their combinations, providing a robust comparative analysis.

Zhao et al. [34] employed a hybrid CNN-LSTM model for classifying PD and HCs. This method combines spatial feature extraction through CNN with temporal pattern through LSTM, achieving an accuracy of 98.61%. Maachi et al. [35] used a 1D-Convolutional Neural Network model on data from vertical ground reaction force (VGRF) signals from foot sensors for classifying PD and HCs, achieving a classification accuracy of 98.7%. Yurdakul et al. [36] implemented a neighborhood representation local binary pattern for classifying PD disease patients and HCs, achieving a classification accuracy of 98.3%.

Xia et al. [37] employed a Dual-Modal Attention-Enhanced Deep Learning Network, achieving 99.31% accuracy in classifying PD and HCs. Balaji et al. [38] utilized the LSTM model for classifying people with PD and HCs using gait data, achieving a classification accuracy of 98.6%. Aversano et al. [39] used a Deep Neural Network architecture for classifying PD patients and HCs using VGRF data, achieving a classification accuracy of 99.37%.

1.2 Research Questions and Aims

This thesis examines movement analysis for the discrimination of NDDs using machine learning, specifically dementia and PD from HCs. The research begins with a review of existing studies for dementia detection through non-invasive methods, followed by empirical studies that apply machine learning techniques to discriminate against dementia and PD from HCs based on movement analysis. The main objective is to evaluate the feasibility of mobility tests in discriminating NDDs (dementia and PD) from HCs.

The following research questions were addressed:

1. What are the non-invasive methods for discrimination between dementia and HCs, using machine learning along with the features that have been utilized? Additionally, investigating the characteristics of the data used for these approaches?
2. How effectively can movement analysis using video-based data discriminate between HCs and those with dementia using machine learning?
3. What are the most relevant movement features during the TUG test under single and dt to discriminate dementia?
4. How does sensor-based movement analysis, by employing traditional machine learning and deep learning models, classify individuals as HCs or those affected by PD, and how do these models compare in terms of performance?

Table 1: The research questions and how the thesis answers them.

Research Question	Paper I	Paper II	Paper III
RQ1	X		
RQ2		X	
RQ3		X	
RQ4			X

The first question is addressed in Paper I of the thesis. Based on the results of this study, the last three questions were developed. Questions 2 and 3 are answered in Paper II and the last question is addressed in Paper III of the thesis, as shown in Table 1.

1.3 Contribution of the Thesis

This thesis employs movement analysis using machine learning algorithms to develop a non-invasive method to discriminate against NDDs (i.e., dementia and PD) compared to HCs.

The contribution of this thesis starts with Paper I, by conducting a systematic review to provide a comprehensive synthesis of machine learning approaches for dementia detection using non-invasive methods such as gait and speech analysis. It identifies gaps in the literature, summarizes the machine learning models, the features used in the literature, and the characteristics of the datasets used. Furthermore, the review provided future directions for early dementia detection using machine learning through gait analysis.

Building on the findings from the systematic literature review, which identified a gap in the application of machine learning to video-based movement analysis for dementia detection. Paper II employed machine learning for UDDGait™ [23] video data for people with dementia and HCs. The review highlighted that while gait analysis is a promising non-invasive approach, existing studies primarily relied on sensor-based approaches or simple statistical analyses, with limited exploration of video-based analysis using machine learning models. To address this gap, this study utilized UDDGait™ video data of individuals with dementia and HCs and applied pose estimation using YOLOv8 to extract keypoints during mobility tests (TUG single and dual tasks: TUGdt-NA and TUGdt-MB). From these keypoints, movement features including velocity, acceleration, stride length, number of steps, cadence, and knee and hip angles were computed and then used in machine learning algorithms (Random Forest (RF), Support Vector Machines (SVM), and Logistic Regression (LR)). Moreover, the current study expands on the findings of the UDDGait studies [23-28] by incorporating more sets of movement features and applying machine learning techniques to discriminate between dementia and HCs.

Paper III contributed by comprehensively comparing deep learning and traditional machine learning models using movement data. It evaluates advanced deep learning architectures (Convolutional Neural Networks (CNN), Long-Short Term Memory (LSTM), and CNN-LSTM), alongside traditional machine learning methods (Decision Tree (DT) and RF). Additionally, the study provides insights into the trade-offs between model complexity and computational efficiency. This comprehensive analysis provides insights into the selection of the models based on the specific needs of clinical or computational contexts.

2 Methodology

This section describes the systematic literature review, the datasets, features, and machine learning algorithms used in the studies.

Paper I conducted a systematic literature review aimed at exploring the non-invasive methods for dementia detection using machine learning, along with the commonly used features and dataset characteristics in previous studies. The PRISMA protocol was followed to search for three selected databases (Scopus, Web of Science, and PubMed) that include studies published between 2017 and 2022. Given the gaps and the potential of gait analysis in literature, Papers II and III are developed.

Figure 1 demonstrates the methodology followed in Papers II and III. The process starts with data acquisition and then preprocessing the data to ensure the dataset's suitability for analysis. Following this, feature engineering is performed to extract relevant movement features. These features are used as inputs for a machine learning algorithm, which is trained to discriminate individuals into distinct groups, such as (HCs and people with dementia) or (HCs and persons with PD).

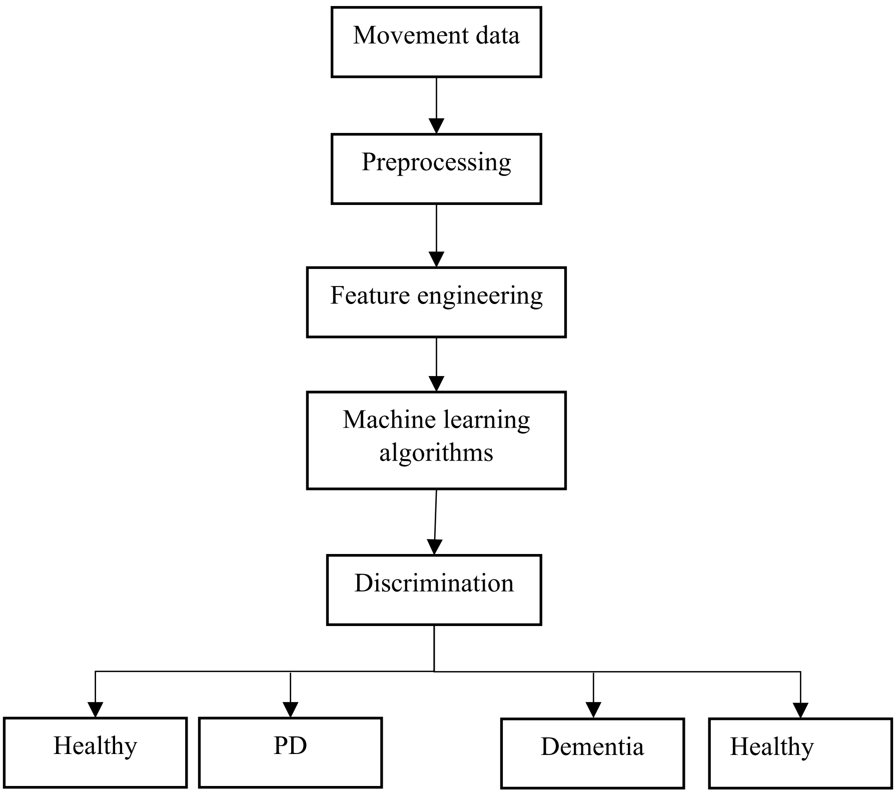

Figure 1: Methodology overview followed in the studies used.

2.1 The Datasets

In this thesis, two primary types of data were used:

1) Video-based data for the study on dementia

2) Sensor-based gait data for the PD study.

