

POLITECNICO DI TORINO

Master's Degree in Biomedical Engineering



Master's Degree Thesis

Development of Human Pose Estimation and Machine Learning-based algorithms for assessing physical exercise proficiency

Supervisors

Prof. Danilo DEMARCHI

Prof. Paolo BONATO

Dr. Giulia CORNIANI

Candidate

Livia COLUCCI

OCTOBER 2023

Abstract

The aim of this Thesis is to create a framework that employs Machine-Learning algorithms to automatically assess proficiency in the practice of Tai Chi Chuan by analyzing video recordings and extracting information through Human Pose Estimation. Tai Chi is a form of low-impact mind-body exercise characterized by slow and fluid movements and whose positive impacts on health, particularly in relation to balance, have been analyzed by numerous studies.

The data employed to achieve the goal of this Thesis was collected from thirty-two older adults aged between 65 and 85 years who were asked to perform six different Tai Chi exercises chosen in collaboration with Tai Chi experts. Study participants were enrolled regardless of their prior Tai Chi experience to acquire data across various proficiency levels. Tai Chi experts scored each exercise through visual examination. After preprocessing, Human Pose Estimation was performed through MediaPipe, an open-source library developed by Google. For each exercise, the (x,y) coordinates of joints trajectories obtained as output of the skeleton tracking were utilized to normalize the skeleton dimensions and automatically segment videos into single repetitions of the Tai Chi exercise. Subsequently, specific data features, designed in collaboration with the Tai Chi experts to effectively capture movement characteristics relevant to proficiency, were extracted and then selected using the minimum Redundancy Maximum Relevance method. To predict proficiency levels (low, medium, high), the selected data features were fed into a balanced 3-class Random Forest classifier, whose performance was evaluated using a Leave-One-Group-Out Cross Validation. Predictions on the single repetition were finally merged to estimate a single score per subject. This process was followed separately for each Tai Chi exercise in the dataset, leading to the development of exercise-specific models. Overall, the trained models consistently achieved an F1 score exceeding 80% in accurately predicting proficiency levels from video recordings of subjects performing a single Tai Chi exercise.

The results of this thesis showcase the viability of employing Human Pose Estimation and Machine Learning algorithms to automatically assess individuals' competence in performing physical exercises. Additionally, it introduces a comprehensive framework for evaluating Tai Chi proficiency through video recordings. Given the evidence of the benefits of the practice of Tai Chi on balance, the framework will enable further investigations of how a practitioner's proficiency level influences the clinical advantages, hence discerning whether there exists a relation between proficiency and the enhancement of balance. Additional applications might concern the analysis of movement abilities of patients or the assessment of the proficiency of subjects performing other kinds of physical exercise.

Acknowledgements

I extend my deepest gratitude to Professor Danilo Demarchi and Professor Paolo Bonato for granting me the opportunity to live such an incredible experience at the Motion Analysis Lab. The unwavering support I received from them and Dr. Giulia Corniani transcends the scope of this Thesis, and I will forever appreciate their assistance.

I also want to express my gratitude to all those who played a role in the Tai Chi study. First of all Professor Peter M. Wayne, who is at the core of this study, and all his collaborators, for their contribution in conceiving the study, scoring the exercises and helping in the design of the features. I'd also like to thank Professor Wayne for permitting the use of his images in this manuscript. Secondly, I'd like to thank Gloria V. Diaz and Stefano Sapienza, who collected the video data that I analyzed. And last but not least, I'd like to extend my gratitude to Anmolika Bolla and Bellerina Hu, who contributed to the manual segmentation of the Grasp the Sparrow's Tail exercise.

The success of this work would not have been achievable without the invaluable contributions of all these remarkable individuals.

Table of Contents

List of Tables	VII
List of Figures	IX
The Motion Analysis Lab	1
1 Introduction	7
1.1 State of the art	8
1.1.1 Health benefits of Tai Chi Chuan	8
1.1.2 Human Pose Estimation	11
1.1.3 Pre-trained Human Pose Estimation platforms	12
2 Materials and Methods	19
2.1 Data collection	20
2.1.1 Tai Chi exercises	21
2.1.2 Scores	26
2.2 Data processing	27
2.2.1 Skeleton tracking with MediaPipe	28
2.2.2 Skeleton dimensions normalization	31
2.2.3 Automatic segmentation	32
2.2.4 Feature Construction	34
2.3 Model development	35
2.3.1 Feature selection and data projections	35
2.3.2 Model Training	37
2.3.3 Estimation of the final proficiency level	38
3 Results	41
3.1 Grasp the sparrow's tail	42
3.1.1 Validation of the automatic segmentation	42
3.1.2 Classification results	44
3.2 Wave the hands like clouds	47

3.2.1	Performance with both hands	47
3.2.2	Performance with a single hand	50
3.3	Withdraw and push	53
3.4	Brush knee twist step	56
3.5	Golden rooster	59
3.6	Raising the power	62
3.7	Summary of the results	64
4	Discussion	65
4.1	Limitations of the study	66
4.2	Future perspectives	67
5	Conclusions	69
A	Look-up Tables	71
	Bibliography	79

List of Tables

1.1	Evidence of the benefit of the conditions analyzed by Tai Chi clinical studies [26].	10
1.2	Backbone architectures for Human Pose Estimation models. [33].	13
2.1	Automatic segmentation for all exercises; moments of interest and detection strategy.	39
3.1	Number of subjects whose video data was available for each exercise.	42
3.2	Features selected for the Grasp the sparrow’s tail exercise.	45
3.3	Optimized hyperparameters for the Grasp the sparrow’s tail exercise.	46
3.4	Classification model performances in terms of F1 micro and F1 weighted scores for the Grasp the Sparrow’s tail exercise.	46
3.5	Features selected for the Wave the hands like clouds exercise performed with both hands.	47
3.6	Optimized hyperparameters for the Wave the hands like clouds exercise performed with both hands.	48
3.7	Classification model performances in terms of F1 micro and F1 weighted scores for the Wave the hands like clouds exercise performed with both hands.	49
3.8	Features selected for the Wave the hands like clouds exercise performed with one hand.	50
3.9	Optimized hyperparameters for the Wave the hands like clouds exercise performed with one hand.	51
3.10	Classification model performances in terms of F1 micro and F1 weighted scores for the Wave the hands like clouds exercise performed with one hand.	52
3.11	Features selected for the Withdraw and push exercise.	53
3.12	Optimized hyperparameters for the Withdraw and push exercise.	54
3.13	Classification model performances in terms of F1 micro and F1 weighted scores for the Withdraw and push exercise.	54
3.14	Features selected for the Brush knee twist step exercise.	57

3.15	Optimized hyperparameters for the Brush knee twist step exercise.	58
3.16	Classification model performances in terms of F1 micro and F1 weighted scores for the Brush knee twist step exercise.	58
3.17	Features selected for the Golden rooster exercise.	60
3.18	Optimized hyperparameters for the Golden rooster exercise.	61
3.19	Classification model performances in terms of F1 micro and F1 weighted scores for the Golden rooster exercise.	61
3.20	Features selected for the Raising the power exercise.	62
3.21	Optimized hyperparameters for the Raising the power exercise.	63
3.22	Classification model performances in terms of F1 micro and F1 weighted scores for the Raising the power exercise.	63
A.1	Look-up table for the features common to all exercises.	71
A.2	Look-up table for the Grasp the sparrow’s tail exercise.	72
A.3	Look-up table for the Wave the hands like clouds exercise.	73
A.4	Look-up table for the Withdraw and push exercise.	74
A.5	Look-up table for the Brush knee twist step exercise.	75
A.6	Look-up table for the Golden rooster exercise.	76
A.7	Look-up table for the Raising the power exercise.	77

List of Figures

2.1	The data analysis pipeline	19
2.2	The Grasp the sparrow’s tail exercise.	22
2.3	The Wave the hands like clouds exercise performed with both hands.	23
2.4	The Wave the hands like clouds exercise performed with one hand.	23
2.5	The Withdraw and push exercise.	24
2.6	The Brush knee twist step exercise.	25
2.7	The Golden rooster exercise.	26
2.8	The Raising the power exercise.	26
2.9	Examples of wrong or poor skeleton tracking quality when multiple people are present in the frame.	28
2.10	Examples of mirroring for one subject performing the Grasp the sparrow’s tail exercise and Wave the hands like clouds.	29
2.11	Examples of mirroring for one subject performing the Withdraw and Push.	29
2.12	Human keypoints tracked by MediaPipe [91].	31
2.13	Examples of poor skeleton tracking from the lateral view videos.	32
2.14	Example of the evaluation of the optimal number of features for the Grasp the sparrow’s tail exercise.	36
3.1	Bar plot of the time differences between the automatic and the manual segmentation	43
3.2	Optimization of the number of features parameter for the Grasp the Sparrow’s tail exercise.	44
3.3	Sammon projection of the data after Feature Selection for the Grasp the sparrow’s tail exercise.	45
3.4	Confusion Matrices presenting the classification results for the Grasp the Sparrow’s tail exercise.	46
3.5	Optimization of the number of features parameter for the Wave the hands like clouds exercise performed with both hands.	47

3.6	Sammon projection of the data after Feature Selection for the Wave the hands like clouds exercise performed with both hands.	48
3.7	Confusion Matrices presenting the classification results for the Wave the hands like clouds exercise performed with both hands.	48
3.8	Optimization of the number of features parameter for the Wave the hands like clouds exercise performed with one hand.	50
3.9	Sammon projection of the data after Feature Selection for the Wave the hands like clouds exercise performed with one hand.	51
3.10	Confusion Matrices presenting the classification results for the Wave the hands like clouds exercise performed with one hand.	51
3.11	Optimization of the number of features parameter for the Withdraw and push exercise.	53
3.12	Sammon projection of the data after Feature Selection for the Withdraw and push exercise.	54
3.13	Confusion Matrices presenting the classification results for the Withdraw and push exercise.	55
3.14	Optimization of the number of features parameter for the Brush knee twist step exercise.	56
3.15	Sammon projection of the data after Feature Selection for the Brush knee twist step exercise.	57
3.16	Confusion Matrices presenting the classification results for the Brush knee twist step exercise.	58
3.17	Optimization of the number of features parameter for the Golden rooster exercise.	59
3.18	Sammon projection of the data after Feature Selection for the Golden rooster exercise.	60
3.19	Confusion Matrices presenting the classification results for the Golden rooster exercise.	61
3.20	Optimization of the number of features parameter for the Raising the power exercise.	62
3.21	Confusion Matrices presenting the classification results for the Raising the power exercise.	63
3.22	Overview of the classifiers' performance in assessing the proficiency level of Tai Chi practitioners measured with the F1 micro score. . .	64

The Motion Analysis Lab

This Thesis was carried out in Harvard Medical School's Motion Analysis Lab at Spaulding Rehabilitation Hospital in Boston, MA (USA). The Motion Analysis Lab focuses on analysing the biomechanics of human movement through robotics and wearable technology with the goal of studying and treating mobility-limiting conditions, including cerebral palsy, stroke, and Parkinson's Disease.

During this time, I was encouraged to explore several projects that are being pursued in the lab.

Muscle synergy projects

I participated in the feasibility studies, experimental design, and preliminary dry runs of three projects related to muscle synergies. A muscle synergy can be defined as a cluster of muscles that activate simultaneously with consistent proportional increases, thereby representing a muscle activation pattern with stable spatial and temporal features [1].

Motor Adaptation with Exoskeleton

Motor adaptation is the ability of humans to modify their gait motor patterns following a change in the environment, and it can be studied by analyzing the change in muscle synergies during perturbations [2]. Since the modification of gait patterns is an important goal of physical therapy, the study of motor adaptation has significant potential for assessing the ability of patients to modify their motor patterns when walking [3].

Cajigas et al. (2017) [3] conducted a study on the effects of mechanical perturbations produced using an exoskeleton system for treadmill-based gait rehabilitation (Lokomat by Hocoma AG, Zurich, Switzerland [4]), that allows the control of the subject's hip and knee flexion and extension. This study focused on the characterization of motor adaptation of healthy subjects whose gait cycle was perturbed during the swing phase. The perturbation was produced by generating torques

using the actuators at the hip and knee joints of the robotic leg. Despite the hypothesis that motor adaptation would have been observed for all the orientations of the perturbation force vector, results of the study revealed instead a selective process of generation of motor adaptation. More specifically, the study participants failed to adapt to step-height perturbations but strongly adapted to step-length perturbations .

Since the Lokomat system does not allow the displacement of the center of mass by constraining the movement of the pelvis, the motor adaptation project I contributed to aims at verifying if, with a different system that does not impose vertical constraints, vertical perturbations of the subject's stride induce a phenomenon of motor adaptation. To this goal, the lab developed its own Exoskeleton. My contributions to this project was mainly the provision of assistance during the dry runs together with my participation in the Exoskeleton's maintenance.

Balance Project

Falls are the second most frequent cause of injury and among the most frequent causes of disability in the elderly, which makes the study of mechanisms of maintaining balance critical [5]. A study by Wojtara et al. (2014) [6] highlighted the possibility of using muscle synergies to assess the balance of subjects. In fact, they showed an association between the lack of consistency in the muscle synergies of participants when their balance was perturbed and the increase in the fall risk . These results were further supported in a study review by Rubega et al. (2021) [5], who pointed out how muscle synergies of the elderly are altered, especially when they are "frequent fallers".

The goal of the balance project is, hence, to study muscle synergies in the condition of weak balance perturbation. To do so, study participants are perturbed through pulls of variable intensity while EMG, motion capture, and force data are collected to record - respectively - the muscle activations, the 3D position of the subject, the Center of Pressure, and the intensity of the pull.

The preliminary study I participated in does not involve the presence of the aforementioned Exoskeleton, but the future plan involves analyzing how the exoskeleton could be employed to help the subject maintain balance. For this project, I played a role in the experimental design and in the preliminary dry runs with and without the Exoskeleton system.

Scaling Stroke

Muscle synergies can be employed to assess the severity of the functional impairment after a stroke [7], which is important to allow the physical therapist to make accurate decisions during the rehabilitation treatment and thus increase the benefit for the

patient. Detecting muscle synergies, though, is time-consuming and expensive as it requires the placement of several EMG electrodes over the examined limb. Hence, the aim of Scaling Stroke is to find an alternative way of assessing the severity of impairment after cortical damage by looking at the kinematics of movement elements. In fact, a study by Miranda et al. (2018) [8] showed how complex upper-limb movement patterns can be modeled through movement elements whose theoretical velocity profile is bell-shaped. The aim of the Scaling Stroke project is thus to verify whether a relation exists between the correlation of the actual velocity profile of movements with the theoretical profile and the muscle synergies patterns during upper limb movements. This would make the measurement of movement kinematics a proxy for muscle synergy analysis in the context of assessing functional impairment after a stroke episode. To this goal, biomechanic and EMG data are collected from stroke survivors who are asked to draw figures (e.g., circles, spirals, ellipses, ...) and perform several tasks with both the affected and the unaffected arm. Motion capture with the Vicon system (Vicon Industries Inc. [9]) provides the kinematic data, while the EMG signal of 16 upper-limb muscles allows the detection and analysis of the muscle activation. For this project, I provided assistance during the preliminary dry runs.

RehabPAL

RehabPAL aims at assessing the increase of the engagement of children with Cerebral Palsy performing physical exercise when playing with the NAO robot (SoftBank Robotics Group [10]).

Cerebral Palsy is a collection of conditions characterized by limitations in movement and posture development due to disturbances that occur in the fetal or infant brain during development. Conditions common to people suffering from Cerebral Palsy include issues related to sensation, cognition, communication, perception, and behavior, as well as bronchopulmonary dysplasia, spasticity, decreased weight bearing, and reduction of coordination secondary to motor-cortex involvement [11]. These conditions affect the subject's ability to participate in physical activity, which is important to decrease the risk for metabolic and cardiovascular diseases, especially in the case of Cerebral Palsy patients, who have lower levels of muscle strength and cardiorespiratory endurance than healthy subjects [12].

To increase the engagement of children suffering from Cerebral Palsy when performing physical exercise, the RehabPAL project exploits the context of a challenge between the kid and the NAO robot. A range of exercises can be selected by the physical therapist and proposed to the child through a Graphical User Interface (GUI). Human Pose Estimation is employed to assess the performance of the child when performing the exercise, and a score is assigned to both the kid and

NAO to determine the winner of the challenge. To verify if this solution leads to an increase in the child's engagement, the data collection is performed when the kid plays with NAO, with a virtual version of NAO that appears on the GUI and without any physical or virtual companion, and a questionnaire is presented to the kid at the end of each study phase.

My contributions to this project lie both in the setup of the study and in the data collection carried out with the kids.

Dephy

The aim of the Dephy project is to enhance rehabilitation therapy by using an orthosis that can adapt to stroke survivors' needs. Such a device is the Dephy Exoboot (Dephy, Inc. [13]), which exists either in a passive or in an active version.

To adapt to the patient's needs, the system allows to set the resistance to dorsiflexion and plantarflexion and the neutral angle. To verify the enhancement of the gait of the subject, we collected motion capture data with the Vicon system (Vicon Industries, Inc. [9]), ground reaction force data with the AMTI force plates (Advanced Mechanical Technology Inc. [14]), accelerometer, gyroscope and magnetometer data with the XSens system (Xsens Technologies BV, now part of Movella Holdings Inc. [15]) and video data with GoPros in different conditions:

1. Habitual condition: the subject wears his/her own orthosis;
2. Simulated condition: the passive Exoboot orthosis parameters are set to mimic those of the subject's orthosis. These parameters are quantified through a rig that allows the mechanical characterization of the study participant's device in terms of stiffness in dorsiflexion and plantarflexion and through visual inspection for the neutral angle;
3. Optimal condition: the parameters of the passive Exoboot orthosis are set to be the ones that mostly enhance the gait of the subject. These parameters are set following the physical therapist's observations;
4. Active condition: the subject wears the active Exoboot, that is able to exert an active force during the study participant's gait.

The kinematic and kinetic data is finally analyzed to assess the quality of gait of the subject in different conditions.

My role in the Dephy project has been involvement during the data collection with stroke survivors.

SwanBio

SwanBio focuses on the balance of patients suffering from Adrenomyeloneuropathy, which is the most prevalent phenotype of X-linked Adrenoleukodystrophy. Adrenomyeloneuropathy is a neurogenetic disorder of the spinal cord that primarily impacts the corticospinal tract and the dorsal columns. The majority of individuals with Adrenomyeloneuropathy exhibit noticeable spasticity, ataxia, and muscle weakness, resulting in challenges related to walking and balance. These difficulties can significantly impact the patients' overall quality of life [16].

To characterize balance, the data of interest to the SwanBio project are the average swing of the center of pressure in the vertical and horizontal directions and the total path of the Center Of Pressure. Measurements were taken in different facilities with two different force plates: the AMTI (Advanced Mechanical Technology Inc. [14]) and the Kistler (Kistler Instrumente AG, Winterthur, Switzerland [17]). My contribution to the project was building a common pipeline to process the data acquired by the two systems.

Other contributions

To conclude, I offered minor contributions to other projects run in the Motion Analysis Lab.

Firstly, I was involved in the test of a sensorized insole provided by an external company. The insole was composed of 18 sensors that measure the kinetics of gait, acting as a portable force plate. The system allows monitoring the gait of the subject through an app and storing both the raw and the processed data.

Finally, I had the chance to learn how to use the Solidworks 3D CAD Design Software to design and print objects required for different studies in the lab. I was asked to design and print a case to store and ship size samples of ring sensors needed by subjects enrolled in a lab's study and a "wand" that could hold reflective markers on the heel of kids performing gait evaluations.

Introduction

Healthy ageing is defined by the World Health Organization as "the process of developing and maintaining the functional ability that enables wellbeing in older age" [18]. As the global population is rapidly ageing, the concept of healthy ageing is getting more and more important to ensure dignity and equality for all people. To achieve this purpose, the World Health Organization has determined ten objectives for the Decade of Healthy Ageing (2020-2030), one of which is the promotion of research that tackles the current and future needs of the elderly population [18].

In the elderly, falls are dangerous events that can lead to fractures, residual disability, chronic pain, and loss of independence. It is estimated that, for older individuals, falls account for 40% of all injury-related deaths, and nearly one out of three elderly individuals experience a fall annually [19, 20]. Inactivity is recognized as a contributing factor to a progressive physical impairment in the elderly that can increase the risk of falls. Indeed, a systematic review and meta-analysis by Papalia et al. (2020) [19] emphasized the idea that physical exercise can significantly reduce the risk of falls among the elderly, helping to preserve muscle mass and enhance balance control. At the same time, physical exercise helps prevent various age-related health issues, including metabolic disorders, cardiovascular diseases, cancer, and diminished bone quality.

In this context, several studies have pointed out the positive impact of the practice of Tai Chi Chuan on balance [20, 21]. In this context, this Thesis deals with the development of Machine-Learning algorithms that automatically assess the proficiency score of older adults performing Tai Chi exercises from video data, whose information is extracted by means of Human Pose Estimation. The developed models will enable further investigation of the relationship between Tai Chi proficiency and increase in balance.

The present manuscript is organized as follows. The first Chapter provides a brief overview of the Tai Chi literature and its beneficial effects on health and of the state of the art methods for Human Pose Estimation. In the second Chapter the Materials and Methods employed in this work are presented. This Chapter is divided into three sections:

1. A presentation of the procedure employed to collect the data, together with a description of the selected Tai Chi exercises and the scores assigned to the subjects;
2. The data processing pipeline, that includes the skeleton tracking with MediaPipe, the normalization of the skeletal dimensions of the study participants, the automatic segmentation, and the procedures to construct and extract the features;
3. The steps followed to develop each classification model including the feature selection, the data projections, the training of the model, and the estimation of the final proficiency level.

The third Chapter presents the results for each of the analyzed exercises. Finally, the fourth and the fifth Chapters respectively focus on the Discussion of the results obtained, including limitations and possible future work, and on the Conclusions of the study.

It is worth mentioning that the results of this Thesis have been showcased at the IEEE-EMBS Body Sensor Networks International Conference held in Boston in October 2023 in a poster presentation.

1.1 State of the art

This section explores the studies and technologies that set the ground for this Thesis. The three main areas that will be covered are Tai Chi and its beneficial effects on health, Human Pose Estimation and an overview of the available skeleton tracking solutions.

1.1.1 Health benefits of Tai Chi Chuan

Tai Chi Chuan is a form of mind-body exercise that originated in the 17th century and that is rooted in the Taoist philosophy, the holistic vision of Chinese medicine, and in the Yin-Yang principle expression [22, 23]. The low-impact exercises of Tai Chi are characterized by slow, continuous, and smooth movements that aim at inducing a state of mind tranquillity similar to what one could experience through meditation [20, 21].

A study from Wayne et al. (2013) [22] deconstructed the multiple potentially therapeutic components of Tai Chi. These components, which can work synergistically or independently, are:

- Awareness, mindfulness, and focused attention: the slow, deliberate movements of Tai Chi and its emphasis on mindfulness of breath, body posture, and

sensory experiences cultivate heightened self-awareness, increase mindfulness and enhance concentration.

- Intention, belief, and expectation: the use of imagery, visualization, and similar cognitive techniques can change one's intent, beliefs, and expectations, potentially playing a substantial role in the therapeutic and physiological benefits of Tai Chi.
- Structural integration, dynamic form, and function: improved coordination and interaction among various structural and physiological systems together with Tai Chi's biomechanically efficient forms and movement patterns are significant elements that could explain its therapeutic impact.
- Active relaxation: the continuous, graceful movements of Tai Chi promote a profound state of relaxation for both the body and the mind, often regarded as a form of "meditation in motion".
- Strengthening and flexibility: Tai Chi offers a moderate level of aerobic conditioning. The coordinated movements of Tai Chi reduce strain, improve strength and balance, and enable greater power with less effort. The deliberate slowness of Tai Chi movements, combined with slightly bent stances and sustained single-leg weight-bearing, effectively provides significant strength training for the lower extremities. Furthermore, the slow, continuous and relaxed motions of Tai Chi incorporate dynamic stretching, which enhances overall flexibility.
- Natural, freer breathing: enhanced and more effective breathing aids in regulating the nervous system, uplifts mood, and is thought to promote the harmonious flow of energy both within the body and between the body and its surroundings.
- Social support, interaction, and community: participating in a group setting holds therapeutic benefits for a range of physical and psychological medical conditions.
- Embodied spirituality, philosophy, and ritual: Tai Chi offers a tangible framework for embracing a holistic Eastern philosophy that harmonizes the body, mind, and spirit in daily life.

According to a bibliometric analysis of published clinical studies between 2010 and 2020 conducted by Yang et al. (2021) [24], the top 10 conditions or diseases analyzed by clinical studies that exploit the practice of Tai Chi are: hypertension, chronic obstructive pulmonary disease, diabetes, knee osteoarthritis, heart failure, depression, osteoporosis and osteopenia, breast cancer, coronary heart disease, and

insomnia. Tai Chi practice can also benefit cardiorespiratory function, pulmonary function, strength and flexibility, balance and motor control, endothelial function and peripheral circulation, body mass index and blood lipid profile, thyroid and immune function, and psychosocial function [23, 25].

A summary of the conditions that benefit from the practice of Tai Chi together with the relative quality of the evidence supporting its positive effect, as reported in the analysis conducted by Houston et al. [26], is presented in Table 1.1.

Quality of evidence	Beneficial effect
Excellent	Preventing falls and improving balance, osteoarthritis, Parkinson disease, rehabilitation for chronic obstructive pulmonary disease, improving cognitive capacity in older adults and increasing aerobic capacity in those with poor fitness.
Good	Depression, cardiac and stroke rehabilitation, dementia and increasing of strength in the lower limbs.
Fair	Improving quality of life for cancer patients, fibromyalgia, hypertension, and osteoporosis, enhancing well-being and improving sleep.

Table 1.1: Evidence of the benefit of the conditions analyzed by Tai Chi clinical studies [26].

Tai Chi and balance

Several studies have analyzed the positive effect of Tai Chi on balance and have found several reasons that explain such benefit [20, 21, 27]:

- Tai Chi training increases trunk control, movements important for gait, flexibility, core stability and muscle strength, especially in the lower extremities, as it requires maintaining a semi-squat position;
- The practice of Tai Chi increases somatic sensation, and coordination of the internal and external space, enhancing the organization of the proprioceptive, visual and vestibular systems;
- Tai Chi requires shifting the weight, swaying the ankles, and stepping forward and backward, hence leading to a positive effect on the ranges of dorsiflexion and plantarflexion of the practitioner;
- Tai Chi exercises strengthen the sensory-motor system and the knee extensor strength;

- Tai Chi can mitigate the balance dysfunction resulting from different conditions, including Parkinson’s disease and stroke.

All these factors point to the idea that the discipline of Tai Chi holds several beneficial effects on health. Practicing this martial art is hence recommendable, especially for the elderly population, to promote healthy ageing.

1.1.2 Human Pose Estimation

Human Pose Estimation is defined as the process of extracting body configurations in images or videos by predicting the locations of the articulated joints of a human body, allowing for a markerless reconstruction of a human skeletal representation either in 2D or in 3D [28]. To do so, state-of-the-art Human Pose Estimation algorithms are trained on large image and/or video datasets where the joints of the individuals have been manually annotated, which makes them able to track unlabeled images or videos. Examples of such datasets are the COCO [29], the MPII Human Pose dataset [30], the Human3.6M [31] and the HumanEva [32]. Given its broad spectrum of possible applications, Human Pose Estimation is today a pivotal and continually advancing field within the domain of Computer Vision [28, 33, 34, 35].

Human Pose Estimation methods can exploit two different approaches: the top-down and the bottom-up. On one hand, the top-down approach begins by detecting each person in the frame and subsequently locates the joints for each individual. On the other, the bottom-up locates all joints first, and then it associates them to the respective subject in the frame [28]. Top-down methods have some major downfalls: firstly, if they do not find the individual in the frame they will not estimate its pose; secondly, their accuracy is at risk when multiple people are present in the frame, and lastly, their computational cost increases as the number of individuals in the frame increases as a single pose estimation is run for every subject detected. Bottom-ups overcome these limitations, but they can encounter difficulties in effectively grouping body parts when there is significant overlap among individuals [34].

As reported in the review by Desmarais et al. (2021) [33]. There are several learning approaches that train Human Pose Estimation algorithms to predict the pose of the subjects in the frame:

- Convolutional Neural Networks (CNN): CNNs are widely used for 3D Human Pose Estimation with various approaches that depend on the representation of the data. Usually, methods that rely on a single-view (monocular methods) use classic convolutions, while methods that take as input multiple views can exploit 3D convolution networks. In the case of video data, better pose

characterization can be achieved through temporal convolutions that use the information given by the past and future frames.

- **Recurrent Neural Networks (RNN):** RNN and, more specifically, Long Short-Term Memory (LSTM) architectures can extract 2D joint trajectories. LSTM networks are sequence-to-sequence models, i.e. they are able to output one sequence from another by passing successively the information stored from the previous inputs.
- **Graph Neural Networks (GNN):** GNN can compute spatial-temporal graphs that encode the pose location and the trajectory in time.
- **Attention mechanism:** the attention mechanism trains the models to detect the information that is most relevant to pose estimation.
- **Adversarial Networks (adversarial learning):** in Adversarial Networks, a discriminator network enhances pose consistency by learning to distinguish between generated poses and those extracted directly from ground-truth data. An adversarial loss component is subsequently incorporated into the generator network, which is responsible for estimating 3D human poses from visual or 2D pose information. As a result, any poses that deviate from established configurations are subject to a penalty.

These training layers part of full backbone architectures that set the basis for pre-trained networks. The characteristics of a selection of such architectures are presented in Table 1.2 [33].

1.1.3 Pre-trained Human Pose Estimation platforms

Several networks have been pre-trained and made available for use. This section presents a non-exhaustive overview of a selection of pre-trained models.

DeepPose (2014)

DeepPose is a Human Pose Estimation framework developed in 2014 by Google researchers [36], whose work introduced the use of Deep Neural Networks for Human Pose Estimation. The DeepPose approach involves framing the problem as a regression task aimed at predicting joint coordinates by utilizing the entire image as input and passing it through a seven-layer generic Convolutional Neural Network. This approach offered two distinct advantages. Firstly, the Deep Neural Network has the capability to encompass the complete context surrounding each body joint. In other words, each joint regressor leverages the entire image as a signal, providing a comprehensive understanding of the joint’s surroundings.

Architecture	Characteristics
Hourglass networks	Hourglass networks consist of a stacked hourglass module interspersed with intermediate supervision. The goal of hourglass networks is to encompass both global and local discriminative features by computing image features at different resolution levels. Hence, each module firstly decreases the resolution through convolutional layers and subsequently up-samples the features while adding them to the corresponding ones of the same dimensions from the preceding stage.
Cascaded pyramid networks	Similarly to hourglass networks, cascaded pyramid networks aim at encompassing both global and local discriminative features by computing image features at different resolution levels. Cascaded pyramid networks follow a two-step architecture for pose prediction. Firstly, poses are predicted using a feature pyramid network, and secondly, results are refined, specifically for keypoints deemed hard to predict by the loss computed in the pyramid network.
High resolution networks	Once again, high resolution networks work with the same image at different resolutions, that are processed simultaneously through convolutional layers that share weights by means of exchange blocks.
Part affinity fields	Part affinity fields serve as a non-parametric representation of the relationships between various body parts. Leveraging these features alongside confidence maps for joint localization, this method accurately predicts human poses while ensuring proper associations among multiple subjects. Additionally, a notable advantage of this approach is its ability to operate in real-time.
Residual networks	Residual networks employ deconvolution layers to generate heatmaps from deep image features, without employing a dedicated procedure for handling hard-to-predict keypoints. Despite its simplicity, this model delivers competitive results while remaining computationally efficient.

Table 1.2: Backbone architectures for Human Pose Estimation models. [33].

Secondly, this methodology is considerably simpler to formulate in comparison to techniques that rely on graphical models. There is no need for explicit design of feature representations and detectors for individual body parts, nor is there a requirement to explicitly engineer a model’s structure and the interactions between different joints as a Convolutional Neural Network is trained for this specific task. Furthermore, to enhance the precision of joint localization, a cascade of Deep Neural Network-based pose predictors was employed as a refinement step: starting with an initial pose estimation derived from the full image, Deep Neural Network-based regressors further refine the joint predictions by utilizing higher-resolution sub-images [36].

The cascade of regression models offered a distinct advantage by enabling the comprehensive incorporation of context and reasoning in the pose estimation process. Consequently, DeepPose achieved superior performance on numerous demanding academic datasets with respect to the previous Human Pose Estimation methods (average Percentage of Correct Parts on the Leeds Sports Dataset [37]: 69%) [36].

Deepercut (2016)

Deepercut [38] marked a significant milestone as the first bottom-up model to truly compete with top-down models successfully. It introduced image-conditioned pairwise terms, enabling the generation of multiple body part configurations and employed an incremental optimization strategy, which efficiently explored the search space, leading to improved accuracy and speed. DeeperCut represented an enhancement of the previous state-of-the-art method, Deepcut [39], in three key aspects: firstly, it incorporated deeper neural networks, taking advantage of advances in Deep Learning from 2016; secondly, it exhibited increased strength through the utilization of novel image-conditioned pairwise terms for body parts; lastly, it achieved higher speed through the combination of these innovations [38]. During its time, DeeperCut delivered the best accuracy, surpassing the previously leading results in multi-person pose estimation. Nevertheless, it did encounter challenges, particularly in cases involving the simultaneous estimation of body articulations for multiple individuals [40].

DeeperCut has been employed over the years to develop systems with a broad range of applications. For example, Saint et al. (2017) [41] used DeeperCut to build an approach for autonomously reconstructing a lifelike and precise representation of an individual’s body shape while clothed, using a 3D scan. One year later, Wei et al. (2018) [42] employed DeeperCut in the context of public security for person and vehicle re-identification. As a last example, Liang et al. (2020) [43] exploited the ability of DeeperCut to estimate individuals’ pose to design a robot that would follow elderly adults to promptly send alerts to family or hospitals in case of fall.

Regional Multi-Person Pose Estimation (AlphaPose, 2016)

Fang et al. (2016) [44] introduced AlphaPose, a method known for enhancing pose estimation, even when faced with imprecise human bounding boxes, through a Regional Multi-person Pose Estimation (RMPE) framework. AlphaPose has three main components:

1. A Symmetric Spatial Transformer Network (SSTN) is attached to a parallel Single Person Pose Estimator (SPPE) that allows the extraction of a high-quality single-person bounding box even starting from an inaccurate.
2. A Parametric Pose Non-Maximum-Suppression (NMS) eliminates redundant poses by comparing pose similarity through a pose distance metric and allows the optimization of the pose distance parameters.
3. A Pose-Guided Proposals Generator (PGPG) that is able to simulate human bounding boxes generation and hence allows for augmentation of the training samples.

This framework has been validated on the MPII dataset [30] and results showed that it outperformed the state-of-the-art methods of its time by achieving a 76.7 mean average precision. However, it's important to note that, in exchange for its superior accuracy, this approach tends to be slower compared to other methods with a frame rate of 5 frames per second [40].

AlphaPose has been employed for the development of several fall detection systems, such as the ones developed by Ramirez et al. (2021) [45], by Zheng et al. (2022) [46], by Ma et al. (2022) [47], by Zhao et al. (2022) [48] and by Inturi et al. (2023) [49]. Other contexts in which AlphaPose was applied include: behavior detection (Tian et al. 2021 [50]), close proximity human-robot interaction (Docekal et al. 2022 [51]) and analysis of gait for patients suffering from knee arthritis (Lv et al. 2022 [52]).

OpenPose (2017)

OpenPose, developed by Cao et al. (2017) [53] is an open-source model using a bottom-up approach and it represents the first real-time attempt at estimating multi-person poses. OpenPose also pioneered the incorporation of association scores through Part Affinity Fields (PAFs), that convey the positions and orientations of limbs across the image space. The concurrent estimation of these bottom-up representations, encompassing both detection and association, effectively captures the global context. This allows a straightforward parsing approach to deliver high-quality outcomes, all while significantly reducing computational costs. OpenPose utilizes a feed-forward network to predict simultaneously both 2D confidence maps

for body part locations and 2D vector fields for part affinities. The association of each body part with a person is achieved through Non-Maximum Suppression, making use of the 2D vector fields that encode the location and orientation of limbs in the image. This innovative approach results in improved overall quality and robustness, especially when dealing with occluded body parts.