Each dataset was utilized to provide features to analyze the movement and was subsequently employed in machine learning for discriminating between HCs and persons with NDDs. Table 2 provides the demographics and the number of participants for both Paper II and Paper III.

Table 2: Demographic characteristics of the datasets for both studies

Variable	Dementia		PD	
	HC (67)	Dementia (64)	HC (73)	PD (93)
Age (mean ± sd)	71.03 ±8.05	77.22 ±7.27	66.3 ±10.67	66.3 ±9.58
Gender (M:F)	37:30	33:31	40: 33	58: 35
Height, mean ± sd(cm)	170.76 ±8.38	167.15 ± 8.12	174 ±0.08	173 ±0.10
MMSE_SR, mean ± sd	29.22 ± 0.91	22.02 ± 3.61	-	-

Sd: Standard deviation; HC: healthy controls; M: Male; F: Female; MMSE_SR: Swedish version Mini-mental state examination; PD: Parkinson's disease

2.1.1 Dementia Disorder Dataset

The UDDGait™ dataset used in Paper II includes recordings of participants performing TUG under single-task and dt conditions. In the single task, participants are instructed to perform a continuous mobility task, starting from a chair with armrests, including sit-to-walk, walking a distance of 3 meters, turning, walking back, and sitting down again on the chair. While in the dt, they perform the TUG with an additional verbal/cognitive task such as naming animals or reciting months in reverse order. The baseline dataset used in the current study comprises 393 videos: 201 videos from 67 HCs and 192 videos from 64 persons with dementia (47 with AD and 17 with unspecified dementia). Videos were recorded at 30 frames per second with a resolution of 1920 x 1080 pixels. The setup of the TUG test is shown in Figure 2.

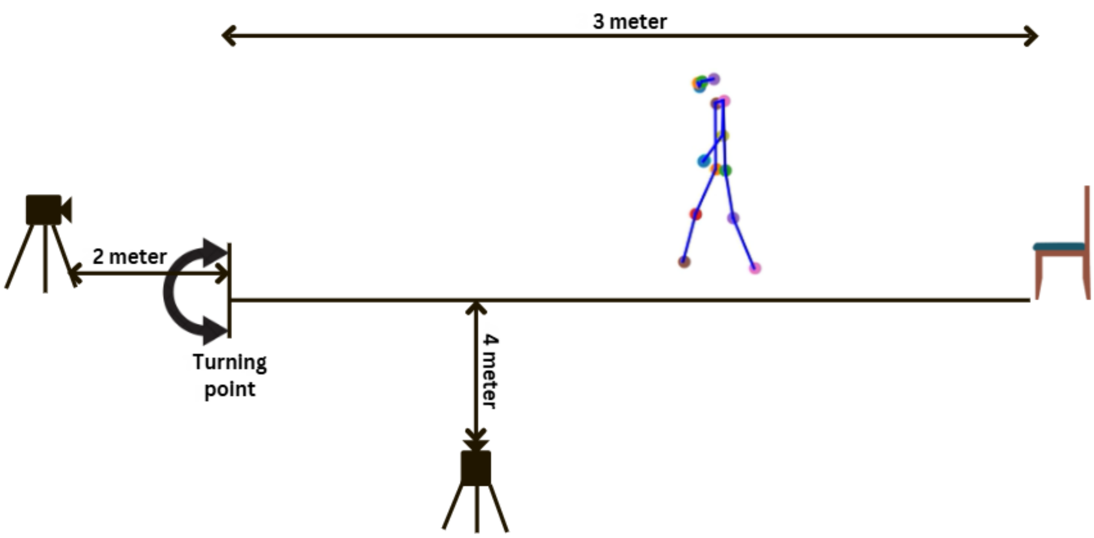

Figure 2: Setup of the TUG test illustrating the placement of the two cameras.

2.1.2 Parkinson's Disease Dataset

The dataset in Paper III comprises gait data from 93 individuals with PD and 73 HCs. The VGRF was recorded for about two minutes as participants walked at their own pace. The data was collected using 16 sensors (8 per foot) measuring force over time and sampled at 100 Hz. The dataset includes 19 parameters: 16 sensor outputs (8 per foot), 2 total VGRF outputs, and a timestamp. The dataset integrates data from three studies: dual-tasking experiments (Galit Yogev et al. [40]), walking under varying conditions (Hausdorff et al. [41]), and treadmill walking (Silvi Frenkel-Toledo et al. [42]). Figure 3 shows the positions of the sensors for both the left and right foot.

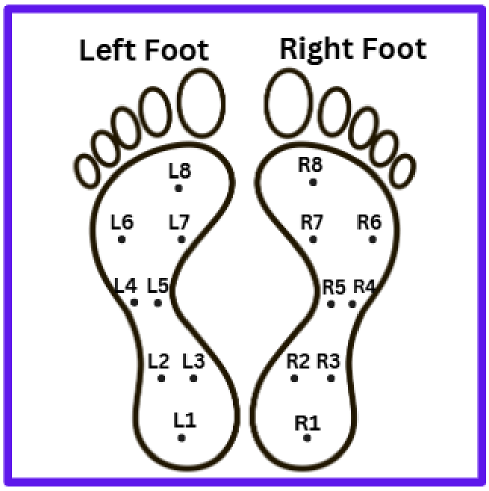

Figure 3: Position of the force sensors under each foot (8 per foot)

2.1.3 Parameters Used in the Studies

Both studies used movement parameters to discriminate between healthy participants from dementia or PD. Paper II for dementia discrimination used movement features such as velocity, stride length, and knee/hip angles during the mobility test (TUG, TUGdt-NA, and TUGdt-MB). Paper III employed VGER and temporal dependencies measured through foot sensors. Table 3 summarizes the features used with their description for both Papers II and III.

The parameters included in Paper II were velocity, acceleration, count of steps, cadence, stride length, total test time, knee angle, and hip angle. Table 3 provides a brief description of those features. Based on the systematic review (Paper I), the study finds that velocity, stride length, gait variability, and whole-body movement enhance the classification accuracy for dementia detection. Hence, the above features were extracted.

Velocity is the displacement of the body's center of gravity (COG) over time. It is calculated based on the displacement between frames upon time interval. Acceleration is defined as the rate of change in velocity.

The number of steps taken during the TUG test is derived based on the peaks of the vertical position in the y coordinate from the ankle keypoints. Cadence is calculated as the number of steps per minute.

Stride length is the distance traveled by two consecutive steps on the same foot while walking. It is computed by identifying peaks in the vertical positions of the left and right ankles, corresponding to each foot's highest point during a step. The horizontal distance between consecutive peaks is calculated and converted into meters through a scaling factor to get the stride length.

Table 3: Description of the features used in both the studies (Papers II and III).

Dementia study (II)		PD (III)	
Feature	Description	Feature	Description
Velocity	The displacement per unit of time for the participants' body COG	Time Stamp	Records the time of measurement in seconds
Acceleration	Rate of change of velocity over time for the participants' body COG	VGRF (Left Foot Sensors)	Vertical Ground Reaction Forces from eight sensors (L1-L8).
Number of steps	The number of steps taken by the participants, calculated based on peaks in the vertical position of keypoints	VGRF (Right Foot Sensors)	Vertical Ground Reaction Forces from eight sensors (R1-R8).
Cadence	Number of steps completed per minute	Total Force (Left Foot)	The combined force recorded from all sensors under the left foot.
Stride length	The distance covered between two successive steps made by the same foot. It begins with the heel strike of one foot and ends when the same foot makes its next heel strike.	Total Force (Right Foot)	The combined force recorded from all sensors under the right foot.
Total time	The duration required to complete the TUG test.	-	-
Knee angle	The measure of the angle between the leg and the thigh, representing the relative orientation of these segments during movement	-	-
Hip angle	The measure of the angle between the pelvis and the thigh, representing the relative orientation of these segments during movement	-	-

COG: Centre of gravity; VGER: Vertical ground reaction force

The knee and hip angles were obtained based on the vector method. For instance, the knee angle is determined using the keypoints of the hip, knee, and ankle. Vectors formed by these points are analyzed with the dot product to measure the angle. The formulas used to calculate velocity, acceleration, and angles are shown in Table 4. The calculated movement parameters were normalized with body height, which facilitates taking into account the individual differences. Moreover, the study adjusted the features with age to account for the confounding effects of age-related changes in gait biomechanics.