OpenPose was exploited for a very broad range of applications, including: gait analysis (Viswakumar et al. 2019 [54] and D’Antonio et al. 2020 [55]), fall detection (Chen et al. 2020 [56] and Lin et al. 2020 [57]), activity recognition (Noori et al. 2019 [58]), ergonomic postural assessment (Kim et al. 2021 [59]), sitting posture recognition (Chen 2019 [60]) and balance assessment (Li et al. 2021 [61]). OpenPose has also been employed in the context of sports, such as to measure the vertical height of jumps (Webering et al. 2021 [62]) or to estimate the pose of athletes performing a baseball swing (Li et al. 2021 [63]).

MediaPipe Pose (2019)

Mediapipe, developed by Google in 2019, is a framework that tackles several Computer Vision tasks, including Human Pose Estimation. MediaPipe comprises three primary components: a framework designed for drawing inferences from sensory data, a suite of tools for assessing performance, and a repository of reusable inference and processing modules [64, 65].

In the context of Human Pose Estimation, MediaPipe Pose, based on the BlazePose pipeline and the ML Kit Pose Detection Application Programming Interface [66], allows to predict the pose of a single individual per frame with a 30 frames per second frame rate, which allows real-time skeleton tracking. To do so, MediaPipe’s pipeline comprises a lightweight body pose detection system, which is subsequently followed by a pose tracking network. The tracker forecasts keypoint coordinates, as well as the person’s presence within the current frame through a face detector, and refines the region of interest accordingly. The pose estimation module of the framework employs an encoder-decoder network architecture that involves heatmaps, offset losses and regression techniques. Skip connections are employed throughout all network stages to maintain a balance between high-level and low-level features, ensuring that gradients from the regression encoder do not propagate back to the heatmap-trained features. This approach has proven to not only enhance heatmap predictions but also significantly improve the accuracy of coordinate regression [67]. Designed primarily for fitness-related applications, Mediapipe has been optimized for multi-platform use, including mobile phones and laptops [66]. However, one limitation is that the estimation process requires the person’s head to always remain visible for accurate results.

In the landscape of fitness-related applications, MediaPipe enabled the development of systems for both general fitness training and tracking (Nath et al. 2023

[68]) and more specific disciplines. For example, this framework has been exploited in the context of Yoga practice, for which MediaPipe was employed in training (Agarwal et al. 2022 [69]), monitoring (Anilkumar et al. 2021 [70]) and pose classification (Garg et al. 2022 [71]) systems. MediaPipe has also been used for applications that are not related to fitness, such as: fall-detection systems (Bugarin et al. 2022 [72]) biomechanical assessment (Lafayette et al. 2022 [73]), gait analysis (Uchida et al. 2023 [74]) and human telerehabilitation (Latreche et al. 2023 [75]).

High-Resolution Net (2020)

In 2020, Wang et al. [76] proposed a pioneering architectural concept known as High-Resolution Net, which maintains high-resolution representations throughout the entire network. This work was driven by the demand for high-resolution representations for position-sensitive information, as in the case of Human Pose Estimation. Nevertheless, many of the latest classification networks decrease the resolution of the representation prior to classification by progressively reducing the spatial dimensions of feature maps and connecting convolutional layers from high-resolution to low-resolution sequentially.

The High-Resolution Net approach commences with a high-resolution convolutional stream followed by a systematical integration of additional high-to-low resolution convolutional streams, all while connecting these multi-resolution streams in parallel. The resulting network comprises four stages, with each stage containing a variable number of streams corresponding to its stage number. Iterative multi-resolution fusions are achieved by consistently exchanging information among these parallel streams.

The high-resolution representations acquired through the High-Resolution Net exhibit both semantic robustness and spatial precision, owing to two key factors. Firstly, this approach connects high-to-low resolution convolutional streams in parallel, as opposed to a sequential series. Consequently, it preserves high resolution throughout the process rather than attempting to recover it from a lower resolution, resulting in potentially more spatially precise learned representations. Secondly, unlike many existing fusion methods that aggregate high-resolution low-level and high-level representations through upsampling low-resolution counterparts, multi-resolution fusions are continually performed. This iterative process enhances high-resolution representations with the assistance of low-resolution counterparts, and vice versa. Consequently, all multi-resolution representations, from high to low resolution, exhibit strong semantic characteristics [76].

MoveNet (2021)

Released in 2021, MoveNet is a real-time pose detection model designed by Google [77, 78] to detect 17 keypoints on a single person through heatmaps. MoveNet

comes in two variants: Lightning and Thunder. While Lightning sacrifices some accuracy compared to Thunder, it offers faster inference times. The architecture of MoveNet consists of a feature extractor and a set of prediction heads, that consist of four parts: the person center heatmap, keypoint regression field, person keypoint heatmap, and 2D per-keypoint offset field. These components work together to predict human keypoints using heatmaps [79].

MoveNet has been used in the context of fitness-related applications, such as for the development of a system that verifies the correctness of posture in the Yoga’s Sun Salutation routine (Girase et al. 2022 [80]) and of an app that performs real-time assessment of physical exercise (Cai, 2022 [81]). At the same time, MoveNet has been applied to other domains such as the detection of violent behaviours or physical bullying (Kozhamkulova et al. 2023 [82]).

OpenPifPaf (2021)

In 2021, Kreiss et al. [83] developed OpenPifPaf, an open-source library for multi-person Human Pose Estimation that aims at detecting, associating, and tracking semantic keypoints within video data even in challenging scenes. The bottom-up approach presented stands out for its efficiency, stable field representation, and impressive accuracy and performance that even surpasses top-down methods.

The OpenPifPaf model architecture comprises a shared base network, either ResNet or ShuffleNetV2, without max-pooling. The core networks of the framework are the Composite Intensity Field, which represents joint intensity, and the Composite Associations Field, which forms associations that allow to track poses. In addition to the Composite Fields, further networks can be incorporated; for example, pose tracking is facilitated by an additional head network that predicts the Temporal Composite Association Field.

OpenPifPaf has been employed in systems that concern pedestrians in public roads with several applications: presence and crossing detection in the context of autonomous driving cars (Belkada et al. 2021 [84] and Proll, 2022 [85]), behaviour prediction (Moseva et al. 2023 [86]) and social distancing (Abdulrahman et al. 2023 [87]).

Materials and Methods

This chapter describes the material and methods that brought to the results presented in this Thesis. Firstly, the data collection setup will be presented together with the Tai Chi exercises that were proposed to the participants and the scores that they were assigned, then the data analysis pipeline will be explored. This pipeline comprises the extraction of the joint trajectories by means of skeleton tracking, the normalization and segmentation of the trajectories, the extraction of the features of interest and their ranking and selection, and the development of the classification model. A summary of the whole pipeline is provided in Figure 2.1.

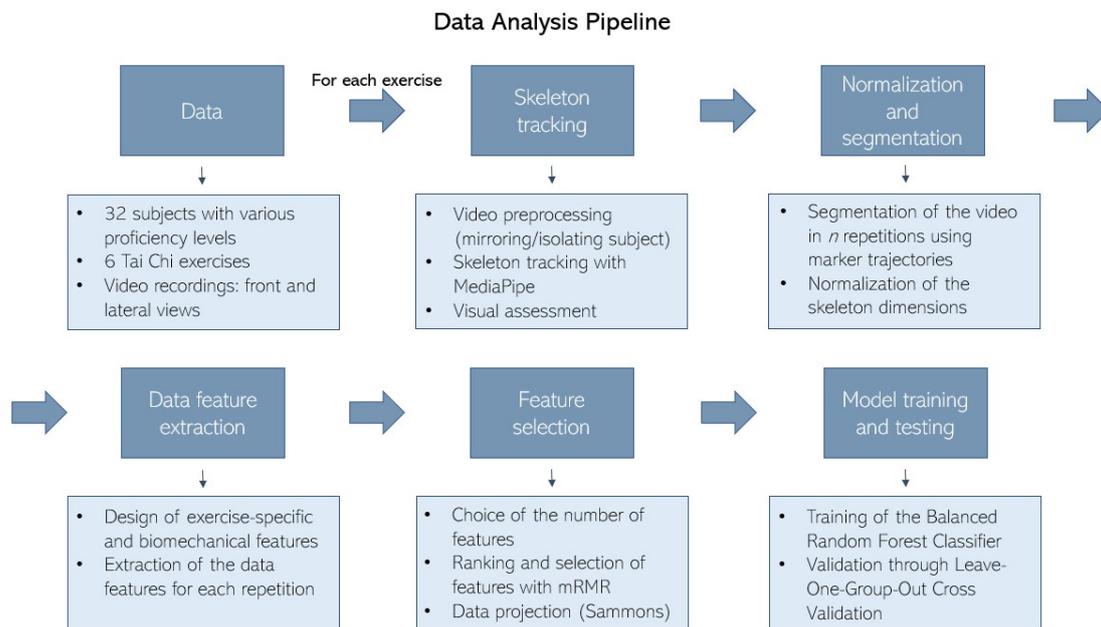


Figure 2.1: The data analysis pipeline

2.1 Data collection

Data were collected from 32 healthy older adults. No constraint was imposed on their previous Tai Chi experience with the goal of acquiring data for a range of proficiency levels.

The inclusion criteria for the study were:

- Subject aged from 60 to 80 years of age;
- Subject capable of ambulating without breaks for 15 minutes without an assistive device.

The exclusion criteria for the study were:

- Chronic neuromuscular conditions (e.g. Parkinson's disease, multiple sclerosis, peripheral neuropathy, stroke) that would prevent the subject from safely participating in the study;
- Acute medical conditions requiring hospitalization in within the 6 months previous to the study;
- Active cancer;
- Self-reported inability to walk unassisted for 15 minutes;
- Musculoskeletal conditions requiring chronic use of pain medications;
- Cognitive impairment (as measured by a Mini-Mental State Examination score < 24).

The recruited subjects were 20 females and 12 males, with a mean age of 70.2 years of age and a standard deviation of 4.88 years of age. In addition to the data of the study participants, video data of two Tai Chi experts was employed to augment the size of the dataset.

A total of six exercises were proposed to the subjects: Raising the power, Golden rooster, Withdraw and push, Grasp the sparrow's tail, Wave the hands like clouds, and Brush knee twist step. After watching a video of a Tai Chi expert performing the exercise, participants were asked to replicate the movements. The tutorial video could be played as many times as the subject desired before being asked to repeat the exercise. More specifically, subjects were asked to perform the movements for a total of 6 to 9 times depending on the exercise, as some of them were more challenging than others. Furthermore, with the exception of the Raising the power exercise, all the others were performed in more than one version: with the left or with the right leg forward. Three variations were performed for the Wave the

hands like clouds exercise: with the left hand moving, the right hand moving, and with both hands moving at the same time.

Video recordings were acquired through two GoPro cameras placed frontally and laterally to the subject. Accelerometer, gyroscope, magnetometer, and altimeter signals were acquired through third-generation Shimmer wearable sensors. Video data and sensor data were collected at the same time.

2.1.1 Tai Chi exercises

The movements of the six Tai Chi exercises chosen together with the Tai Chi Experts and proposed to the participants are described in this section [88, 89]. For what concerns those exercises that have more than one version (Goolden rooster, Grasp the sparrow tail, Brush knee twist step and Withdraw and Push), the left side only is described.

Grasp the sparrow's tail

In the Grasp the sparrow's tail exercise, emphasis is placed on the central role of the waist in coordinated movements. Rotating the waist while maintaining a fixed stance is beneficial for enhancing flexibility in the hips and lower spine, as well as promoting internal organ massage and integration of leg, torso, and arm movements. The exercise also involves performing large, circular arm movements to improve the range of motion, enhance circulation to the upper body, develop dynamic balance, and provide a moderate level of aerobic and strength training.

The exercise starts with a forward-weighted stance, with the left knee stable over the left foot and the torso facing forward (aligned with the front toes). The right foot is pointed 30 degrees to the right. Both palms are held in front of the chest, facing each other, resembling the act of grasping a small ball or bird, often referred to as "grasping a sparrow's tail."

The first step of the exercise is to shift the weight forward while the navel and the torso turn slightly to the left. While still facing to the left, the weight is shifted entirely onto the right leg, and then the waist is rotated to the right. While still facing to the right, the weight is transferred back to the left leg, returning to the forward bow stance, while the torso turns back to its original starting position with the navel pointing forward. For what concerns the arm movement, the starting position for the right arm is centered in front of the chest in a pushing motion. It should be coordinated with the turning of the torso to the left, avoiding any arm extension or changes in elbow angle. The right arm is then released downward as the weight shifts backward, creating a large circular motion. As the waist turns to the right, the right arm is circled back up to shoulder height, ensuring that the palm faces the same direction as the navel. Finally, the shoulders and wrist are

relaxed, returning the palm to the initial push gesture as the torso returns to the starting stance. The starting position of the left arm is a round shape at chest height, similar to cradling a large balloon or sphere against the chest. This gesture has to be maintained as hips turn to the left. As the weight is shifted back, the left arm relaxes downward in a circular motion, sweeping across the groin area. Then, the left arm swings back up to the same starting gesture, which remains consistent throughout the movement. [89]

The movements of the exercise in its left side version are shown in Figure 2.2.



Figure 2.2: The Grasp the sparrow's tail exercise.

Wave the hands like clouds

Wave the hands like clouds is a fundamental Tai Chi exercise that combines leg and waist movements with arm motions in a seamless flow. It employs circular movements to enhance blood circulation to the extremities while balancing both sides of the body.

The starting position of the exercise is with the feet parallel and shoulder-width apart. Weight shifts to the right leg, turning the waist slightly to the right, with attention to the knee, which should be stable and aligned with the foot without twisting. While maintaining a slight rightward orientation with the waist and the head, weight shifts to the left leg, aligning it with the left foot. The waist then turns slightly to the left without twisting the knee or torso. For what concerns the arms movements, hands should be envisioned as soft calligraphy paintbrushes. The right hand paints an oval that starts with a smooth brush stroke from right to left at navel height as the weight shifts from the right leg to the left leg. The arm movement is completed by a turn of the waist slightly to the left and the raise of the right hand to chest height without raising the right shoulder. As the weight shifts back to the right, a second brush stroke is created with the back of the right hand. The oval is finished by lowering the right hand to navel height on the right side. The same movement, but in the opposite direction, is performed with the left hand. To conclude the movement, hands are paused in front of the shoulder, turned to face downward, and allowed to gently float down. Throughout

the exercise, breathing should be deep, continuous, and relaxed.

The movements of the Wave the hands like clouds exercise are continuous, without abrupt starts or stops. Someone observing the practitioner should perceive a fluid, uninterrupted flow. [89] During the execution of this exercise, principles of balance and yin and yang become evident. The movements of the limbs are always synchronized but consistently opposite in terms of position and direction of the upper limbs, symbolizing balance and harmony throughout the performance [88].

This exercise was performed by the study participants with both hands simultaneously (Figure 2.3) and with one hand at a time (Figure 2.4). To allow for a more consistent analysis, these versions of the exercise have been analyzed separately.



Figure 2.3: The Wave the hands like clouds exercise performed with both hands.



Figure 2.4: The Wave the hands like clouds exercise performed with one hand.

Withdraw and push

The Withdraw and push exercise introduces the challenge of relying on one leg, the rear one, for support and balance while smoothly transferring weight from front to back. It coordinates simple leg and torso movements with continuous changes in arm positions, promoting dynamic balance, enhancing strength and flexibility in both the legs and arms, and providing a moderate level of aerobic activity.

The exercise starts with the feet aligned, shoulder-width apart, with the left foot in front. The left knee moves over the center of the left foot, ensuring it doesn't extend past the base of the toes, and about 60-70 percent of the weight

is distributed forward. The belly button, nose, and toes of the left foot should all point in the same direction. Hands raise to approximately chest height in a pushing-like gesture with an angle at the elbow joint not exceeding 90 degrees. Shoulders and arms are relaxed. Subsequently, a circular motion pattern involving the upper and lower body is initiated. 100 percent of the weight is shifted back onto the right leg while the arms relax forward and downward in a circular movement. The angle at the elbow joint will become more pronounced. Then, following the circular and descending movement of the arms, they are raised in front of the body as weight shifts forward until the arms reach the original pushing position. The exercise ends by letting the arms fall to the sides while the weight shifts backward and the front foot slides back to be parallel to the back foot at shoulder width [89]. The movements of the exercise in its left side version are presented in Figure 2.5.

Throughout the exercise, breathing should be kept deep, slow, and comfortable without imposing any particular rhythm or pattern [89]. From a choreographic perspective, participants generally did not find this exercise to be overly challenging despite the requirement of a high degree of fluidity in movements and attention to maintaining the alignment of the torso over the lower limbs throughout the weight transfer [88].



Figure 2.5: The Withdraw and push exercise.

Brush knee twist step

The Brush knee twist step exercise aims to develop the ability to take slow and deliberate forward steps while managing the challenging balance demands created by constantly changing arm positions. This exercise involves coordinating both leg and arm movements, facilitating the integration of the upper and lower body.

The performance starts with the right leg in the back and the left leg bearing the weight in the front. The left hand is placed close to the left thigh, while the right hand is positioned in front of the right shoulder with the fingers pointing upward, creating a pushing gesture. Now, the whole body moves forward, balancing on the left leg by bringing the right knee alongside the left knee while slightly lifting the toes of the right foot off the ground. Subsequently, the right foot returns to

its original position, and as all of the weight is shifted back to the right leg, the waist is turned to the right. The left toes move alongside the right ones, arms are released, and the right arm moves in a circular motion [89]. The movements of the exercise in its left side version are presented in Figure 2.6.

It’s worth noting that participants and the Tai Chi experts identified this exercise as the most challenging within the protocol due to its intricate choreography, the requirement for synchronized movements, balance, and the fluid transition between successive actions [88].



Figure 2.6: The Brush knee twist step exercise.

Golden rooster

The Golden rooster exercise starts with the subject standing with the feet close together and the arms relaxed alongside the body. The left leg moves forward, and the whole body follows until the right leg moves forward as well. As soon as the two feet come close gain, the right leg and right hand are raised simultaneously while the left hand descends, with the palm facing backward. The practitioner briefly stands on one leg only before lowering the raised knee. As the limb lowers, the right hand sweeps in a circular motion and moves towards the right thigh. During this motion, the right elbow should not move. As the hands are both lowered, the weight is shifted backward. The exercise ends by moving the left leg backward so that the feet are once again next to each other. The movements of the exercise in its left side version are presented in Figure 2.7.

In the execution of this exercise, the concept of maintaining proper body alignment is of utmost importance to preserve balance when lifting the leg [88].

Raising the power

The Raising the power exercise harmonizes the movements of the upper and lower body by coordinating simple motions of the arms and legs. It also enhances the strength and flexibility of the ankles, knees, hips, and back. Moreover, it presents a balance challenge by continually changing the position of the arms relative to the



Figure 2.7: The Golden rooster exercise.

torso while also drawing attention to the energy differences between the left and right sides of the body.

The exercise starts with the feet shoulder-width apart. Knees are bent slowly and gently, not exceeding a 10 percent bend, simultaneously allowing the hands to rise in front of the body. Wrists are bent and relaxed, with the hands approximately shoulder-width apart and the fingers hanging downward. As the wrists reach shoulder height, the palms and then the fingers are gradually opened. Subsequently, elbows descend, permitting the relaxed wrists to float down the front of the body as they lower. Simultaneously, legs are slowly straightened [89]. The movements of the exercise are presented in Figure 2.8.

An essential aspect of this exercise involves maintaining proper alignment, ensuring that the head remains centered over the torso and the torso stays aligned with the lower limbs during both the descent and ascent phases [88].



Figure 2.8: The Raising the power exercise.

2.1.2 Scores

Five different metrics were used to evaluate the performance of the participants. Tai Chi experts assigned Low, Medium, or High scores for each of the five metrics to each subject for each exercise. Notably, a single value for each metric was assigned concurrently to all repetitions and variations of the same exercise, taking into account the overall performance and the improvement over repetitions. The

five metrics are:

- **Gross competency:** the gross competency score assesses the subject’s capability to accurately execute the choreography of the proposed exercise.
- **Expression of the Yin-Yang principle:** the Yin-Yang philosophy is at the core of the Tai Chi art; hence, effectively embodying this principle within the dynamics of execution is crucial to proficiency. Expression of the Yin-Yang can be detected through actions like shifting body weight or synchronizing breath with movement.
- **Alignment and posture:** this score aims at evaluating proper body segment alignment throughout the movement, with a particular focus on verticality. The Tai Chi practitioner should aim at keeping the head centered over the torso, balancing the torso over the hips, aligning the hips over the base of support and positioning each knee joint over the central axis of the respective foot.
- **Flow and dynamic integration:** this criterion the three important factors of slowness, coordination, and the capacity to transition seamlessly from one position to the next. Ideally, the sequence of movements should convey a sense of unity and simultaneity, aiming to maintain balance to the greatest extent possible.
- **Range of motion:** this score assesses the extent of movement in the waist and limbs, as well as the flexibility of the body.

Previous analyses [88, 90] pointed out how the Yin Yang, Alignment and posture, Flow and dynamic integration, and Range of motion scores all correlate positively with the Gross competency score. Hence, the presented analysis only focuses on the estimation of the Gross competency score as a proxy of the proficiency level in Tai Chi of the study participant.

2.2 Data processing

This section illustrates the steps that allowed to extract features that embody biomechanical characteristics of movement that relate to Tai Chi proficiency from the raw video data. After preprocessing, joint trajectories will be tracked with MediaPipe [64], normalized and segmented into single exercise repetitions to allow the extraction of the constructed features.

2.2.1 Skeleton tracking with MediaPipe

The first step for enabling good-quality skeleton tracking was to preprocess recorded videos to crop additional people present in the frame, if any. Indeed, whenever the study participant did not seem sufficiently stable, physical therapists would get closer to provide assistance. Since the chosen pre-trained Human Pose Estimation method, MediaPipe [64], has been developed to track the skeleton of only one individual in the frame, its performances decrease greatly whenever more than one person is present in the video frame. Hence, any additional individual in the frame had to be covered before being able to launch the skeleton tracking. Examples of how MediaPipe's performances decrease when there are multiple individuals in the frame can be observed in Figure 2.9.

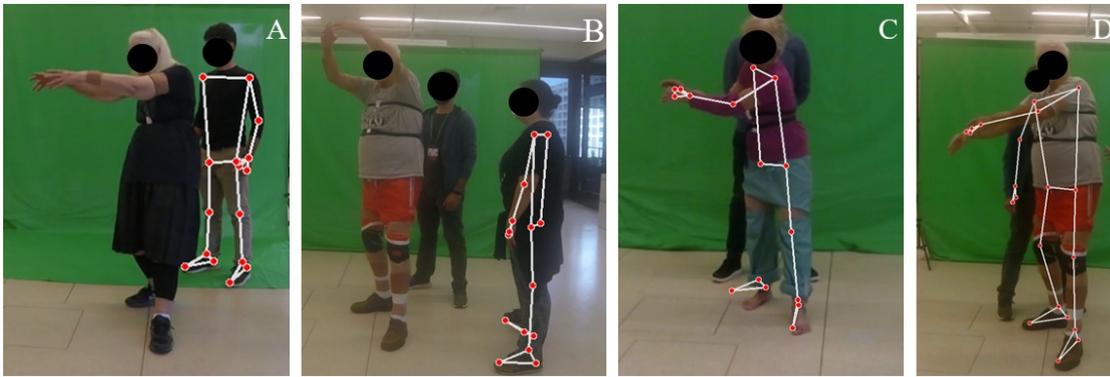


Figure 2.9: Examples of wrong person detected (A, B) or poor skeleton tracking quality (C, D) when multiple people are present in the frame.

Secondly, a subset of the videos was mirrored. This step was needed for different reasons depending on the exercise:

- For Grasp the sparrow's tail, Wave hands like clouds (performed with one hand only), Brush knee twist step and Golden Rooster, mirroring was needed as these exercises had two mirrored versions (referred to as the "right" and the "left" version in Section 2.1.1). Thus, "right" version videos were mirrored prior to skeleton tracking to allow for a more consistent extraction of the features (Figure 2.10.A, 2.10.B, 2.10.C);
- For Wave the hands like clouds (performed with both hands) and Raising the power, mirroring was performed to augment the size of the dataset. In fact, these exercises did not have a mirrored version, hence the amount of data available for their analysis was smaller compared to the other exercises (Figure 2.10.D, 2.10.E, 2.10.F).

The only exercise that was not mirrored at all was the Withdraw and Push exercise (more details in Section 2.1.1), which was analyzed by means of the lateral view. In this case, mirroring was not possible since, as described more in detail later in the Discussion (Chapter 4), cameras recording from the lateral view were not moved when the subject performed the two versions of the exercise. Hence, mirroring the right side video would have provided even more inconsistency to the analysis than not mirroring it, as shown in Figure 2.11.

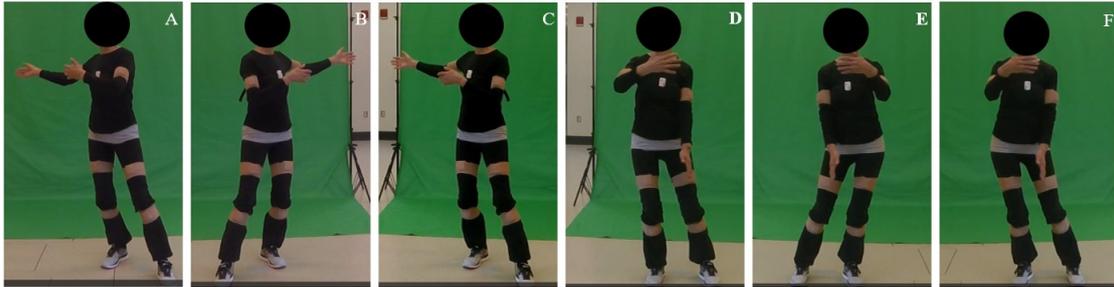


Figure 2.10: Examples of mirroring for one subject performing the Grasp the sparrow’s tail exercise. (A) shows the performance on the left side, (B) shows the performance on the right side, (C) shows the performance on the right side, mirrored to look like the left side. (D, E) show the unmirrored performance in different moments of the Wave the hands like clouds exercise, while (F) shows the mirrored movements at the same time as (E), demonstrating how mirroring this exercise does not impact the joint trajectories.

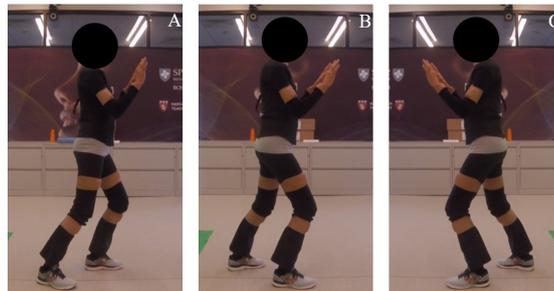


Figure 2.11: Examples of mirroring for one subject performing the Withdraw and Push. (A) shows the performance on the left side, (B) shows the performance on the right side, (C) shows the performance on the right side mirrored, demonstrating how for this exercise mirroring does not make sense.

The pose of the subject was estimated through MediaPipe [64]. This method has several characteristics that make it suitable to use in this work:

- As shown in Figure 2.12, MediaPipe allows tracking the joints of up to 33 keypoints of the human skeleton, which is more than sufficient for the purpose of the current analysis. The output of the skeleton tracking is the (x,y,z) coordinates of the joints, together with a visibility score. The (x,y) coordinates range between 0 and 1, as they are normalized with respect to the image dimensions, with the origin of the reference system located in the upper-left corner of the image. The z coordinate provides information on the depth of the joints relative to the midpoint of the hips, where a smaller value indicates that the landmark is closer to the camera. Nevertheless, the z coordinate was discarded from the analysis after assessing its low reliability. Lastly, the visibility score tells the user if the joint is present in the frame and not occluded; it ranges between 0 and 1, where a score of 1 means that the limb is perfectly visible [91].
- The videos analyzed in this Thesis were collected through GoPros, whose sampling frequency is 29.97 frames per second. MediaPipe's sampling frequency of 30 frames per second [67] is thus a crucial feature as it allows to track the coordinates of the joints for each frame of the videos.
- MediaPipe allows to set confidence scores for the pose detection and the pose tracking: the *min_detection_confidence* and the *min_tracking_confidence* [92]. These parameters range between 0 and 1 with a default value of 0.5; as their value increases, the velocity of the pose estimation decreases, but there's a gain in the confidence of the tracking. In this work, these parameters were both set to 0.9.
- As mentioned in Section 1.1.3, MediaPipe was designed primarily for fitness-related applications, and it has already been employed with success both in this context and in the domain of rehabilitation. Hence, this platform matches the requirements of the current study.
- MediaPipe is exceptionally user-friendly, as it only requires the installation of the library and meets straightforward software and hardware prerequisites, making it easily accessible for operation.

To conclude, MediaPipe's features match the requirements of the current work while being easy to install and run. At the same time, MediaPipe's sampling frequency of 30 frames per second [67] allows for tracking the joint trajectories of the subject in real time. This feature has not been exploited in this Thesis work but, together with MediaPipe's easiness of use and few hardware requirements, might be of interest for future applications of the study.

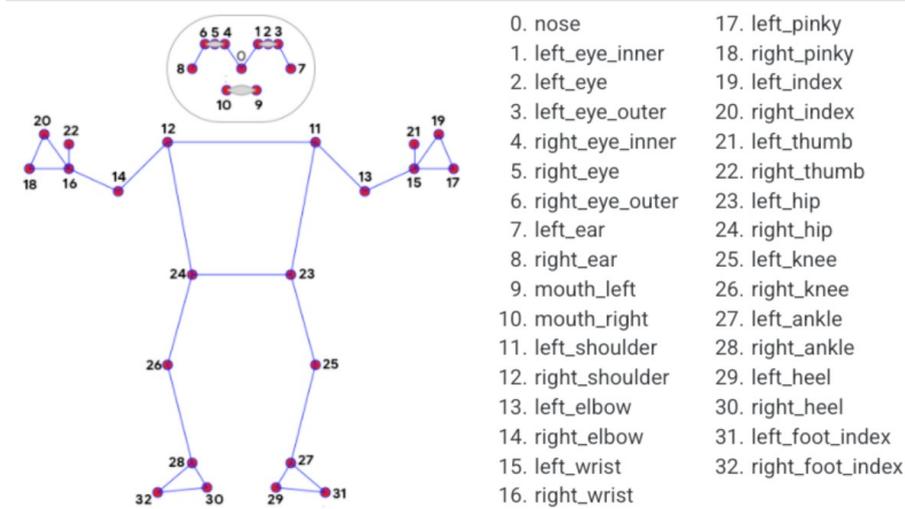


Figure 2.12: Human keypoints tracked by MediaPipe [91].

Assessment

After extracting the joint trajectories, circular markers have been juxtaposed to each frame of the video to allow for a visual assessment of the skeleton tracking quality. This assessment was particularly important for those videos that had been pre-processed by covering the presence of physical therapists in the frames (as discussed previously in this Chapter and shown in Figure 2.9) and for the videos recorded from a lateral view. In fact, the position of the camera led to an inevitable occlusion of part of the subject’s body. For this reason, the majority of the joints on the right side of the body recorded from the lateral view were excluded from the analysis. Examples of poor skeleton tracking quality due to the view are provided in Figure 2.13.

2.2.2 Skeleton dimensions normalization

To avoid the influence of the anthropometric measurements of the subjects on the analysis, skeletal dimensions were normalized with respect to the length of the trunk.

Literature search and conversations with the Motion Analysis Lab’s clinicians revealed that anthropometric scaling is usually performed by dividing the dimensions of the skeleton by the length of a limb or the dimensions of the head [93]. Due to the positioning of the cameras and the variable start of the recording, which did not provide a consistent view of the subject’s limbs at the beginning of the video, and the location of the keypoints extracted by MediaPipe, that would not allow to extract the length of the head, the choice of normalizing by the length of

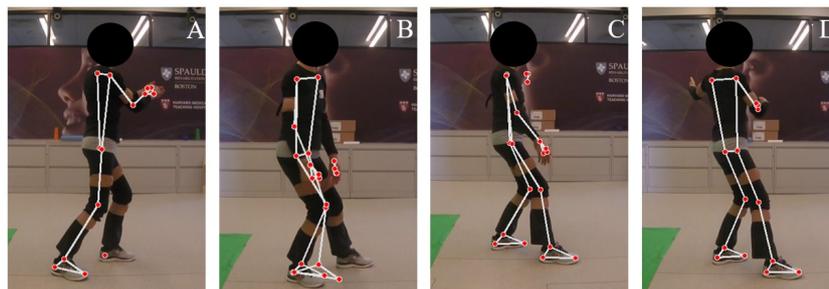


Figure 2.13: Examples of poor skeleton tracking from the lateral view videos. (A, B) show one subject performing Grasp the sparrow’s tail on the left side, (C, D) show one subject performing Grasp the sparrow’s tail on the right side. In both cases, there are instances of poor skeleton tracking quality.

the trunk was made. Furthermore, for every subject, the origin of the reference system was shifted to be the position of the left hip so that each subject would be placed at the center of the frame and all joint trajectories would be relative to the same point. Lastly, to ensure that the skeleton tracking was stable enough when extracting the trunk length, the coordinates of the joints that were employed in the normalization step were extracted at the two-hundredth frame (around 6.7 seconds after the start of the video).

2.2.3 Automatic segmentation

In a single video recording, the exercise was repeated several times by the subject. Despite having a single proficiency score per exercise per subject, it was chosen to segment each exercise execution in its repetitions to augment the dimensions of the dataset. Furthermore, for each exercise, moments of interest for the extraction of features were selected. The moments of interest to be detected for each exercise, in addition to the start and the stop of the repetitions, are:

- Grasp the sparrow’s tail: the start and the stop of the swing;
- Wave the hands like clouds: the peak of the raise of the arms;
- Push: the start and the stop of the push;
- Raising the power: the peak of the raise of the arms;
- Brush knee twist step: the peak of the wrists when the legs come together;
- Golden rooster: the peak of the rise of the knee.

All these moments were detected by analyzing the trajectories of the joints. Specific detection strategies were designed for each exercise to accommodate their different characteristics. A summary of these strategies is presented in Table 2.1. As the exercise repetitions were performed continuously by the study participants, the start of a new repetition was considered to be the end of the previous one.

Because the recordings began before the exercise started and ended after the exercise finished, cross-correlation was employed to ensure that the frames identified as the start of the repetitions actually captured the exercise execution rather than random movements. More in detail, the assumption was made that among the identified intervals, the one in the middle contained a complete exercise repetition. The cross-correlation was thus calculated between this repetition and all the others to eliminate those with a cross-correlation below a threshold. To maintain a consistent threshold across all subjects, the choice was made to normalize the cross-correlation. This normalization ensured that when comparing two identical repetitions (resulting in auto-correlation), the maximum correlation value would be equal to 1. The function defined for this goal follows Equation (2.1), based on Matlab’s *xcorr* function with the normalization option set to be *normalized* [94].

$$\hat{R}_{xy, norm}(m) = \frac{1}{\sqrt{\hat{R}_{xx}(0)\hat{R}_{yy}(0)}}\hat{R}_{xy}(m) \quad (2.1)$$

Where $\hat{R}_{xy, norm}(m)$ is the normalized cross-correlation, $\hat{R}_{xx}(0)$ is the value of the auto-correlation of the first signal at a delay equal to 0, $\hat{R}_{yy}(0)$ is the value of the auto-correlation of the second signal at a delay equal to 0 and $\hat{R}_{xy}(m)$ is the non-normalized cross-correlation between the two signals.

Automatic segmentation also revealed a lack of synchronicity between the frontal and the lateral recordings for several subjects, which led to the decision of considering only one of the views for the extraction of features. With the exception of the Withdraw and Push exercise (more details in Section 2.1.1), all the others have been analyzed by means of videos recorded from a front view exclusively. On the other hand, as most features of the Withdraw and Push exercise were extracted from the lateral view, this exercise was analyzed through the lateral view recordings.

Validation

Automatic segmentation was validated for the Grasp the Sparrow tail exercise.

To validate the algorithm, frames of interest have been detected manually to provide a ground truth for the analysis. Such ground truth was employed in two ways. Firstly, to assess the rough difference in time between the frames detected automatically and those detected manually, and secondly, to analyze the impact of

these time differences on the values of the features and thus on the classification model performances.

Before being able to compare the time differences between the automatic and the manual segmentation, these two had to be aligned in case a different number of repetitions was detected with the two methods. In particular, these differences might occur at the beginning or at the end of the exercise, when the automatic segmentation needs to discriminate if the detected interval contains the exercise performance or random movements. As mentioned earlier in this Section, the metric employed to decide whether to include or not a repetition was the maximum of a normalized cross-correlation between the analyzed repetition and the one in the middle, which was compared to a threshold. To guarantee that the detected interval was indeed an exercise repetition, the threshold was set to be high so that dubious intervals would be automatically discarded. This choice might lead to the mistaken exclusion of repetitions that are detected manually instead. To ensure that this error does not impact the validation results, the repetitions detected with one method only will be discarded from the analysis.