Paper III utilized gait features, which are the VGER measured nearly two minutes for each foot (L1-L8 for the left foot and R1-R8 for the right foot), using 16 sensors (8 for each foot) to capture forces as a function of time at a sampling rate of 100 Hz. Additionally, the total force for each foot represents the combined force from all sensors, providing an overall measure of vertical force for the left and the right foot.

Table 4: Formulas used for calculation of velocity, acceleration, and angles

Feature	Formula
Velocity	$\sqrt{(\Delta x)^2 + (\Delta y)^2}$ (1)
Acceleration	$\frac{Velocity}{\Delta t}$ (2)
Angle	$tan^{-1}(\frac{a_x.b_y - a_y.b_x}{a_x.b_x + a_y.b_y})$ (3)

2.2 Data Preprocessing

Before applying the machine learning algorithms, both datasets underwent a preprocessing step to ensure the data readiness for analysis.

In Paper II, the recorded videos were segmented to include only the portions relevant to the TUG test, specifically focusing on the single-task and dt (animal naming and reciting months in reverse order) components. This step ensured that irrelevant parts of the recordings, such as initial instructions by the clinician, were excluded. Subsequently, pose estimation techniques were applied to the segmented videos to extract keypoints representing body movements. These keypoints were then used to derive features that are used as inputs for machine learning models. To enhance model performance, all features were standardized to have a mean of 0 and a standard deviation of 1 for SVM and LR, which are sensitive to the scale of the input features.

In Paper III, to prepare the dataset for analysis, all features were first standardized to a mean of zero and a standard deviation of one. This standardization ensures a uniform scale that is crucial for facilitating the convergence of deep learning models and enhancing overall model performance. Next, multicollinearity was addressed by making a correlation heatmap to identify highly correlated features, such as L1-L3, L4-L8, R1-R3, and R4-R8. Representative features (Time, L1, L4, R1, R4, Total Force Left, and Total Force Right) were selected to remove the redundancy in the features. Given the imbalance present in the dataset between participants with PD and HCs, data balancing was applied

2.2.1 Converting Videos into Features

To convert the movement of participants from videos into features for analysis, pose estimation was implemented. Pose estimation involves detecting and tracking the positions of a person's body parts through images or videos. This is accomplished by finding key features of the human body, such as joint landmarks. In Paper III, the YOLOv8 model is utilized for 3D pose estimation (x, y coordinates along with visibility) to identify 17 keypoints, for detecting the participants' movement.

The 17 keypoints identified are: Nose, Left Eye, Right Eye, Left Ear, Right Ear, Left Shoulder, Right Shoulder, Left Elbow, Right Elbow, Left Wrist, Right Wrist, Left Hip, Right Hip, Left Knee, Right Knee, Left Ankle, and Right Ankle. Figure 4 illustrates the sequence of the TUG test, captured through video analysis with pose estimation. The figure highlights the keypoints of a participant performing the test, starting from sitting, then standing, walking, turning, walking back, and sitting again. These keypoints were extracted through pose estimation to detect the participant's movements throughout the entire test.

YOLOv8 is a pre-trained pose estimation model trained on the COCO dataset [43]. The visibility score measures how the keypoints are visible in the images, with values between 0 and 1, which indicates the highest level of visibility. In Paper III, the average visibility score for the YOLOv8 pose model for the dataset is 0.77.

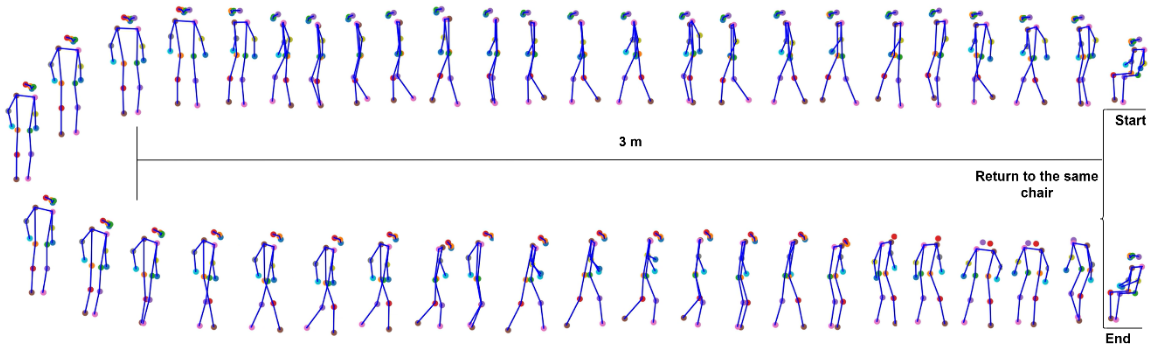

Figure 4: Visualization of the TUG test.

The YOLOv8 pose model was used to identify and locate skeletal landmarks from video frames, capturing keypoints that were employed for feature extraction. By determining joint positions, it allows for the analysis of movement features through the extraction of relevant features from these keypoints. These features are then applied in machine learning models for discrimination between people with dementia and HCs.

2.2.2 Cross-Validation

Paper II employed two cross-validation methods, Leave-One-Out Cross-Validation (LOOCV) and 5-Fold Cross-Validation. These approaches are used to assess the performance of the machine learning models in differentiating between HCs and people with dementia. LOOCV involves leaving out all the observations from a single individual as the test set while using the remaining data for training. This ensures that no data from the same individual is present in both training and testing, which helps in preventing data leakage and enhancing the robustness of the evaluation. The 5-fold cross-validation method divides the dataset into five equal partitions; one-fold is used for testing and the remaining four for training.

In paper III, a 70/30 split was utilized to divide the dataset into training and testing sets; 70% of the data was used for training and 30% for testing to evaluate the performance of the machine learning models. The study tested with different splits such as 80/20, 70/30, and CV, and because of the limitation of the conference, the study mentioned only 70/30.

2.3 Machine Learning Models

Machine learning algorithms were used to discriminate individuals as either healthy or affected by dementia or PD. These models were trained on the movement features extracted from the datasets. Paper II employed traditional machine learning algorithms (RF, SVM, and LR). Paper III implemented both machine and deep learning models (RF, DT, CNN, LSTM, CNN-LSTM).

In both Paper II and III, hyperparameter optimization was performed to fine-tune the parameters of the machine learning algorithms, aiming to improve the model performance. To assess the performance of the machine learning models, evaluation metrics were used, including accuracy, precision, recall, and F1-score. These metrics helped assess how well the models differentiated individuals as healthy or affected by NDDs.

Different machine learning algorithms were employed in this thesis. SVM and LR were implemented in paper II. RF was utilized in Papers II and III. Moreover, DT, CNN, LSTM, and CNN-LSTM models were implemented in paper III.

2.3.1 Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm used for classification and regression problems. It works by identifying a hyperplane that best separates different classes of data with the maximum margin, making it effective in high-dimensional spaces [44].