To examine the impact of errors in the segmentation on the values of the features, features have been extracted both with the manually detected frames and with the automatically detected ones, and the relative error was computed for each feature for every repetition. Features from the automatic segmentation whose difference with respect to the ones from the manual segmentation was higher than $\pm 10\%$ were considered to be wrong. The impact of this error was further analyzed by training a classifier on the features extracted from the manually segmented trials to observe how the model performances would be affected.

2.2.4 Feature Construction

Features have been designed with the goal of extracting the biomechanical characteristics of movement that were relevant to Tai Chi proficiency. To this aim, the collaboration with the Tai Chi experts was crucial as they were able to point out which aspects of movement were of interest for the purpose of scoring the exercise. This collaboration led to the construction of look-up tables where the movement characteristics were associated with a feature that could be extracted from the (x,y) coordinates of the joint trajectories obtained from MediaPipe. The constructed features were grouped into two main categories:

1. Static features: features to extract in a single frame of interest, namely one of those detected in the segmentation step;
2. Dynamic features: features to extract from the trajectory of the joints in a specific time interval of the movement. In this case, only statistical measures

(mean, median, standard deviation, variance, and range) or single-value quantities (Pearson correlation coefficient, distances) were extracted from the time series.

All exercises have their own look-up table and, hence, their own set of features, but they share common ones as well. Furthermore, we extracted the mean, the median, the standard deviation, the variance, and the range of the displacement as well as the mean velocity, the mean acceleration, and the mean jerk of every keypoint. The Look-up tables for each exercise (Tables A.2 - A.1) can be found in Appendix A.

2.3 Model development

This section presents the steps followed to analyze the data. Data feature vectors extracted in the data processing phase were first ranked and selected through the minimum Redundancy Maximum Relevance method [95, 96]. Results of the feature selection were visually assessed through Sammons mapping [97, 98] before being fed into a 3-class balanced Random Forest classifier [99] from the Imbalanced-learn library [100]. All exercises were analyzed separately, and the developed models estimate a proficiency level for a single repetition of each exercise. Hence, the final step of the analysis is that of merging the scores of the single repetitions to estimate a single score per subject.

2.3.1 Feature selection and data projections

Due to the way that features were constructed, feature selection was a crucial step in this analysis for several reasons:

- The number of datapoints was too small with respect to the number of features, so feature selection was needed to mitigate the curse of dimensionality;
- The same biomechanical aspect of movement was characterized in more than one way, hence feature selection was needed to deal with redundancy;
- Reducing the number of features would help maintain the interpretability of the model;
- As the dataset was small, feature selection aided in preventing overfitting.

Feature selection was performed via the minimum Redundancy Maximum Relevance (mRMR) algorithm, designed in 2005 by Peng et al. [95] and coded in Python by Smazzanti et al. [96]. This choice was made after testing different feature selection algorithms, including the Random Forest intrinsic feature selection

method, the wrapper-based Recursive Feature Elimination method, and the mRMR, which is a filter-based approach. Out of these methods, the mRMR stood out not only for its computational efficiency but also because it offered the flexibility to fine-tune the number of features. In contrast, both the Recursive Feature Elimination and the Random Forest intrinsic feature selection method did not provide a feature ranking, preventing the selection of a subset of the chosen features.

To evaluate the optimal number of features from the ranked list outputted by the mRMR method, several models were trained and validated with the Leave-One-Group-Out Cross-Validation technique from the Scikit-Learn library [101]. Each model was trained on an increasing number of features, and performances were evaluated in terms of F1 micro score and F1 weighted score. This procedure allowed us to find the optimal number of features as the one after which the performances of the model were not positively influenced anymore by the addition of new features. An example of this procedure is shown in Figure 2.14, which shows the evolution of the F1 weighted and F1 micro scores as the number of features increases both in the case of the classification of the single repetition and in the case of the estimation of the final score (more details in Paragraph 2.3.3).

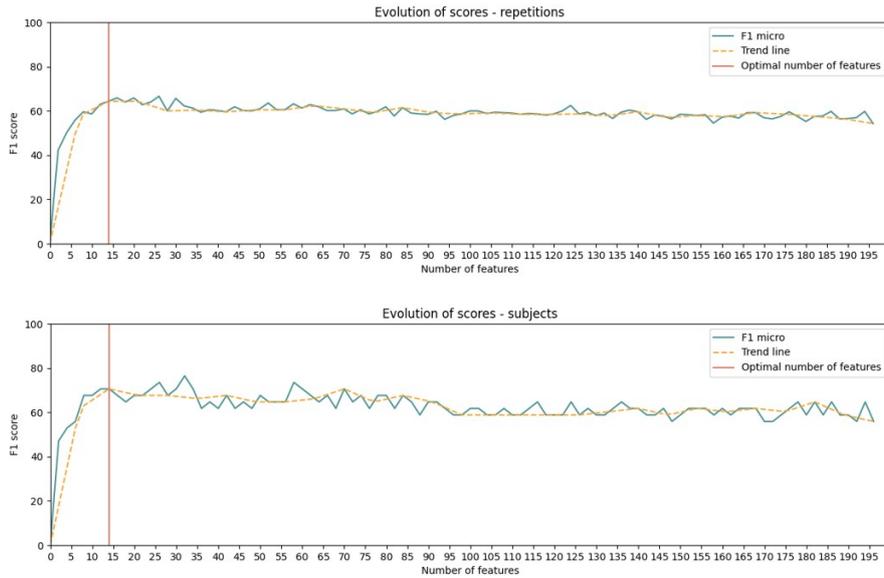


Figure 2.14: Example of the evaluation of the optimal number of features for the Grasp the sparrow’s tail exercise. In this case, 14 features have been selected.

Due to the small dimensions of the Dataset, feature selection was performed on the whole set of data without creating a Training and a Test set.

Results of the Feature Selection were analyzed via Sammons mapping, conceived by Sammon in 1969 [97] and coded in Python by Pollard et al. in 2014 [98].

2.3.2 Model Training

The chosen classification model was the Balanced Random Forest classifier, proposed by Chen et al. (2004) [99] and implemented in the Imbalanced-learn library [100]. This choice was due to the high imbalance of the dataset, which requires to be addressed to prevent the model from becoming unreliable. Dataset imbalance is addressed in the bootstrap phase of the model training: instead of randomly sampling elements of the training set to build the decision trees, the Balanced Random Forest Classifier follows different strategies that allow the representation of all classes by an equal number of samples in the training set [99]. In the Imbalanced-learn library implementation [100], these strategies, which represent a tunable parameter of the algorithm, are:

- "majority": only the majority class is resampled;
- "not minority": all classes are resampled but the minority class;
- "not majority": all classes are resampled but the majority class;
- "all": all classes are resampled.

Due to the small dimensions of the dataset, cross-validation was employed to tune the parameters of the model. To prevent the same subject's different repetitions from being split between the training and test sets, which could lead to overfitting and unreliable model performance evaluations, we used Leave-One-Group-Out cross-validation on each individual repetition.[101]. This method is similar to a K-fold Cross-Validation, but it ensures that the folds are composed of repetitions coming from the same subject. Thus, there were as many folds as the number of subjects who performed the exercise. From the creation of the folds moving forward the cross-validation procedure goes on as usual, so each group is employed once as test set while all the others are employed as training set.

The Leave-One-Group-Out Cross-Validation procedure was exploited in the context of the Grid Search [101], which automatizes the selection of the model hyperparameters to find those that maximize a metric of choice of the user. Due to the imbalance of the dataset, it was chosen to maximize the F1 weighted score. In fact, this metric accounts for the dataset imbalance by weighting the F1 score per each class by its number of elements. Among the parameters of the Balanced Random Forest classifier [100], those that were tuned with the Grid Search were:

- "n_estimators": number of decision trees;
- "criterion": function that measures the quality of the split, either the Gini impurity or the information gain;

- "max_depth": the trees' maximum depth;
- "max_features": the maximum number of features analyzed when searching the best split, either the square root of the number of features or its logarithm in base 2.
- "sampling_strategy": the strategy employed to resample the classes and ensure that they'll be represented in an equal number in the training set, as mentioned earlier in this Section;
- "class_weight": whether weights are associated with the classes; the possibilities are those of associating no weight, associating a weight computed on the frequency of the elements of a class within all the input data or within each tree's bootstrap sample.

Once the best hyperparameters were found with the Grid Search procedure, the classifier performances were evaluated by training and testing the model with the Leave-One-Group-Out Cross-Validation. This procedure was iterated 10 times to then average the results and get mean cross-validated scores that mitigate the variability of a single iteration's results. The performances of the classifier were evaluated both in terms of the F1 weighted score, which takes into account the dataset imbalance, and in terms of F1 micro score, which refers to the overall performances of the model.

2.3.3 Estimation of the final proficiency level

When assigning the score, the Tai Chi experts considered all the repetitions and provided only one score per exercise per subject. Hence, to mimic this procedure, the last step of the analysis was to merge the scores estimated for the single repetitions into a single score per subject.

This was done by looking at the score that was estimated most frequently. In case two different scores were both selected as frequently and as often, the final score was estimated to be the one whose probability in the tree prediction was the highest.

The performances of the model were also evaluated with respect to its ability to estimate a single score per subject. Again, the metrics employed were the F1 weighted score and the F1 micro score.

Exercise	Moment of interest	Detection strategy
Grasp the sparrow's tail	Start of the repetition	Zero-crossing of the difference in x coordinate between the left wrist and the nose.
	Start of the swing	Start of the increase of the y coordinate of the left wrist, detected by looking at the change in slope.
	Stop of the swing	Stop of the decrease of the y coordinate of the left wrist, detected by looking at the change in slope.
Wave the hands like clouds	Start of the repetition	Intersection of the x and y trajectories (minus the mean) of the left wrist.
	Peak of the hands before descent	Intersection of the x and y trajectories (minus the mean) of the right wrist.
Withdraw and push	Start of the repetition	Peaks of the y trajectory of the most visible wrist.
	Start of the push	First intersection of the x and y trajectories (minus the mean) of the most visible wrist.
	Stop of the push	Second intersection of the x and y trajectories (minus the mean) of the most visible wrist.
Brush knee twist step	Start of the repetition	Start of the increase in x coordinate of the back foot, detected by looking at the change in slope.
	Peak of the hands	Peaks of the y trajectory of the front hand.
Golden rooster	Start of the repetition	Start of the increase in x coordinate of the back foot, detected by looking at the change in slope.
	Peak of the knee	Peaks of the y trajectory of the back knee.
Raising the power	Start of the repetition	Start of the increase in y coordinate of the middle point between the hands, detected by looking at the change in slope.
	Peak of the hands	Peaks of the y trajectory of the middle points between the hands.

Table 2.1: Automatic segmentation for all exercises; moments of interest and detection strategy.

Results

This Chapter presents the Results of the classification models developed for each exercise. The results include:

- The number of features selected;
- The features selected;
- The Sammon projection of the dataset after Feature Selection;
- The hyperparameters optimized by the Grid Search;
- The mean Cross-Validated F1 weighted and F1 micro scores, both for the classification of the single repetitions and for the estimation of a single score per subject;
- The final Confusion Matrices, both for the classification of the single repetitions and for the estimation of a single score per subject;

For what concerns the Grasp the sparrow's tail exercise, additional results regarding the validation of the automatic segmentation are reported. These results are not available for the other exercises as the segmentation was performed exclusively automatically, hence there was no ground truth to perform the analysis. These results include both analyses performed: the first one on the raw time difference between the frames detected manually and those detected automatically and the second one on the impact of these differences on the value of the features and hence on the classifier results. To this goal, a second classification model was trained for the Grasp the sparrow's tail exercise with the features extracted with the manual segmentation. All analyses performed on the other models will be presented for this one as well to further validate the performances of the automatic segmentation.

Regarding the Wave the hands like clouds exercise, two models were developed as well to analyze separately the instances in which the subject was performing the exercise with one hand only or with both hands.

Lastly, not all exercises have been performed by the same number of subjects. For example, the Brush knee twist step and the Golden rooster exercises were sometimes too challenging for the study participants as they respectively required to move the feet close together lift the knee, and stay on one leg only. Another example concerns the Withdraw and push exercise, for which the video data of the Tai Chi experts is missing as there was no recording of the lateral view. Furthermore, for those exercises that had more than one version, some subjects only performed one, decreasing the amount of data available. Table 3.1 summarizes the number of video data available and hence the number of subjects analyzed for each exercise.

Exercise	Study participants	Tai Chi experts	Total number of subjects	Total number of repetitions
Grasp the sparrow's tail	32	2	34	528
Wave hands like clouds (both hands)	32	2	34	390
Wave hands like clouds (single hand)	32	2	34	391
Withdraw and push	31	0	31	469
Brush knee twist step	28	2	30	427
Golden rooster	27	2	29	308
Raising the power	32	2	34	546

Table 3.1: Number of subjects whose video data was available for each exercise.

3.1 Grasp the sparrow's tail

This Section presents the results obtained in the analysis of the Grasp the sparrow's tail exercise. Before moving to the results of the classifier, the validation of the automatic segmentation will be presented.

3.1.1 Validation of the automatic segmentation

As mentioned before, the automatic segmentation was validated in two ways: by looking at the raw differences in time between the frames detected manually and those detected automatically and by looking at the impact of those differences on the features and hence on the classifier's performances.

As foreseen in Paragraph 2.2.3, before being able to start the analysis the repetitions segmented with the two methods had to be aligned. In fact, the total amount of automatically detected repetitions was equal to 528 while the total amount of manually detected repetitions was 540. Considering that 3 automatically detected repetitions were not detected manually, 15 repetitions were missed by the automatic segmentation. These 15 repetitions, together with the 3 ones that were not detected manually, were discarded from both the raw time difference analysis and from the analysis on the impact of the features but they will surely have an effect on the performances of the trained models.

Time differences

The results of the time difference calculations are presented in Figure 3.1. 95% of the time differences are smaller than 1.5 seconds, which can be considered to be a good result as one of the key features of Tai Chi movements is their slowness. Nevertheless, there are five instances of time differences greater than 2.5 seconds, namely for one start of the repetition, one start of the swing, and three stops of the swing, on four different subjects. These errors should be addressed as such a big time difference could have a strongly negative impact on the value of the extracted features that rely on these moments of interest.

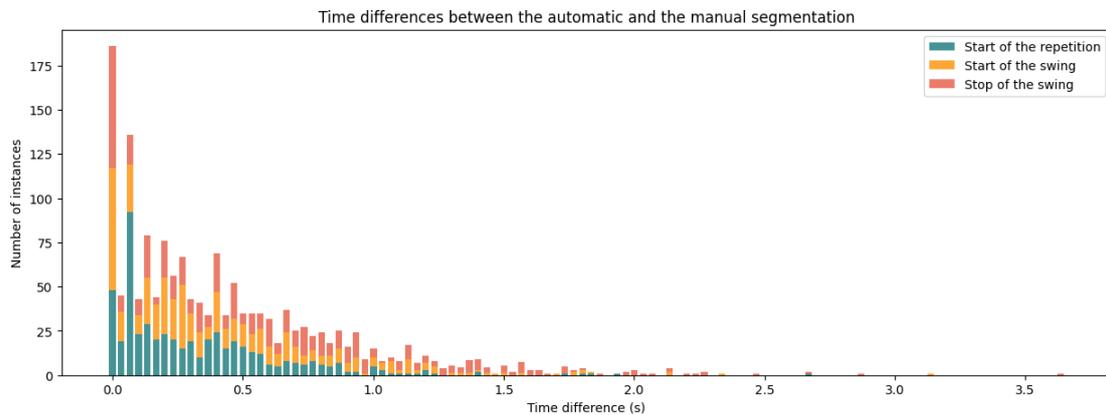


Figure 3.1: Bar plot of the time differences between the automatic and the manual segmentation. The time differences have been calculated by taking the absolute value of the difference between the automatically detected time frame and the manually detected one. Results for the time differences relative to the start of the repetition (green), the start of the swing (orange) and the stop of the swing (red) were computed separately and then merged on a single graph to provide an overall evaluation of the automatic segmentation performance.

Impact on the features

16.48% of the features extracted from the automatically segmented data have an error higher than $\pm 10\%$ of the respective feature extracted from the manually segmented data. The majority of these features were static ones, hence the ones that were extracted in the single frame detected. To further assess the impact of this error on the classifier performances, a model was trained with the features extracted from the manually segmented joint trajectories. This classifier reached an F1 micro score of 68.16% when classifying the single repetitions and an F1 micro score of 81.47% on the estimation of the final score per subject. Compared to the results obtained from the classifier trained on the features extracted from the automatically detected trials, presented in Section 3.1.2, a decrease in F1 micro score on the classification of the repetitions of 1.92%, that translates into a decrease in F1 micro score on the subjects' proficiency level estimation of 6.18%.

3.1.2 Classification results

As shown in Figure 3.2, for the Grasp the sparrow's tail exercise, 14 features were selected. Those features, together with a brief explanation, are presented in Table 3.2. Further details on the features can be found in Table A.2. The list of optimized hyperparameters is presented in Table 3.3; further information on the meaning of those parameters can be found in Section 2.3.2. The projection of the features after Feature Selection is presented in Figure 3.3. Lastly, the performances of the model, in terms of mean cross-validated F1 micro and F1 weighted scores, both for the classification of the repetitions and for the estimation of the final score are presented in Table 3.4, while Figure 3.4 shows the Confusion Matrices for both cases.

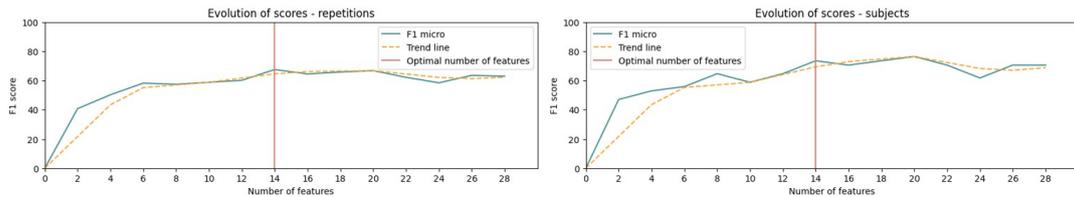


Figure 3.2: Optimization of the number of features parameter for the Grasp the Sparrow's tail exercise.