The SVM algorithm was fine-tuned using grid search, which explored combinations of parameters to identify the most effective settings for model performance. The parameters tested included different values of the regularization parameter (C), kernel types (linear, RBF, Polynomial), gamma settings (scale, auto), and polynomial degrees (2, 3, 4). The best-performing parameters identified for TUGdt-NA were (RBF), C = 1, and gamma = scale). For the TUGdt-MB were (linear kernel, C = 0.1, and gamma = scale). For the TUG single-task were (linear kernel, C=0.7, and gamma = scale).

2.3.2 Logistic Regression (LR)

Logistic regression is a machine learning method used for modeling binary outcomes by estimating the probability of an event occurring based on given predictors. It calculates odds ratios to measure the effect of predictors on the event likelihood. LR applies a logistic function to ensure the output values fall between 0 and 1 [45]. In paper II, LR was implemented with maximum iteration set to 1000.

2.3.3. Random Forest (RF)

RF is an ensemble machine learning algorithm that combines multiple decision trees and aggregates their predictions by averaging. RF is effective for classification and regression tasks and can handle datasets with high dimensionality [46].

In Paper II the RF configurations for the single-task TUG included 50 trees, a maximum depth of 3, minimum samples per leaf, and a split set at 5. In the TUGdt-NA, the model used 50 trees, a maximum depth of 10, and required only 1 sample per leaf and at least 10 to split. For the TUGdt-MB (reciting months), the setup was adjusted to 200 trees, a maximum depth of 3, with 10 minimum samples per leaf and to split. In Paper III, the RF algorithm was implemented with 100 trees and 42 random states for reproducibility.

2.3.4 Decision Trees (DT)

A decision tree is a data mining method used for classification and prediction. It segments a population into branch-like segments, forming an inverted tree structure. This non-parametric method can handle large and complex datasets [47]. The DT was implemented in Paper III with a random state of 42.

2.3.5 Convolutional Neural Network (CNN)

CNN is a type of neural network that includes multiple layers, with each layer made up of interconnected neurons. These layers are comprised of convolutional layers for extracting detailed features through filters, pooling layers designed to reduce dimensionality, and fully connected layers [48]

The CNN used in Paper III was configured with an input layer that accepts two-dimensional input with a single channel, followed by six convolutional layers using 3x3 kernels. The initial two layers each contain 32 filters, while the subsequent four layers increase to 64 filters each. These layers employ the Rectified Linear Unit (ReLU) activation function to enable non-linear processing of features. To reduce spatial dimensions, max-pooling layers with a kernel size of 2 and a stride of 2 are used. Dropout layers with a rate of 0.5 are integrated to prevent overfitting by randomly deactivating neurons during training. The model concludes with two fully connected layers, with the first containing 8 neurons and the second consisting of 2 neurons, designed for binary classification.

2.3.6 Long Short-Term Memory (LSTM)

LSTM is a type of recurrent neural network (RNN) that is specialized for processing sequences of data. It has the capability to learn long-term dependencies through its unique architecture involving four interactive operations (three sigmoid and one tanh), enabling robust memory capabilities [49].

The LSTM model employed in Paper III is structured to process sequential data effectively. It begins with an input layer configured for sequences with a batch size of 32, a feature size of 7, and a sequence length of 400. Following this, there are two LSTM layers, each with a hidden size of 64, which are crucial for capturing the temporal dependencies present in the data. A dropout layer with a rate of 0.5 is incorporated to mitigate the risk of overfitting. Following that, the architecture contains two fully connected layers; the first layer reduces the data's dimensionality to 128, and the subsequent layer further compresses it to a dimension of 2 for the binary classification.

2.3.7 The hybrid CNN-LSTM

The CNN-LSTM model leverages the strengths of CNN and LSTM networks to process data that contain spatial and temporal components. This hybrid architecture utilizes CNN layers for effective spatial feature extraction from input data, followed by LSTM layers that capture temporal dependencies and sequences in the data.

The CNN-LSTM in Paper III model was designed with an input layer to handle data sequences characterized by a batch size of 32, a feature size of 7, and a sequence length of 400. It has a one-layer CNN with an output channel of 64, a kernel size of 3, and padding of 1, using the ReLU activation function followed by max pooling with a kernel size of 2. This CNN setup is aimed at extracting spatial features, which are then fed into a two-layer LSTM setup with 128 hidden units and an input size of 64, designed to capture temporal dependencies. A dropout layer with a rate of 0.2 to prevent overfitting was performed. The learning rate is set to 0.0008. The model concludes with two fully connected layers of sizes 128 and 2, respectively.

3 Results and Discussion

This thesis explores the non-invasive methods to discriminate people with NDDs from HCs, specifically focusing on dementia and PD, emphasizing movement analysis as a useful and automated approach.

The first research question was to identify the non-invasive and affordable (relatively easy setup) methods for the detection of dementia, the features commonly used, and the characteristics of the datasets employed in previous studies. To address this, Paper I conducted a systematic literature review on machine learning approaches for dementia detection. This systematic review highlighted the potential of non-invasive methods, such as gait and speech analysis, for early dementia detection. Machine learning techniques (SVM and LR) were the most frequent algorithms used in the literature. Movement features such as walking speed, stride length, and variability were identified as important features for early detection. The review showed that features extracted from the whole-body movements enhanced classification performance. Moreover, dt paradigms, such as walking while performing a cognitive task, show to improve prediction accuracy.

The review points out the limited number of studies utilizing non-invasive methods, highlighting that no research has applied machine learning to data collected through a simple camera for movement analysis. Moreover, it also indicated that data used in previous studies are small in size and emphasize the need for larger, more heterogeneous datasets.

Based on the gaps in Paper I, movement analysis using machine learning through video-based data is proposed in Paper II. This study developed an automated system for dementia discrimination using movement analysis through video-based pose estimation and machine learning. Using video data from 64 people with dementia and 67 HCs performing the TUG tests under single and dt (TUGdt-NA, and TUGdt-MB) conditions. The study used YOLOV8 pose estimation to extract keypoints and identify features for analysis.

Fourteen movement features were extracted and used for machine learning algorithms. This study identified significant differences in the movement parameters between the two groups. Features such as test duration, stride length, velocity, and step counts showed statistical significance between groups. This study extracted hip and knee angles that have not been explored in the literature. Using machine learning algorithms (SVM, LR, and RF), the study achieved the highest discrimination performance with SVM in the naming animal dt with an accuracy of 88.5% and recall of 94% using 5-fold cross-validation. The results of the machine learning algorithms are shown in Table 5 for 5-fold cross-validation and Table 6 for LOOCV. These results indicate that the dt approaches enhance the machine learning model performance. The results highlight the importance of task selection to enhance discrimination accuracies. There is a 10% increase in the dt animal naming using 5-fold cross-validation and 6% in the reciting months in reverse order compared to the single task using SVM. RF did not show a big performance, and the SVM responded much stronger to the nonlinearity in the dataset.

The use of video-based pose estimation provides an alternative to the traditional methods for dementia discrimination from HCs, as video systems are relatively easy to set up and use. Compared to prior research, this study is the first to employ YOLOv8 for pose estimation using simple camera data and then employing machine learning, providing an alternative to the Kinect-based methods used in the literature [29, 33]. Moreover, Åberg et al. [50] validated the extraction of gait parameters using marker-free video recordings of TUG tests, confirming the method's accuracy and reliability. Similarly, Halvorsen et al. [51] demonstrated the feasibility of deep-learning-based identification of heel keypoints from video-recorded gait, utilizing convolutional neural networks to improve marker-less gait analysis.

Table 5: Evaluation metrics of machine learning models using 5-fold cross-validation

Model	Task	Accuracy	Recall	Precision	F1-Score
SVM	Single task	80.9%	80.5%	81.8%	81.2 %
	Dual-task animal naming	88.5%	92.5%	86.1%	89.2%
	Dual-task reciting months	85.5%	88.8%	84.2%	86.1%
LR	Single task	83.9%	88%	81.9%	84.9%
	Dual-task reciting months	85.5%	88%	84.2%	86.1%
	Dual-task animal naming	85.5%	92.5%	81.5%	86.7%
RF	Single task	78%	81%	78%	78%
	Dual-task reciting months	79%	80%	76%	78%
	Dual-task animal naming	79%	78%	79%	78%

Table 6: Evaluation metrics of machine learning models using LOOCV cross-validation

Model	Task	Accuracy	Recall	Precision	F1-Score
SVM	Single task	83.9%	83.5%	84.8%	84.2 %
	Dual-task animal naming	85.4%	94%	80.7%	86.6%
	Dual-task reciting months	86.2%	89.5%	84.5%	86.9%
LR	Single task	82.4%	86.5%	80.5%	83.4%
	Dual-task reciting months	84.7%	89.5%	82.1%	85.7%
	Dual-task animal naming	85.5%	91%	82.4%	86.5%
RF	Single task	80%	83.5%	78.8%	81.1%
	Dual-task reciting months	83.9%	86.5%	82.8%	84.6%
	Dual-task animal naming	81.6%	82%	82%	82%

Paper III implemented various machine learning and deep learning models for differentiating PD and HCs. The study employed gait sensor-based data from 93 PD participants and 73 HCs. The results showed that the RF model achieved the highest accuracy and recall (96%), demonstrating its robustness in handling complex and non-linear relationships in the dataset. The CNN-LSTM hybrid model followed closely with an accuracy of 95.49% and a recall of 91.5%, effectively leveraging its ability to combine spatial and temporal features. The results of the machine learning models are shown in Table 7. These findings emphasize the advantages of ensemble and hybrid approaches in achieving higher recall and accuracy, which is crucial for discriminating between the groups. The test accuracies for CNN, LSTM, and CNN-LSTM are shown in Figures 5, 6, and 7.

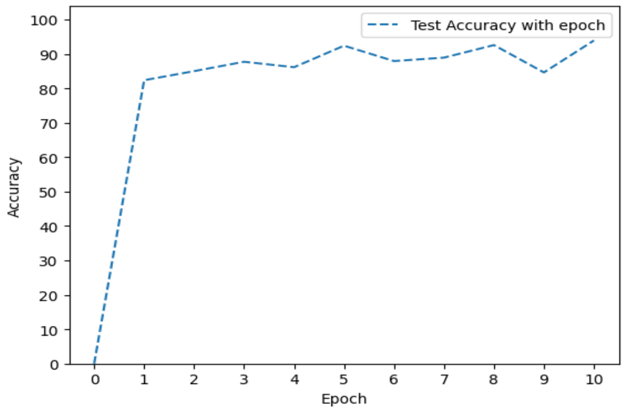

Figure 5: Test accuracy of the CNN model

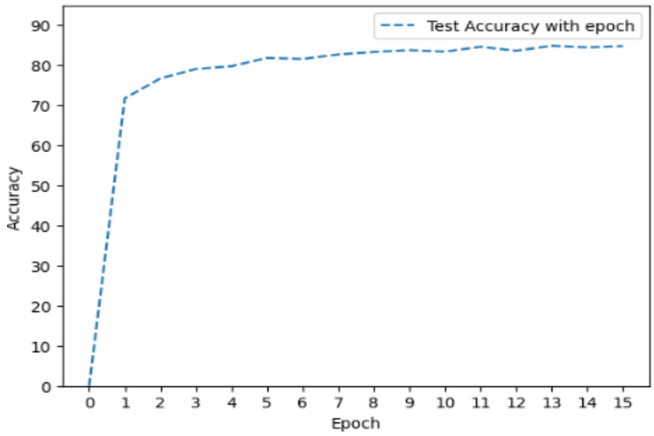

Figure 6: Test accuracy of the LSTM model

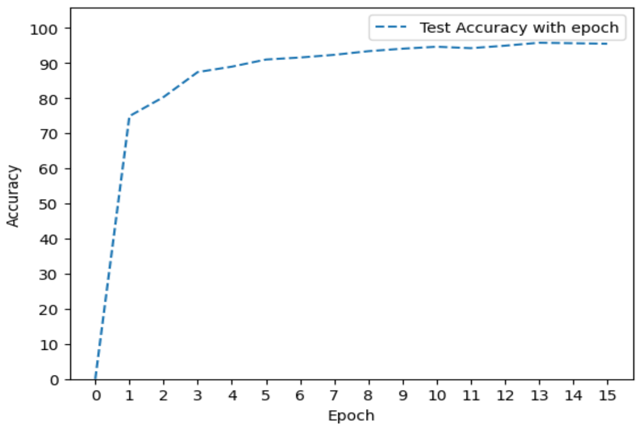

Figure 7: Test accuracy of the CNN-LSTM model

This study highlights the trade-off between accuracy and time complexity, with CNN having the highest training time (2291.13 seconds) and DT being the least (189.06 seconds), while RF achieved a balance with higher accuracy and moderate training time (200.70 seconds). The CNN-LSTM model achieved comparable accuracy to RF; however, it has a higher training time (322.80). This study is the first to compare machine learning and deep learning models for PD disease discrimination in terms of accuracy and time complexity. This comparison is valuable as it provides insights into the trade-offs between computational efficiency and predictive performance, aiding in the selection of the most efficient model based on specific needs and clinical requirements.

Table 7: Evaluation metrics of machine learning for PD study

Metrics	Accuracy	Precision	recall	F1 score
CNN	93.9%	99.4%	88.5%	93.6%
LSTM	84.6%	84.0%	86.0%	85.0%
CNN-LSTM	95.4%	98.0%	91.5%	94.7%
DT	83.0%	83.0%	83.0%	83.0%
RF	96.0%	96.3%	96.0%	96.2%

The findings from the systematic literature review paper, supported by empirical studies, (Paper III and II) demonstrate the potential of movement analysis and machine learning algorithms as an affordable, automated alternative for the discrimination of NDDs, particularly dementia and PD.

A major obstacle in the medical domain is data collection, primarily due to privacy concerns. Medical data collection involves significant challenges, such as obtaining informed consent from participants, anonymizing data, and ensuring secure storage. While these privacy protections are essential to uphold ethical standards, they complicate the data collection process and limit access to large datasets, thereby making it more challenging for researchers to obtain and utilize data effectively. As a result, this thesis used the video-based dataset for dementia and the publicly available dataset for PD. Comparing the performance metrics of both datasets, the sensor-based data achieved higher results compared to the video data. This could be due to its higher precision and ability to capture detailed movement.

In this thesis, the discrimination was dementia against HCs and PD against HCs, it does not include a direct comparison of dementia against PD. This method ensures condition-specific insights from the available datasets. Moreover, comparing dementia and PD would require a dataset containing both conditions, which was not accessible for this thesis.

In terms of validation method, LOOCV and 5-fold cross-validation were applied in the dementia study to ensure robust model evaluation. In the PD study, a 70/30 train-test split was used based on experimenting with different splits.

The machine algorithms selection between dementia and PD studies are mainly driven by the nature of the datasets available, the features extracted, and the size of the datasets. In the dementia study, video-based movement analysis was utilized, and the traditional machine learning models (SVM and LR) were employed as they provided better performance, given the structure and nature of the extracted features. Notably, more complex models might not necessarily lead to improved performance, as evidenced by RF performing worse than SVM and LR. This suggests that in the context of dementia discrimination, deep learning models may not offer a significant advantage or even lead to worse performance than simple algorithms unless larger datasets with rich features are utilized. In contrast, the PD study used sensor-based gait data, which contains more temporal and spatial complexity. The larger dataset and high-dimensional feature space made deep learning models more beneficial for capturing movement patterns. RF was used in both studies as it has the capability to handle complex and non-linear relationships, but its performance varied depending on the dataset and feature set. The differences in model selection highlight the importance of tailoring machine learning methods to the specific characteristics of the data and the nature of the features extracted.