Selected Features
Range of x displacement of the left hip
Mean of y displacement of the right ankle
Position of the right finger at the end of swing
Range of y displacement of the left shoulder
Median of the angle of the right elbow from the stop of swing to the end of the repetition
Range of x displacement of the right hip
Mean of the angle of the right elbow from the stop of swing to the end of the repetition
Range of x displacement of the right wrist
Median of the curvature of the left wrist during the swing
Variance of the angle of the right elbow from the stop of swing to the end of the repetition
Range of x displacement of the right knee
Correlation between the wrist and the hips movement
Range of y displacement of the head
Synchronicity of the hips and shoulders movement (x trajectory)

Table 3.2: Features selected for the Grasp the sparrow’s tail exercise.

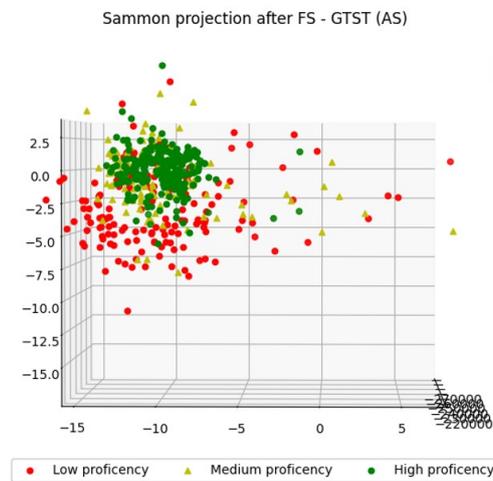


Figure 3.3: Sammon projection of the data after Feature Selection for the Grasp the sparrow’s tail exercise.

Hyperparameter	Optimized value
n_estimators	100
criterion	'gini'
max_depth	None
max_features	'sqrt'
sampling_strategy	'not majority'
class_weight	None

Table 3.3: Optimized hyperparameters for the Grasp the sparrow’s tail exercise.

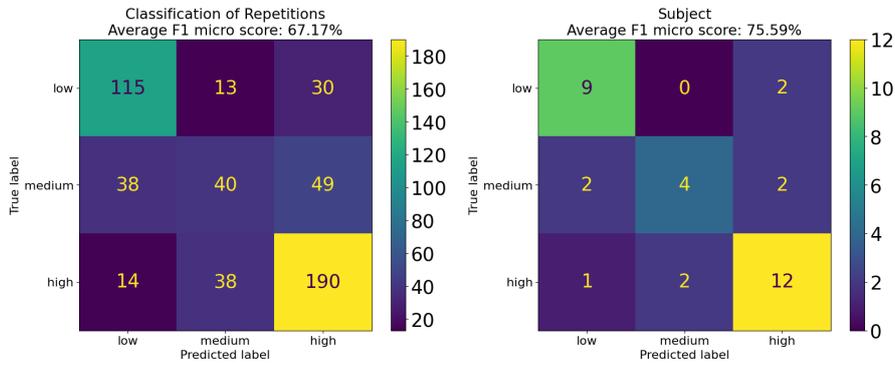


Figure 3.4: Confusion Matrices presenting the classification results for the Grasp the Sparrow’s tail exercise.

Metric	Target	Result (mean \pm std)
F1 micro score	Single repetition	67.17% \pm 1.35%
	Subject	75.59% \pm 2.96%
F1 weighted score	Single repetition	65.84% \pm 1.3%
	Subject	74.66% \pm 2.94%

Table 3.4: Classification model performances in terms of F1 micro and F1 weighted scores for the Grasp the Sparrow’s tail exercise.

3.2 Wave the hands like clouds

This Section presents the results obtained in the analysis of the Wave the hands like clouds exercise. As mentioned earlier in this Chapter, the instances in which the movement was performed with both hands were analyzed separately than the instances in which the movement was performed with one hand only, leading to the development of two different models.

3.2.1 Performance with both hands

As shown in Figure 3.5, for the Wave the hands like clouds exercise performed with both hands, 4 features were selected. Those features, together with a brief explanation, are presented in Table 3.5. Further details on the features can be found in Table A.3. The list of optimized hyperparameters is presented in Table 3.6; further information on the meaning of those parameters can be found in Section 2.3.2. The projection of the features after Feature Selection is presented in Figure 3.6. Lastly, the performances of the model, in terms of mean cross-validated F1 micro and F1 weighted scores, both for the classification of the repetitions and for the estimation of the final score are presented in Table 3.7, while Figure 3.7 shows the Confusion Matrices for both cases.

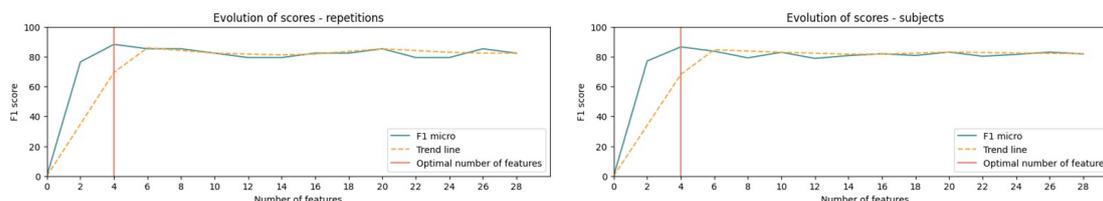


Figure 3.5: Optimization of the number of features parameter for the Wave the hands like clouds exercise performed with both hands.

Selected Features
Similarity of the trajectories of the right and left wrists
Mean of x displacement of the right wrist
Range of x displacement of the right hip
Range of y displacement of the left elbow

Table 3.5: Features selected for the Wave the hands like clouds exercise performed with both hands.

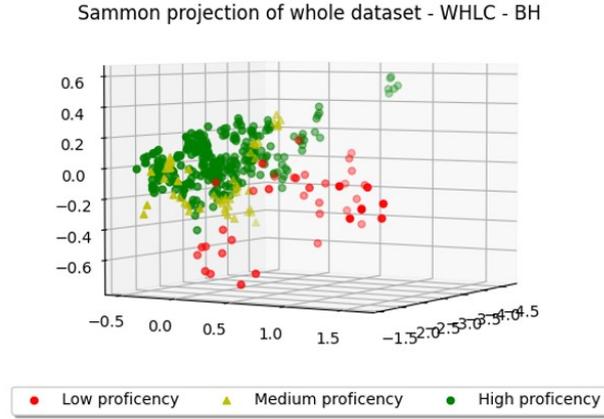


Figure 3.6: Sammon projection of the data after Feature Selection for the Wave the hands like clouds exercise performed with both hands.

Hyperparameter	Optimized value
n_estimators	50
criterion	'gini'
max_depth	10
max_features	'sqrt'
sampling_strategy	'all'
class_weight	'balanced'

Table 3.6: Optimized hyperparameters for the Wave the hands like clouds exercise performed with both hands.

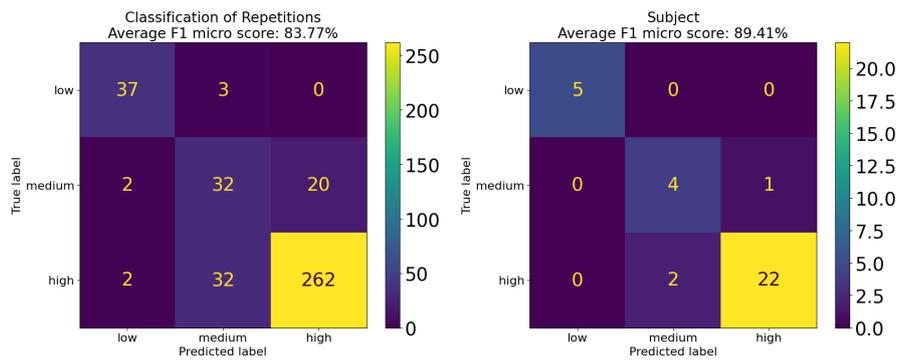


Figure 3.7: Confusion Matrices presenting the classification results for the Wave the hands like clouds exercise performed with both hands.

Metric	Target	Result (mean \pm std)
F1 micro score	Single repetition	83.77% \pm 0.87%
	Subject	89.41% \pm 1.95%
F1 weighted score	Single repetition	84.42% \pm 0.83%
	Subject	89.49% \pm 2.0%

Table 3.7: Classification model performances in terms of F1 micro and F1 weighted scores for the Wave the hands like clouds exercise performed with both hands.

3.2.2 Performance with a single hand

As shown in Figure 3.8, for the Wave the hands like clouds exercise performed with one hand only, 8 features were selected. Those features, together with a brief explanation, are presented in Table 3.8. Further details on the features can be found in Table A.3. The list of optimized hyperparameters is presented in Table 3.9; further information on the meaning of those parameters can be found in Section 2.3.2. The projection of the features after Feature Selection is presented in Figure 3.9. Lastly, the performances of the model, in terms of mean cross-validated F1 micro and F1 weighted scores, both for the classification of the repetitions and for the estimation of the final score are presented in Table 3.10, while Figure 3.10 shows the Confusion Matrices for both cases.

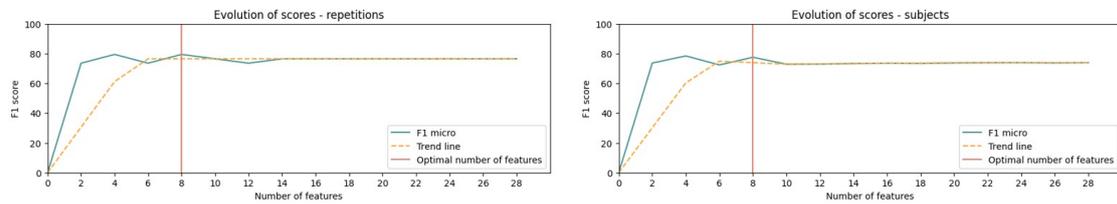


Figure 3.8: Optimization of the number of features parameter for the Wave the hands like clouds exercise performed with one hand.

Selected Features
Range of x displacement of the right hip
Variance in difference in height of the elbows
Range of x displacement of the right knee
Range of y displacement of the head
Range of x displacement of the left hip
Synchronicity in movement of the left hip and the left shoulder (x trajectory)
Mean of the distance between knees
Range of x displacement of the left knee

Table 3.8: Features selected for the Wave the hands like clouds exercise performed with one hand.

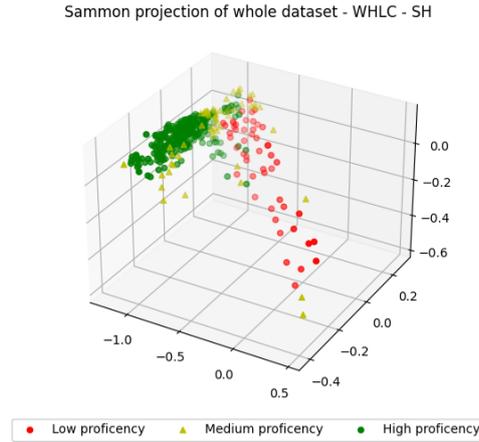


Figure 3.9: Sammon projection of the data after Feature Selection for the Wave the hands like clouds exercise performed with one hand.

Hyperparameter	Optimized value
n_estimators	200
criterion	'gini'
max_depth	40
max_features	'log2'
sampling_strategy	'all'
class_weight	'balanced'

Table 3.9: Optimized hyperparameters for the Wave the hands like clouds exercise performed with one hand.

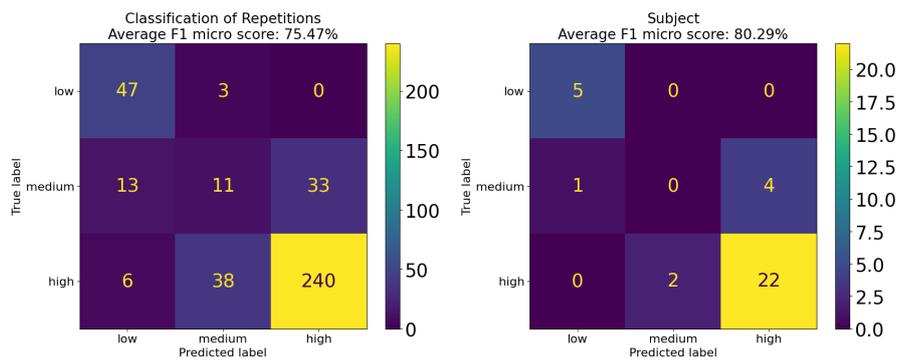


Figure 3.10: Confusion Matrices presenting the classification results for the Wave the hands like clouds exercise performed with one hand.

Metric	Target	Result (mean \pm std)
F1 micro score	Single repetition	75.47% \pm 0.6%
	Subject	80.29% \pm 2.3%
F1 weighted score	Single repetition	75.27% \pm 0.65%
	Subject	78.22% \pm 2.64%

Table 3.10: Classification model performances in terms of F1 micro and F1 weighted scores for the Wave the hands like clouds exercise performed with one hand.

3.3 Withdraw and push

This Section presents the results obtained in the analysis of the Withdraw and push exercise.

As shown in Figure 3.11, for the Withdraw and push exercise, 4 features were selected. Those features, together with a brief explanation, are presented in Table 3.11. Further details on the features can be found in Table A.4. The list of optimized hyperparameters is presented in Table 3.12; further information on the meaning of those parameters can be found in Section 2.3.2. The projection of the features after Feature Selection is presented in Figure 3.12. Lastly, the performances of the model, in terms of mean cross-validated F1 micro and F1 weighted scores, both for the classification of the repetitions and for the estimation of the final score are presented in Table 3.13, while Figure 3.13 shows the Confusion Matrices for both cases.

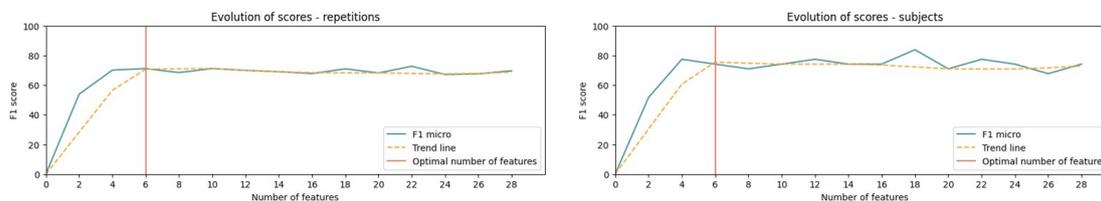


Figure 3.11: Optimization of the number of features parameter for the Withdraw and push exercise.

Selected Features
Median of the angle of the front ankle
Variance of the weight in the back leg (difference in x coordinate of toe, knee and ankle)
Standard deviation of the weight in the front leg (difference in x coordinate of knee and ankle)
Mean of the weight in the front leg (difference in x coordinate of knee and ankle)
Mean of the head alignment
Height of the left wrist at the start of the push

Table 3.11: Features selected for the Withdraw and push exercise.

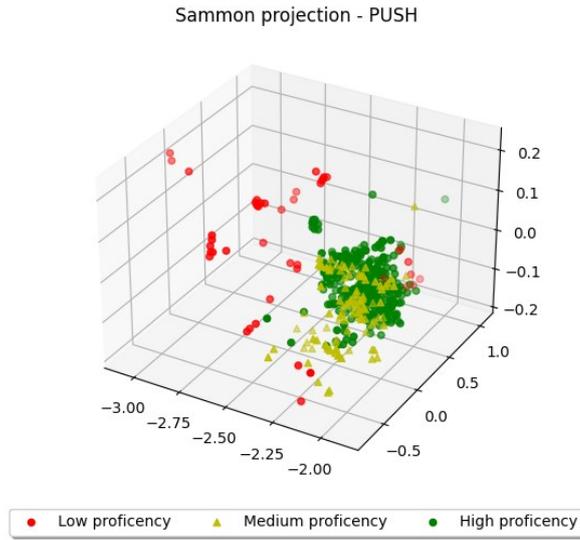


Figure 3.12: Sammon projection of the data after Feature Selection for the Withdraw and push exercise.

Hyperparameter	Optimized value
n_estimators	50
criterion	'gini'
max_depth	None
max_features	'log2'
sampling_strategy	'not majority'
class_weight	None

Table 3.12: Optimized hyperparameters for the Withdraw and push exercise.

Metric	Target	Result (mean \pm std)
F1 micro score	Single repetition	79.25% \pm 1.63%
	Subject	81.29% \pm 4.03%
F1 weighted score	Single repetition	77.54% \pm 1.9%
	Subject	79.36% \pm 4.64%

Table 3.13: Classification model performances in terms of F1 micro and F1 weighted scores for the Withdraw and push exercise.

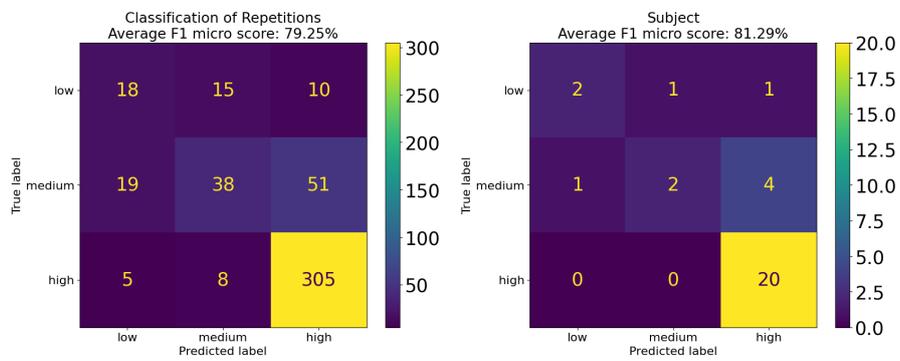


Figure 3.13: Confusion Matrices presenting the classification results for the Withdraw and push exercise.

3.4 Brush knee twist step

This Section presents the results obtained in the analysis of the Brush knee twist step exercise.

As shown in Figure 3.14, for the Brush knee twist step exercise, 4 features were selected. Those features, together with a brief explanation, are presented in Table 3.14. Further details on the features can be found in Table A.5. The list of optimized hyperparameters is presented in Table 3.15; further information on the meaning of those parameters can be found in Section 2.3.2. The projection of the features after Feature Selection is presented in Figure 3.15. Lastly, the performances of the model, in terms of mean cross-validated F1 micro and F1 weighted scores, both for the classification of the repetitions and for the estimation of the final score are presented in Table 3.16, while Figure 3.16 shows the Confusion Matrices for both cases.