4. Conclusions and Future Work

4.1 Conclusion

This thesis investigated movement analysis for the discrimination of NDDs, specifically dementia and PD. The research highlights the potential of machine learning algorithms in distinguishing between HCs and people affected by these conditions using movement features extracted from video and sensor-based data. This thesis contributes by investigating the non-invasive, simple setup, and affordable methods for NDDs through mobility analysis using machine learning.

The findings emphasize that gait abnormalities, such as reduced walking speed, shorter stride length, and increased gait variability, can serve as early indicators of cognitive and motor decline. Moreover, the use of dt in dementia analysis further underscores the importance of task selection in enhancing discrimination accuracy. For PD, the comparison of machine learning and deep learning models demonstrates the advantages of hybrid approaches in capturing complex gait patterns and proving better performance.

This thesis automated the movement analysis of dementia and PD using simple video and sensor data for discrimination between NDDs. By employing these methods, the study contributes to a better understanding of movement and gait abnormalities and the use of machine learning in these conditions to support the development of affordable, automated, and easy setup methods for further research and exploration.

4.2 Limitation

The limitations of this thesis are that the datasets used in both empirical studies were relatively small, and further research with larger, more diverse datasets is needed to improve the generalizability and robustness of the models. Additionally, for the video-based study, the potential sensitivity of pose estimation models to variations in video quality and environmental conditions. Although the analysis was conducted using video recordings captured in a controlled setting, real-world applications may encounter challenges such as poor lighting and visual obstructions, which could potentially compromise the accuracy and reliability of extracted features.

Another limitation is the absence of a direct comparison between dementia and PD groups. The current studies focus separately on dementia against HCs and PD against HCs, but comparing these two conditions would provide more insights into the differences in gait and movement abnormalities and a better understanding of the unique characteristics of each condition. Moreover, comparing the features between the two studies is limited due to the discrepancy in the collection methods.

4.3 Future Work

Future research could focus on incorporating larger and more diverse datasets for both dementia and PD. It could also include different stages of cognitive impairment of dementia, such as MCI and SCI. Additionally, exploring the extraction of more comprehensive features, such as parameters derived from whole-body movements, to capture a broader range of movement characteristics.

Further exploration of speech analysis utilizing the TUGdt tests. By analyzing speech patterns during the dt (i.e., animal naming and reciting months in reverse order), this may allow to identify signs of cognitive decline that are not captured by movement analysis separately. Moreover, employing a multimodal fusion of speech and movement features could improve the accuracy of the results. By integrating data from both modalities, it might be possible to capture a more holistic view of the cognitive and motor impairments, which could provide more robust, accurate, and generalizable results.

5. Summary of the papers

This section provides a summary of the papers included in this thesis. It starts with the systematic literature review, followed by empirical studies on NDDs (Dementia and PD). Each summary includes briefly the study's aim, methodology, results, and conclusions.

5.1 Summary of Paper 1: Machine Learning Approaches for Dementia Detection Through Speech and Gait Analysis: A Systematic Literature

The paper aims to review machine learning methods for detecting dementia through speech and gait analysis, emphasizing non-invasive techniques. It also focuses on identifying the key features used and the characteristics of the datasets included in the studies. This systematic literature review followed the PRISMA guidelines. The review was conducted on studies published between 2017 and 2022, extracted from Scopus, Web of Science, and PubMed. Keywords such as "dementia", "gait analysis", and "speech analysis" and their synonyms guided the search to make sure all relevant literature was included. The studies were selected based on the eligibility criteria. For a comprehensive search, the review included all types of papers (journal papers, conference proceedings, and book chapters). The search was conducted on 13 January 2023. A total of 40 studies met these criteria and were subsequently synthesized. The summaries of the datasets, instruments used, types of gait tasks, types of speech tasks, algorithms, features, and best accuracies present in the literature for the selected studies are shown in Table 3 and Table 5 in Paper I.

The systematic review revealed that gait analysis through non-invasive methods for dementia detection using machine learning is underexplored in the literature, with only limited studies employing it. Moreover, none of the reviewed studies utilized machine learning on video data to analyze gait for dementia detection. The reviewed studies implemented supervised machine learning algorithms, with SVM and LR being the most commonly used algorithms. The studies showed that features including gait speed and stride length distinguish dementia from healthy individuals. Moreover, whole-body movement parameters (such as joint movement of hand and legs) and dt (performing a cognitive task while walking) were shown to enhance detection accuracies, suggesting that incorporating these could be crucial for improving the performance of the algorithms. In terms of datasets for gait studies, the review emphasizes the need for larger and more diverse datasets, as most of the datasets used in the literature are small and limited.

In terms of speech studies, speech analysis for dementia detection was more extensively studied compared to gait studies, with a variety of machine learning techniques employed, including deep learning. Diverse acoustic and linguistic features, such as vowel duration, pitch, phoneme characteristics, pauses, and speech rhythm, were utilized in the literature for detecting dementia. Additionally, automatic transcription methods show comparable performance to manual transcription, highlighting the potential for scalable and automated screening methods. However, the datasets used are often small and language-specific, underscoring the need for larger, multilingual, and more representative datasets to improve the robustness and applicability of speech-based dementia detection models.

The review underscores the potential of machine learning in enhancing early detection of dementia through non-invasive automated systems. The review recommended that future research focuses on developing larger, more diverse datasets for both gait and speech analysis to enhance the robustness and generalizability of machine learning models. It emphasizes the need for multimodal approaches that integrate gait and speech features to improve accuracy in the detection of dementia.

5.2 Summary of Paper II: Motion Analysis for Dementia Detection Based on Pose Estimation and Machine Learning

This study aimed to develop an automated method for the discrimination of dementia from HCs through movement analysis and machine learning using video data. This approach offers a non-invasive alternative to traditional diagnostic methods such as cerebrospinal fluid analysis or brain imaging.

Data were collected from 64 individuals with dementia and 67 HCs, the participants performed the TUG test under single and dt (TUGdt-NA and TUGdt-MB) conditions. In the dt, participants performed the TUG with an added cognitive task, which is naming animals or reciting months in reverse order. The dataset was collected through the UDDGait™, which is a longitudinal study in Sweden.

The study employed the YOLOV8 pose model to estimate 17 keypoints from video frames, capturing the movement during the TUG test. These keypoints, were used to extract features for analyzing movement and discriminating dementia cases from HCs using machine learning. A total of 14 movement features were extracted and analyzed before and after adjusting for age to reduce confounding effects. Statistical Significance was observed in the features such as test duration, number of steps taken, stride lengths, and gait velocities between the two groups.

The study utilized three machine learning algorithms (RF, LR, and SVM to differentiate between HCs and people with dementia. Fine-tuning for all the models is performed to achieve the highest performance. Cross-validation methods (LOOCV and 5-fold) were applied to ensure robust model performance. The results showed that SVM in the dt conditions achieved the highest performance compared to other algorithms with an accuracy of 88.5%, recall of 92.5%, and an F1 score of 89.2% in the TUG animal-naming task with 5-fold cross-validation. Dual-task tests consistently outperformed single-task tests across all models, demonstrating the importance of task selection.

The recall rates, particularly in SVM and LR (92.5% for animal naming using 5-fold cross-validation), underscore the model's ability to identify dementia cases effectively, minimizing the risk of missed detection.

This study is the first study to use video data through machine learning for the discrimination of dementia from HCs, offering an accessible and non-invasive alternative to sensor-based and traditional approaches. Compared to previous studies, this research extracted unique features, such as knee and hip angles, which have not been explored in previous research.

Future directions include expanding datasets to include various cognitive impairments (e.g., MCI, SCI) and integrating multimodal features from speech to enhance the accuracy of the models.

5.3 Summary of Paper III: Parkinson's Disease Classification through Gait Analysis: A Comparative Study of Deep Learning and Machine Learning Algorithms

This study aimed to compare the classification performance of various deep learning models (CNN, LSTM, and CNN-LSTM) with traditional machine learning algorithms (RF and DT) for distinguishing individuals with PD from HCs using sensor-based gait data. This comparison aims to determine the most effective approach for PD classification in terms of accuracy and time complexity.

The dataset used for the study includes 93 subjects with PD and 73 HCs. It is gait data collected through sensors that record the VGER. It consists of 16 sensors embedded under both feet, with 8 sensors per foot, measuring VGER while walking at a natural pace for approximately two minutes.

The preprocessing in this study involved standardizing all features to have a mean of zero and a standard deviation of one, ensuring consistency for model convergence. High multicollinearity among features was addressed by feature selection (Time, L1, L4, R1, R4, “Total Force Left,” “Total Force Right”) to reduce redundancy. To address the imbalance present in the dataset, data balancing was applied. The data was split into 70% training set and 30% testing set.

Deep learning models and machine learning algorithms were developed and trained to classify the subjects as healthy or PD. The models included CNN, LSTM, and an ensemble of CNN and LSTM (CNN-LSTM), compared to RF and DT. The outcome of this study showed that RF outperformed all other models, achieving the highest accuracy of 96% and recall of 96%, showing it is capabilities in handling complex, non-linear relationships within medical datasets. This was followed by the CNN-LSTM hybrid model, which achieved an accuracy of 95.49%. The CNN-LSTM model leveraged the spatial feature extraction from CNN and temporal data through LSTM, showing the strength of combining deep learning techniques for complex medical data classification.

The study also compared the training times of the models, noting that deep learning models (CNN and CNN-LSTM) required significantly more computational time compared to machine learning models like DT and RF. This highlights the trade-off between computational cost and performance accuracy, particularly in medical diagnostics.

In conclusion, this study demonstrates that both deep learning and machine learning models offer valuable tools for PD classification, with RF and CNN-LSTM showing the best performance.

For future directions, expanding the dataset and utilizing the UPDRS features within the dataset to classify varying severity levels of PD and differentiate between various degrees of severity.

References

1) "World Health Organization (WHO)," Mar. 14, 2024. https://www.who.int/news/item/14-03-2024-over-1-in-3-people-affected-by-neurological-conditions--the-leading-cause-of-illness-and-disability-worldwide (accessed Nov. 01, 2024).
2) J. D. Steinmetz et al. "Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021," *The Lancet Neurology*, vol. 23, no. 4, pp. 344–381, Mar. 2024, doi: 10.1016/s1474-4422(24)00038-3.
3) B. M. Kelser, E. M. Teichner, R. C. Subtirelu, and K. N. Hoss, "A review of proposed mechanisms for neurodegenerative disease," *Frontiers in Aging Neuroscience*, vol. 16, Oct. 2024, doi: 10.3389/fnagi.2024.1370580.
4) World Health Organization: WHO and World Health Organization: WHO, "Dementia," Mar. 15, 2023. https://www.who.int/news-room/fact-sheets/detail/dementia
5) World Health Organization: WHO and World Health Organization: WHO, "Parkinson disease," Aug. 09, 2023. https://www.who.int/news-room/fact-sheets/detail/parkinson-disease
6) S. A. Gale, D. Acar, and K. R. Daffner, "Dementia," *The American Journal of Medicine*, vol. 131, no. 10, pp. 1161–1169, Feb. 2018, doi: 10.1016/j.amjmed.2018.01.022.
7) J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, Mar. 2008, doi: 10.1136/jnnp.2007.131045.
8) G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. R. D'Orazio, "Human GAIT Analysis in Neurodegenerative Diseases: a review*," IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 229–242, Jun. 2021, doi: 10.1109/jbhi.2021.3092875.
9) T. Skillbäck et al., "Slowing gait speed precedes cognitive decline by several years," *Alzheimer S & Dementia*, vol. 18, no. 9, pp. 1667–1676, Feb. 2022, doi: 10.1002/alz.12537.
10) M. Al-Hammadi, H. Fleyeh, A. C. Åberg, K. Halvorsen, and I. Thomas, "Machine Learning Approaches for Dementia Detection Through Speech and GAIT Analysis: A Systematic Literature review," Journal of Alzheimer S Disease, vol. 100, no. 1, pp. 1–27, Jun. 2024, doi: 10.3233/jad-231459.
11) W. Maetzler and J. M. Hausdorff, "Motor signs in the prodromal phase of Parkinson's disease," *Movement Disorders*, vol. 27, no. 5, pp. 627–633, Mar. 2012, doi: 10.1002/mds.24973.
12) D. Podsiadlo and S. Richardson, "The Timed 'Up & Go': A Test of Basic Functional Mobility for Frail Elderly Persons," *Journal of the American Geriatrics Society*, vol. 39, no. 2, pp. 142–148, Feb. 1991, doi: 10.1111/j.1532-5415.1991.tb01616.x.
13) M. Belghali, N. Chastan, D. Davenne, and L. M. Decker, "Improving Dual-Task walking paradigms to detect prodromal Parkinson's and Alzheimer's diseases," *Frontiers in Neurology*, vol. 8, May 2017, doi: 10.3389/fneur.2017.00207.
14) H. B. Åhman et al., "Dual-Task performance and neurodegeneration: correlations between timed Up-and-Go Dual-Task test outcomes and Alzheimer's disease cerebrospinal fluid biomarkers," *Journal of Alzheimer S Disease*, vol. 71, no. s1, pp. S75–S83, May 2019, doi: 10.3233/jad-181265.
15) P. Khera and N. Kumar, "Role of machine learning in gait analysis: a review," *Journal of Medical Engineering & Technology*, vol. 44, no. 8, pp. 441–467, Oct. 2020, doi: 10.1080/03091902.2020.1822940.
16) S. M. Alfayeed and B. S. Saini, Human Gait Analysis Using Machine Learning: A Review. 2021. doi: 10.1109/iccike51210.2021.9410678.2021 *International Conference on Computational Intelligence and Knowledge Economy (ICCIK*E), Amity University Dubai, UAE. https://doi.org/10.1109/ICCIKE51210.2021.9410678
17) G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. R. D'Orazio, "Human GAIT Analysis in Neurodegenerative Diseases: a review," *IEEE Journal of Biomedical and Health Informatics,* vol. 26, no. 1, pp. 229–242, Jun. 2021, doi: 10.1109/jbhi.2021.3092875.
18) H. Zheng, M. Yang, H. Wang, and S. McClean, "Machine learning and statistical approaches to support the discrimination of neuro-degenerative diseases based on GAIT analysis," *in Studies in computational intelligence,* 2009, pp. 57–70. doi: 10.1007/978-3-642-00179-6_4.
19) Ç. B. Erdaş, E. Sümer, and S. Kibaroğlu, "Neurodegenerative disease detection and severity prediction using deep learning approaches," *Biomedical Signal Processing and Control*, vol. 70, p. 103069, Aug. 2021, doi: 10.1016/j.bspc.2021.103069.
20) P. C. Negi, S. Negi, and N. Sharma, "Gait Analysis-Based Identification of Neurodegenerative Diseases Using Machine Learning Techniques". *2022 International Conference on Advances in Computing, Communication and Materials (ICACCM)*, 1–6. https://doi.org/10.1109/icaccm56405.2022.10009413.
21) S. Sundari and V. C. Jadala, "Real-Time Neurological Disease Prediction with 3D Single Pose Estimation using MediaPipe," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4s, pp. 595-607, 2024, Nov. 2023.
22) T. Connie, T. B. Aderinola, T. S. Ong, M. K. O. Goh, B. Erfianto, and B. Purnama, "Pose-Based Gait Analysis for Diagnosis of Parkinson's Disease," *Algorithms*, vol. 15, no. 12, Art. no. 474, Dec. 2022, doi: 10.3390/a15120474.
23) Y. Cedervall et al., "Timed Up-and-Go Dual-Task Testing in the Assessment of Cognitive Function: A Mixed methods observational study for development of the UDDGAIT Protocol," *International Journal of Environmental Research and Public Health*, 17(5), 1715. https://doi.org/10.3390/ijerph17051715.