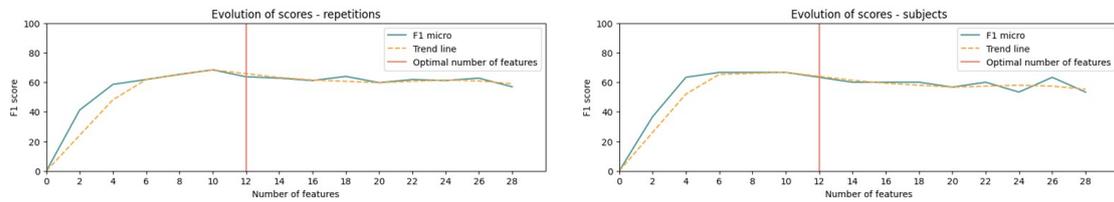


Figure 3.14: Optimization of the number of features parameter for the Brush knee twist step exercise.

Selected Features
Range of x displacement of the left elbow
Median of the pelvis displacement from the beginning to the middle of the repetition
Height of the front wrist at the end of the repetition
Median of the angle of the right knee during the repetition
Range of x displacement of the right ankle
Range of y displacement of the right knee
Range of y displacement of the right wrist
Height of the back wrist at the end of the repetition
Standard deviation of the distance between knees
Height of the front wrist at the beginning of the repetition
Mean velocity of the head (y trajectory)
Range of x displacement of the left wrist

Table 3.14: Features selected for the Brush knee twist step exercise.

Sammon projection after FS - BKTS

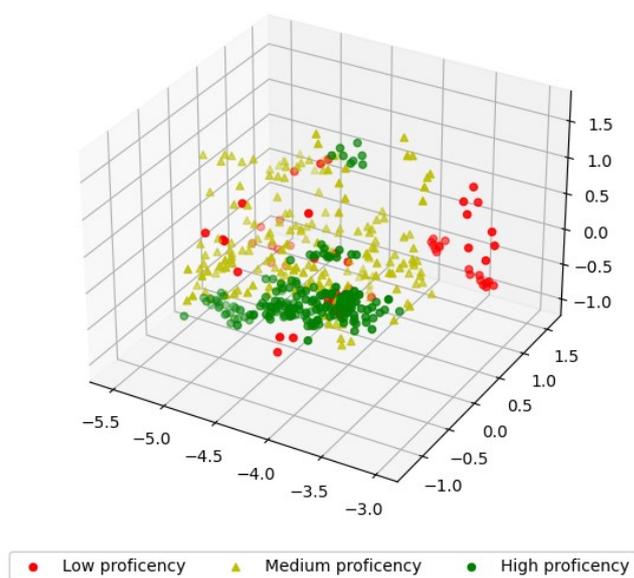


Figure 3.15: Sammon projection of the data after Feature Selection for the Brush knee twist step exercise.

Hyperparameter	Optimized value
n_estimators	50
criterion	'gini'
max_depth	20
max_features	'sqrt'
sampling_strategy	'all'
class_weight	'balanced_subsample'

Table 3.15: Optimized hyperparameters for the Brush knee twist step exercise.

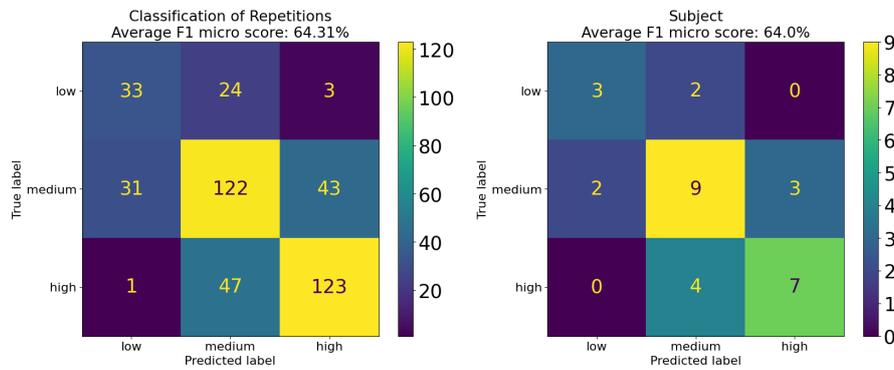


Figure 3.16: Confusion Matrices presenting the classification results for the Brush knee twist step exercise.

Metric	Target	Result (mean \pm std)
F1 micro score	Single repetition	64.31% \pm 1.59%
	Subject	64.0% \pm 2.91%
F1 weighted score	Single repetition	64.45% \pm 1.55%
	Subject	63.96% \pm 2.96%

Table 3.16: Classification model performances in terms of F1 micro and F1 weighted scores for the Brush knee twist step exercise.

3.5 Golden rooster

This Section presents the results obtained in the analysis of the Golden rooster exercise.

As shown in Figure 3.17, for the Golden rooster exercise, 4 features were selected. Those features, together with a brief explanation, are presented in Table 3.17. Further details on the features can be found in Table A.6. The list of optimized hyperparameters is presented in Table 3.18; further information on the meaning of those parameters can be found in Section 2.3.2. The projection of the features after Feature Selection is presented in Figure 3.18. Lastly, the performances of the model, in terms of mean cross-validated F1 micro and F1 weighted scores, both for the classification of the repetitions and for the estimation of the final score are presented in Table 3.19, while Figure 3.19 shows the Confusion Matrices for both cases.

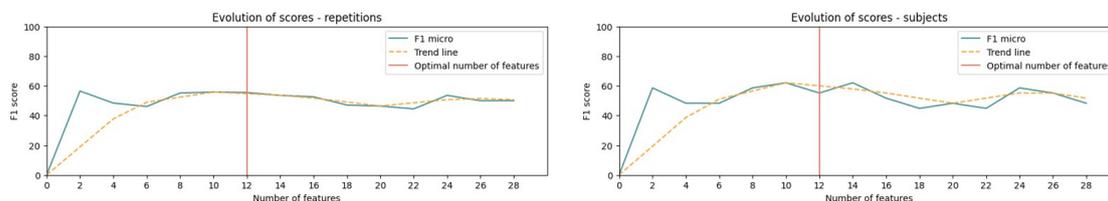


Figure 3.17: Optimization of the number of features parameter for the Golden rooster exercise.

Selected Features
Mean of the velocity of the raised arm relative to the knee
Mean of the y displacement of the right ankle
Variance of the alignment of the head with respect to the midpoint of the shoulders
Median of the distance between the knees during the repetition
Alignment of the right knee to the right ankle at the knee peak
Separation between the raised arm and the knee at knee peak
Height of the left wrist at knee peak
Standard deviation of the distance between the knees during the repetition
Mean of the acceleration of the raised arm relative to the knee
Standard deviation of the alignment of the head with respect to the midpoint of the shoulders
Alignment of the right wrist to the elbow at knee peak
Mean of y displacement of the right elbow

Table 3.17: Features selected for the Golden rooster exercise.

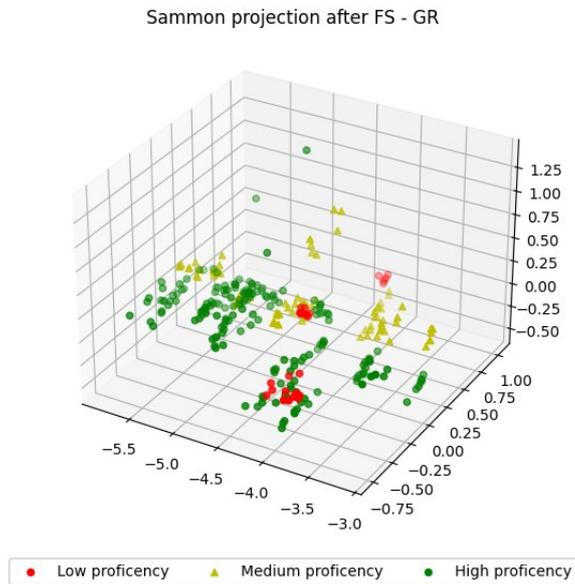


Figure 3.18: Sammon projection of the data after Feature Selection for the Golden rooster exercise.

Hyperparameter	Optimized value
n_estimators	50
criterion	'gini'
max_depth	40
max_features	'sqrt'
sampling_strategy	'not majority'
class_weight	None

Table 3.18: Optimized hyperparameters for the Golden rooster exercise.

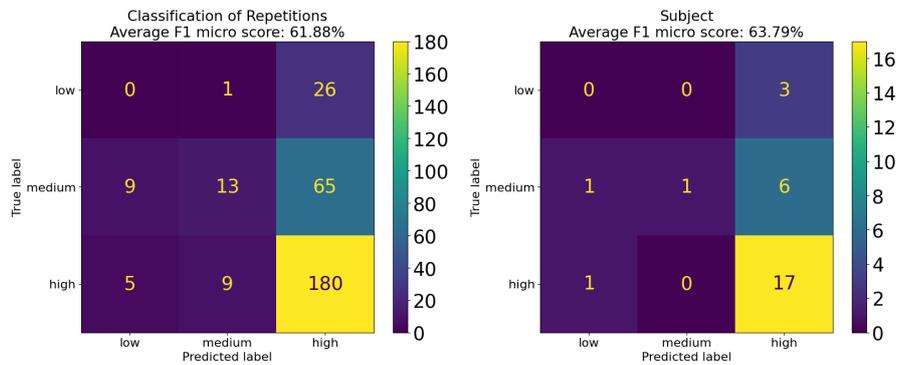


Figure 3.19: Confusion Matrices presenting the classification results for the Golden rooster exercise.

Metric	Target	Result (mean \pm std)
F1 micro score	Single repetition	61.88% \pm 0.57%
	Subject	63.79% \pm 3.18%
F1 weighted score	Single repetition	54.46% \pm 1.22%
	Subject	54.15% \pm 3.52%

Table 3.19: Classification model performances in terms of F1 micro and F1 weighted scores for the Golden rooster exercise.

3.6 Raising the power

This Section presents the results obtained in the analysis of the Raising the power exercise.

As shown in Figure 3.20, for the Raising the power exercise, 4 features were selected. Those features, together with a brief explanation, are presented in Table 3.20. Further details on the features can be found in Table A.7. The list of optimized hyperparameters is presented in Table 3.21; further information on the meaning of those parameters can be found in Section 2.3.2. The projection of the features after Feature Selection is presented in Figure ???. Lastly, the performances of the model, in terms of mean cross-validated F1 micro and F1 weighted scores, both for the classification of the repetitions and for the estimation of the final score are presented in Table 3.22, while Figure 3.21 shows the Confusion Matrices for both cases.

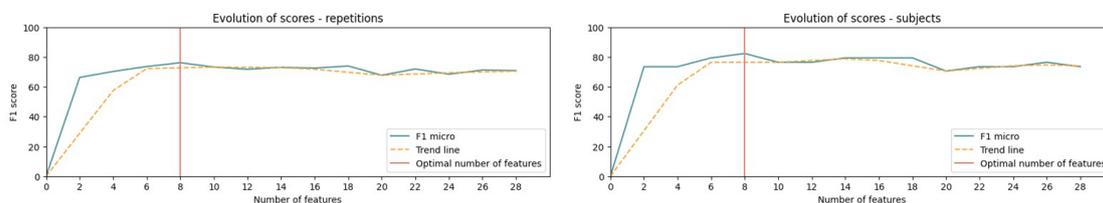


Figure 3.20: Optimization of the number of features parameter for the Raising the power exercise.

Selected Features
Synchronicity of movement of the between right wrist and right hip
Mean of the distance between the knees
Range of y displacement of the right hip
Standard deviation of the neck angle
Synchronicity of movement of the midpoint between wrists and hips
Height of the left wrist relative to the left shoulder at the peak of hands
Synchronicity of movement of the between left wrist and left hip
Range of y displacement of the left hip

Table 3.20: Features selected for the Raising the power exercise.

Hyperparameter	Optimized value
n_estimators	50
criterion	'gini'
max_depth	10
max_features	'log2'
sampling_strategy	'not minority'
class_weight	'balanced_subsample'

Table 3.21: Optimized hyperparameters for the Raising the power exercise.

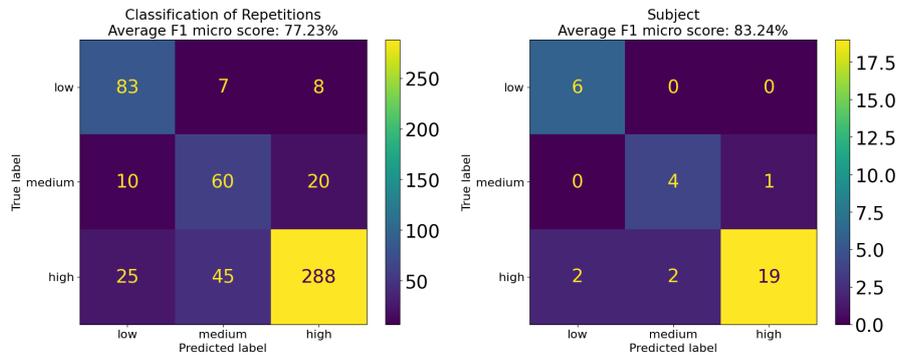


Figure 3.21: Confusion Matrices presenting the classification results for the Raising the power exercise.

Metric	Target	Result (mean \pm std)
F1 micro score	Single repetition	77.23% \pm 1.21%
	Subject	83.24% \pm 2.3%
F1 weighted score	Single repetition	77.93% \pm 1.17%
	Subject	83.66% \pm 2.2%

Table 3.22: Classification model performances in terms of F1 micro and F1 weighted scores for the Raising the power exercise.

3.7 Summary of the results

In summary, all of the developed models, except for those concerning the Brush Knee Twist Step and the Golden Rooster exercises, consistently achieve an F1 micro score of over 75% when estimating the proficiency level of Tai Chi practitioners. In contrast, the Brush Knee Twist Step and the Golden Rooster exercises exhibit an F1 micro score just above 60%.

An overview of these results, which will be discussed in Chapter 4, is provided in Figure 3.22.

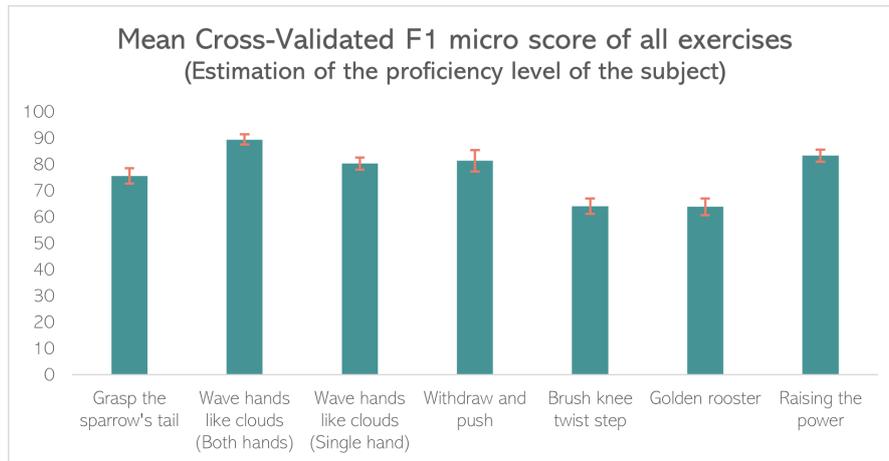


Figure 3.22: Overview of the classifiers' performance in assessing the proficiency level of Tai Chi practitioners measured with the F1 micro score.

Discussion

In this Chapter, the results obtained and presented in Chapter 3 will be analyzed.

The validation of the automatic segmentation for Grasp the sparrow's tail showed that the quality of the segmentation step has an impact on the classification performances. Nevertheless, it's worth pointing out that there is a remaining degree of incertitude on the ground truth due to the possibility of human errors. In fact, having multiple annotators provides a better estimate and allows to be robust against human error, but in this work each video was manually segmented by a single person. However, the classifier trained on the features extracted from the manually segmented trajectories demonstrated how performances over 80% could be reached for the Grasp the sparrow's tail exercise. Hence, working on a more robust segmentation for this exercise should increase the performance of the classifier.

The performance of the developed models, as measured by the F1 micro score for estimating subject scores, generally exceeds 75-80%. However, there are two exercises, namely Brush knee twist step and Golden rooster, for which the model performance is comparatively lower. This decline in model performance can be attributed to the greater complexity of these two exercises in comparison to the others. The intricacy of these exercises poses a challenge for the extracted features to adequately characterize the biomechanical aspects of movement, resulting in a decrease in model performance. This observation is reinforced by the fact that, for the classification of these two exercises, a higher number of features is deemed optimal compared to all other exercises except for Grasp the sparrow's tail.

Furthermore, due to the heightened complexity of these exercises, the execution by study participants exhibits greater variability. This increased variability adversely affects the classifier's ability to accurately group subjects with similar scores. Additionally, as discussed in more detail in Section 4.1, the scoring method does not provide a single score per repetition and thus doesn't properly reflect the variability observed in the exercise execution. Particularly in the case of more complex exercises like Brush knee twist step and Golden rooster, there is a higher likelihood of mislabeled repetitions, leading to a decrease in the classification

performance of the models.

Regarding the Golden rooster exercise, it appears that dataset imbalance negatively affected the classification performance despite the utilization of the Balanced Random Forest classifier. Specifically, most misclassified subjects have low or medium proficiency scores and are frequently misclassified as having high proficiency, the class with the highest number of instances, showing that the model is probably overfitting.

4.1 Limitations of the study

This study presented several limitations. Most of these limitations concern the way that video data were collected, but also have to do with the dimension and the lack of balance of the dataset and the way that scores were assigned.

During different sessions of data collection, there was a lack of consistency in the placement of the cameras with respect to the subject, both in terms of distance between and alignment of the subject and the cameras. The fact that the distance between the front camera and the subject was variable was compensated by the normalization of the skeletal dimensions. On the other hand, the fact that sometimes there was an angle between the field of view of the camera and the subject could not be compensated for. This introduced an avoidable increase in variability within the dataset.

Furthermore, laterally recorded videos were recorded from the same side of the subject regardless of the version of the exercise being performed. In other words, if the study participant was executing the exercise in one version on its mirrored one the camera would still be in the same position. This limitation concerns the Withdraw and push exercise exclusively, as for the others, the lateral view was discarded, but the result was that for this exercise, it made no sense to mirror one of the two versions of the exercise. Hence, a lack of consistency was introduced in the analysis.