24) A. C. Åberg, L. E. Larsson, V. Giedraitis, L. Berglund, and K. Halvorsen, "Dual-Task Interference of Gait Parameters During Different Conditions of the Timed Up-and-Go Test Performed by Community-Dwelling Older Adults," *J. Aging Phys*. Act., vol. 31, no. 5, pp. 823-832, 2023, doi: 10.1123/japa.2022-0304.
25) H. B. Åhman et al., "Dual-Task performance and neurodegeneration: correlations between timed Up-and-Go Dual-Task test outcomes and Alzheimer's disease cerebrospinal fluid biomarkers," *Journal of Alzheimer S Disease*, vol. 71, no. s1, pp. S75–S83, May 2019, doi: 10.3233/jad-181265.
26) H. B. Åhman et al., "Dual-Task tests Predict Conversion to Dementia—A prospective Memory-Clinic-Based cohort study," *International Journal of Environmental Research and Public Health*, vol. 17, no. 21, p. 8129, Nov. 2020, doi: 10.3390/ijerph17218129.
27) A. C. Åberg et al., "Prediction of conversion to dementia disorders based on timed up and go dual-task test verbal and motor outcomes: a five-year prospective memory-clinic-based study," *BMC Geriatrics*, vol. 23, no. 1, Sep. 2023, doi: 10.1186/s12877-023-04262-w.
28) N. Löfgren et al., "Extracted step parameters during the Timed Up and Go test discriminate between groups with different levels of cognitive ability - A cross-sectional study.," *Research Square (Research Square)*, Mar. 2024, doi: 10.21203/rs.3.rs-4068945/v1.
29) M. Seifallahi, A. H. Mehraban, J. E. Galvin, and B. Ghoraani, "Alzheimer's disease detection using comprehensive analysis of timed up and go test via Kinect V.2 camera and machine learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1589–1600, Jan. 2022, doi: 10.1109/tnsre.2022.3181252.
30) Z. You, Z. You, Y. Li, S. Zhao, H. Ren, and X. Hu, "Alzheimer's Disease Distinction Based On Gait Feature Analysis," *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM*), pp. 1-6, 2020. doi: 10.1109/HEALTHCOM49281.2021.9398984.
31) Z. Zhang et al., "Deep Learning Based Gait Analysis for Contactless Dementia Detection System from Video Camera," *2022 IEEE International Symposium on Circuits and Systems (ISCAS),* pp. 1–5, Apr. 2021, doi: 10.1109/iscas51556.2021.9401596.
32) J. Kondragunta, R. Seidel, and G. Hirtz, "Machine learning based identification of elderly persons with cognitive impairment using dynamic time warping," *Current Directions in Biomedical Engineering*, vol. 6, no. 3, pp. 360–363, Sep. 2020, doi: 10.1515/cdbme-2020-3093.
33) K. Aoki et al., "Early detection of lower MMSE scores in elderly based on Dual-Task GAIT," *IEEE Access*, vol. 7, pp. 40085–40094, Jan. 2019, doi: 10.1109/access.2019.2906908.
34) A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu, "A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data," *Neurocomputing,* vol. 315, pp. 1–8, Mar. 2018, doi: 10.1016/j.neucom.2018.03.032.
35) I. E. Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait*," Expert Systems With Applications*, vol. 143, p. 113075, Nov. 2019, doi: 10.1016/j.eswa.2019.113075.
36) O. C. Yurdakul, M. S. P. Subathra, and S. T. George, "Detection of Parkinson's Disease from gait using Neighborhood Representation Local Binary Patterns," *Biomedical Signal Processing and Control*, vol. 62, p. 102070, Jul. 2020, doi: 10.1016/j.bspc.2020.102070.
37) Y. Xia, Z. Yao, Q. Ye, and N. Cheng, "A Dual-Modal Attention-Enhanced deep learning network for quantification of Parkinson's disease characteristics*," IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 42–51, Oct. 2019, doi: 10.1109/tnsre.2019.2946194.
38) B. E, B. D, V. K. Elumalai, and V. R, "Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network," *Applied Soft Computing*, vol. 108, p. 107463, May 2021, doi: 10.1016/j.asoc.2021.107463.
39) L. Aversano, M. L. Bernardi, M. Cimitile, and R. Pecori, "Early Detection of Parkinson Disease using Deep Neural Networks on Gait Dynamics," 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, Jul. 2020, doi: 10.1109/ijcnn48605.2020.9207380.
40) G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff, "Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?," *European Journal of Neuroscience*, vol. 22, no. 5, pp. 1248–1256, Sep. 2005, doi: 10.1111/j.1460-9568.2005.04298.x.
41) J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, "Rhythmic auditory stimulation modulates gait variability in Parkinson's disease," *European Journal of Neuroscience*, vol. 26, no. 8, pp. 2369–2375, Oct. 2007, doi: 10.1111/j.1460-9568.2007.05810.x.
42) S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff, "Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease," *Movement Disorders*, vol. 20, no. 9, pp. 1109–1114, May 2005, doi: 10.1002/mds.20507.
43) Ultralytics, "Pose," Ultralytics YOLO Docs, Mar. 17, 2025. https://docs.ultralytics.com/tasks/pose/
44) P. Chen, C. Lin, and B. Schölkopf, "A tutorial on ν-support vector machines," *Applied Stochastic Models in Business and Industry*, vol. 21, no. 2, pp. 111–136, Mar. 2005, doi: 10.1002/asmb.537.
45) M. P. LaValley, "Logistic regression," *Circulation,* vol. 117, no. 18, pp. 2395–2399, May 2008, doi: 10.1161/circulationaha.106.682658.
46) G. Biau and E. Scornet, "A random forest guided tour," *Test,* vol. 25, no. 2, pp. 197–227, Apr. 2016, doi: 10.1007/s11749-016-0481-7.

47) Y. Y. Song and Y. Lu, "Decision tree methods: Applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130-135, 2015. doi: 10.11919/j.issn.1002-0829.215044.
48) K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv (Cornell University), Jan. 2015, doi: 10.48550/arxiv.1511.08458.
49) R. C. Staudemeyer and E. R. Morris, "Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks," arXiv (*Cornell University*), Jan. 2019, doi: 10.48550/arxiv.1909.09586.
50) A. C. Åberg et al., "Extraction of gait parameters from marker-free video recordings of Timed Up-and-Go tests: Validity, inter- and intra-rater reliability," Gait & Posture, vol. 90, pp. 489–495, Aug. 2021, doi: 10.1016/j.gaitpost.2021.08.004.
51) K. Halvorsen, W. Peng, F. Olsson, and A. C. Åberg, "Two-step deep-learning identification of heel keypoints from video-recorded gait," *Medical & Biological Engineering & Computing*, Sep. 2024, doi: 10.1007/s11517-024-03189-7.

Appendix 1: Paper I

Appendix 2: Paper II

Appendix 3: Paper III