An additional limitation of the study concerns the way the scores were assigned. Firstly, scores were assigned by different Tai Chi experts, producing a possible small variability in the scores. Secondly, the provided scores were one per subject only, while the developed models classified each repetition. On one hand, this choice was made to augment the size of the dataset but on the other, it has the downfall of assuming that all repetitions of one subject have the same score, namely the subject's. Unfortunately, this is not always true: if the study participant's performance was increasing during the trial, the Tai Chi experts would assign the best score reached. Hence, the first repetition, with a lower proficiency level, would still receive a high score. This procedure adds a good amount of noise to the data, making it harder for the classifier to create boundaries between the classes. In the

future, dubious repetitions might be flagged to be re-labeled by the Tai Chi experts to try to decrease this noise component.

Lastly, the small dimensions of the dataset make it harder to generalize the results obtained in this Thesis. This limitation can be minimized by collecting more data or including data coming from different studies in this analysis.

4.2 Future perspectives

There are several future perspectives for this work. Firstly, it would be of great interest to include the 3rd dimension in the analysis and get information on depth. This can be achieved by collecting data with a depth camera or by employing skeleton reconstruction techniques. Using skeleton tracking methods different than MediaPipe might be of interest as well as multiperson Human Pose Estimation platforms could minimize the tracking errors due to the presence of multiple individuals in the frame.

Furthermore, the results of the validation of the automatic segmentation suggest that it could be appropriate to validate the segmentation strategies for the other exercises as well. The downfall of this choice is that manual labeling would be needed to provide the ground truth for the analysis. However, the results of this analysis might once again encourage to work on making the segmentation more robust to increase the model performances.

In addition, future analyses could include the data coming from the lateral view recordings after synchronization with the frontal ones. In fact, using the lateral data would provide a gain in information that could increase the performance of the classifiers. This is especially true for the analysis of Tai Chi exercises, as one key aspect in the practice of the discipline is the shift of weight, and features designed to characterize this aspect of movement often rely on the lateral view. Nevertheless, videos capturing the lateral side of the subject are not always recorded. Developing a pipeline that only relies on the frontal view is hence more general and applicable to other cases of study.

Lastly, the pipeline developed in this study can be applied to a longitudinal study of the Motion Analysis Lab, whose data collection was recently concluded. This project followed the progress of Tai Chi naive study participants who were trained for 12 weeks and evaluated 4 times during the study, with the goal of understanding whether proficiency matters in terms of motor gains. Hence, in the context of assessing whether there's a relation between the increase in balance and the increase in Tai Chi proficiency, the pipeline developed in this study can play a crucial role by being able to estimate the latter automatically.

Conclusions

In the context of investigating the beneficial effect of Tai Chi on health, and specifically on balance, the goal of this Thesis was to develop Machine Learning algorithms to automatically assess Tai Chi proficiency from video data analyzed with Human Pose Estimation.

To reach this aim, previously collected data from 32 older adults with a wide range of proficiency levels was analyzed. After extracting the joint trajectories with MediaPipe, normalizing the skeleton dimensions, and segmenting the trials, specifically designed features were extracted. These features were ranked, selected, and used to train a Balanced Random Forest classifier to classify the individual exercise repetition into 3 levels of proficiency. From this information, a final score per subject was estimated by picking the score that was assigned most frequently and with the highest confidence.

The overall performances of the developed models in terms of F1 micro score exceeds 80% with the exception of the two most complex exercises. In these cases, the extracted features were unable to capture the exercise complexity and hence allowed the models to estimate the proficiency level of the study participants correctly.

Despite the small dimension of the dataset, this study serves as proof of concept that Human Pose Estimation and Machine Learning can be employed to assess Tai Chi proficiency. The pipeline developed in this work will be employed in a longitudinal study on Tai Chi that investigates whether exercise proficiency matters in terms of motor gains.

Look-up Tables

This Appendix contains the full Look-up tables for all Exercises (Tables A.1 - A.7). These tables associate the aspect of movement of interest to a measurement strategy and the respective extracted features. These look-up tables are the result of a collaboration with the Tai Chi experts, who provided the aspect of movement they were looking at when assigning the proficiency scores to the exercises, and the engineering team, who converted these observations into quantifiable features that could be extracted from the (x,y) trajectories of the joints.

These tables include only the features that have been considered in this Thesis work. Further features were designed but not extracted as they would have required to exploit the second view (namely the lateral view for all exercises but the Withdraw and push), which was discarded in this work as explained in Paragraph 2.2.3.

Aspect of movement	Measurement	Extracted values
Whether the ankles are shoulder width-apart in the bow stance	Ratio of the distance between the ankles and the distance between the shoulders	Mean, median, standard deviation and variance of the time series
Whether the left or right knee is excessively bent from side to side	Observation of the knee angle	Mean, median, standard deviation, variance and range of the time series
Assessment of the separation of the legs in the frontal plane	Observation of the distance between the knees	Mean, median, standard deviation, variance and range of the time series

Table A.1: Look-up table for the features common to all exercises.

Aspect of movement	Measurement	Extracted values
Whether the left or right wrist is at collarbone height at the beginning of repetition, the beginning of swing and the end of swing	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the hands are centered with the body at the beginning of repetition	Difference in x coordinate between the left or right wrist and the midpoint between the shoulders	Feature value in the moment of interest
Whether one of the hands barely peeking out behind the other at the beginning of repetition	Difference in visibility between the wrists	Feature value in the moment of interest
Whether the body midline is moved towards one side of the body, indicating a weight shift at the beginning of repetition	Difference in x coordinate between the midpoint of the ankles and the midpoint of the shoulders	Feature value in the moment of interest
Whether the fingers of the left or right hand are curved with the respective wrist or splayed at the end of swing	Difference in x coordinate between the wrist and the index knuckle	Feature value in the moment of interest
Whether the subject is moving in a figure 8 pattern, shifting then turning the body throughout the movement	Dynamic time warping between the figure 8 pattern of the subject and an 8 curve	Distance measure
Whether the subject is shifting the weight forward as the arms move, and shifting back when the arms are still throughout the movement	Evaluation of the pelvis shift in the sagittal plane	Pearson Correlation Coefficient between the left hip x trajectory and wrist x trajectory
Whether the hips and the shoulders are turning synchronously during the swing	Evaluation of the synchronicity of hips and shoulder movements	Pearson Correlation Coefficient between the hips x and y trajectories and the shoulders x and y trajectories
Whether the elbow bends to return the wrist to the body center between the end of the swing and the end of the repetition	Observation of the elbow angle time series	Mean, median, standard deviation and variance of the time series
Whether the elbow sinks to return the wrist to the body center between the end of the swing and the end of the repetition	Observation of the elbow y-trajectory time series	Mean, median, standard deviation and variance of the time series
Whether the head is centered with respect to the trunk throughout the movement	Observation of the difference between the x-trajectory of the head and the x-trajectory of the midpoint between the shoulders	Mean, median, standard deviation and variance of the time series

Table A.2: Look-up table for the Grasp the sparrow's tail exercise.

Aspect of movement	Measurement	Extracted values
Whether the left or right wrist is at collarbone height at its peak height	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the head is centered with respect to the trunk throughout the movement	Observation of the difference between the x-trajectory of the head and the x-trajectory of the midpoint between the shoulders	Mean, median, standard deviation and variance of the time series
Whether the hips and the shoulders are aligned	Evaluation of the ratio between the slope of the shoulders markers and the slope of the hips markers	Mean, median, standard deviation and variance of the time series
Whether the subject is moving in a figure 8 pattern, shifting then turning the body throughout the movement	Dynamic time warping between the figure 8 pattern of the subject and an 8 curve	Distance measure
Whether the subject is shifting the weight forward as the arms move, and shifting back when the arms are still throughout the movement	Evaluation of the pelvis shift in the sagittal plane	Pearson Correlation Coefficient between the left and right hip x trajectory and the respective wrist and elbow x trajectories
Whether the left or right foot is parallel in the foot stance	Ratio of the difference in y coordinate of the left or right ankle and the respective toe and the difference in x coordinate of the left or right ankle and the respective toe	Mean, median, standard deviation and variance of the time series
Whether the circular pattern of the wrists is smooth and symmetrical	Estimation of the curvature of the left and right wrist trajectory	Mean, median, standard deviation, variance, minimum, maximum and range of the time series
Whether the hand movement on the left side of the body matches the movement on the right side	Evaluation of the wrists trajectories during the repetition	Pearson Correlation Coefficient between the wrists x trajectories
Whether the elbows are at the same height	Observation of the difference in y coordinate of the elbows	Mean, median, standard deviation and variance of the time series

Table A.3: Look-up table for the Wave the hands like clouds exercise.

Aspect of movement	Measurement	Extracted values
Whether the left or right wrist is at collarbone height at the start of the push	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the left or right wrist is at collarbone height at the end of the push	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the left or right wrist move towards the hips in the middle of the push	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the subject shifts the body forwards for the first half of the exercise	Evaluation of the most visible hip x displacement during the push phase	Mean, median, standard deviation and variance of the time series
Whether the subject shifts the body backwards for the second half of the exercise	Evaluation of the most visible hip x displacement from the end of the push to the end of the repetition	Mean, median, standard deviation and variance of the time series
Whether the arm extension movement is synchronous with the leg shift movement	Observation of the wrist and the hip x trajectory	Pearson Correlation Coefficient between the hip x trajectory and wrist x trajectory
Whether the head is unnecessarily bent forwards or backwards throughout the movement	Observation of the difference between the x-trajectory of the head and the x-trajectory of the midpoint between the shoulders	Mean, median, standard deviation and variance of the time series
Assessment of the weight in the front leg during the push by looking at whether the knee is in front of the ankle, but no further than the base of the toes	Difference in x coordinate between the front leg's toe, knee and ankle and knee and ankle only during the push portion of the movement	Mean, median, standard deviation and variance of the time series
Assessment of the weight in the back leg during the push by looking at whether the knee is in behind the ankle and the toes	Difference in x coordinate between the back leg's toe, knee and ankle and knee and ankle only during the push portion of the movement	Mean, median, standard deviation and variance of the time series
Whether the subject is leaning back throughout the movement	Evaluation of the ratio between the slope of the shoulders markers and the slope of the hips markers	Mean, median, standard deviation and variance of the time series

Table A.4: Look-up table for the Withdraw and push exercise.

Aspect of movement	Measurement	Extracted values
Whether the front wrist is at collarbone height at the beginning of the repetition	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the back wrist is at pelvis height at the beginning of the repetition	Difference in y coordinate between the analyzed wrist and the respective hip	Feature value in the moment of interest
Whether the left or right wrist is at collarbone height in the middle of the repetition	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the front wrist is at pelvis height at the end of the repetition	Difference in y coordinate between the analyzed wrist and the respective hip	Feature value in the moment of interest
Whether the left or right wrist is at collarbone height at the end of the push	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the back wrist is at collarbone height at the end of the repetition	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the head is unnecessarily bent from side to side	Observation of the neck angle time series throughout the repetition	Mean, median, standard deviation and variance of the time series
Whether the head is centered with respect to the trunk throughout the movement	Observation of the difference between the x-trajectory of the head and the x-trajectory of the midpoint between the shoulders	Mean, median, standard deviation and variance of the time series
Whether the hips and the shoulders are aligned	Evaluation of the ratio between the slope of the shoulders markers and the slope of the hips markers	Mean, median, standard deviation and variance of the time series
Whether the weight shifts laterally from one leg to the other from the beginning of the repetition to its middle and from the middle of the repetition to its end	Observation of the x displacement of the pelvis	Mean, median, standard deviation and variance of the time series

Table A.5: Look-up table for the Brush knee twist step exercise.

Aspect of movement	Measurement	Extracted values
Whether the raised elbow is centered over the knee at the peak of the knee	Difference in x coordinate between the analyzed elbow and the respective knee	Feature value in the moment of interest
How far apart the raised elbow is from the raised knee at the peak of the knee	Difference in y coordinate between the analyzed elbow and the respective knee	Feature value in the moment of interest
Whether the raised wrist is at collarbone height at the peak of the knee	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the raised ankle is aligned to the respective knee at the peak of the knee	Difference in x coordinate between the analyzed ankle and the respective knee	Feature value in the moment of interest
Whether the raised wrist is aligned to the respective elbow at the peak of the knee	Difference in x coordinate between the analyzed wrist and the respective elbow	Feature value in the moment of interest
Whether the subject is stable when raising the knee	Evaluation of the navel trajectory from the beginning of the repetition to the peak of the knee	Mean, median, standard deviation, variance and range of the navel displacement, mean of the velocity, mean of the acceleration, mean of the jerk both for the x and in the y trajectories.
Whether the arm being raised moves in a medial trajectory when raising the knee	Observation of the trajectory of the wrist being raised from the beginning of the repetition to the peak of the knee	Mean, median, standard deviation, variance and range of the time series
Whether the head is unnecessarily bent from side to side	Observation of the neck angle time series throughout the repetition	Mean, median, standard deviation and variance of the time series
Whether the head is centered with respect to the trunk throughout the movement	Observation of the difference between the x-trajectory of the head and the x-trajectory of the midpoint between the shoulders	Mean, median, standard deviation and variance of the time series
Whether the hips and the shoulders are aligned	Evaluation of the ratio between the slope of the shoulders markers and the slope of the hips markers	Mean, median, standard deviation and variance of the time series

Table A.6: Look-up table for the Golden rooster exercise.

Aspect of movement	Measurement	Extracted values
Whether the left or right is at collarbone height at the peak of the wrists	Difference in y coordinate between the analyzed wrist and the respective shoulder	Feature value in the moment of interest
Whether the left or right is at pelvis height at the beginning of the repetition	Difference in y coordinate between the analyzed wrist and the respective hip	Feature value in the moment of interest
Whether the knees are excessively bent at the peak of the wrists	Evaluation of the knee angle	Feature value in the moment of interest
Whether the overall movement is smooth and synchronised	Observation of the hips and wrists trajectories	Pearson Correlation Coefficient between the x trajectory of the hip and of the respective wrist
Whether the head is unnecessarily bent from side to side	Observation of the neck angle time series throughout the repetition	Mean, median, standard deviation and variance of the time series
Whether the hips are consistently aligned throughout the movement	Observation of the slope of the hips	Mean, median, standard deviation, variance and range of the time series
Whether the hips and the shoulders are aligned	Evaluation of the ratio between the slope of the shoulders markers and the slope of the hips markers	Mean, median, standard deviation and variance of the time series
Whether the hips and the knees are aligned	Evaluation of the ratio between the slope of the knees markers and the slope of the hips markers	Mean, median, standard deviation and variance of the time series
Whether the hands trajectories are symmetric throughout the movement	Observation of the distance of the wrists from the body midline	Mean, median, standard deviation and variance of the time series
Whether the head is centered with respect to the trunk throughout the movement	Observation of the difference between the x-trajectory of the head and the x-trajectory of the midpoint between the shoulders	Mean, median, standard deviation and variance of the time series

Table A.7: Look-up table for the Raising the power exercise.

Bibliography

- [1] Gelsy Torres-Oviedo and Lena H Ting. «Muscle synergies characterizing human postural responses». In: *Journal of neurophysiology* 98.4 (2007), pp. 2144–2156 (cit. on p. 1).
- [2] Giacomo Severini, Alexander Koenig, Catherine Adans-Dester, Iahn Cajigas, Vincent CK Cheung, and Paolo Bonato. «Robot-driven locomotor perturbations reveal synergy-mediated, context-dependent feedforward and feedback mechanisms of adaptation». In: *Scientific reports* 10.1 (2020), p. 5104 (cit. on p. 1).
- [3] Iahn Cajigas, Alexander Koenig, Giacomo Severini, Maurice Smith, and Paolo Bonato. «Robot-induced perturbations of human walking reveal a selective generation of motor adaptation». In: *Science Robotics* 2.6 (2017), eaam7749 (cit. on p. 1).
- [4] Sašo Jezernik, Gery Colombo, Thierry Keller, Hansruedi Frueh, and Manfred Morari. «Robotic orthosis lokomat: A rehabilitation and research tool». In: *Neuromodulation: Technology at the neural interface* 6.2 (2003), pp. 108–115 (cit. on p. 1).
- [5] Maria Rubega, Roberto Di Marco, Marianna Zampini, Emanuela Formaggio, Emanuele Menegatti, Paolo Bonato, Stefano Masiero, and Alessandra Del Felice. «Muscular and cortical activation during dynamic and static balance in the elderly: A scoping review». In: *Aging Brain* 1 (2021), p. 100013 (cit. on p. 2).
- [6] Tytus Wojtara, Fady Alnajjar, Shingo Shimoda, and Hidenori Kimura. «Muscle synergy stability and human balance maintenance». In: *Journal of neuroengineering and rehabilitation* 11.1 (2014), pp. 1–9 (cit. on p. 2).
- [7] Vincent CK Cheung, Andrea Turolla, Michela Agostini, Stefano Silvoni, Caoimhe Bennis, Patrick Kasi, Sabrina Paganoni, Paolo Bonato, and Emilio Bizzi. «Muscle synergy patterns as physiological markers of motor cortical damage». In: *Proceedings of the national academy of sciences* 109.36 (2012), pp. 14652–14656 (cit. on p. 2).

- [8] Jose Garcia Vivas Miranda et al. «Complex upper-limb movements are generated by combining motor primitives that scale with the movement size». In: *Scientific reports* 8.1 (2018), p. 12918 (cit. on p. 3).
- [9] Inc. Vicon Industries. *Vicon - Award Winning Motion Capture Systems*. URL: <http://www.vicon.com/> (visited on 09/2023) (cit. on pp. 3, 4).
- [10] SoftBank Robotics Group. *NAO: Personal Robot Teaching Assistant*. URL: <http://us.softbankrobotics.com/nao> (visited on 07/2023) (cit. on p. 3).
- [11] Kristen L Carroll, Jennifer Leiser, and Theodore S Paisley. «Cerebral palsy: physical activity and sport». In: *Current Sports Medicine Reports* 5.6 (2006), pp. 319–322 (cit. on p. 3).
- [12] Olaf Verschuren, Mark D Peterson, Astrid CJ Balemans, and Edward A Hurvitz. «Exercise and physical activity recommendations for people with cerebral palsy». In: *Developmental Medicine & Child Neurology* 58.8 (2016), pp. 798–808 (cit. on p. 3).
- [13] Dephy Inc. *Dephy - Beyond nature*. URL: <http://www.dephy.com/> (visited on 09/2023) (cit. on p. 4).
- [14] Advanced Mechanical Technology Inc. *Multi-Axis Force Plates - AMTI*. URL: <https://www.amti.biz/product-line/force-plates/> (visited on 09/2023) (cit. on pp. 4, 5).
- [15] Movella Holdings Inc. *Xsens - 3D Motion Tracking - 3D Motion Tracking Technology*. URL: <http://www.movella.com/products/xsens> (visited on 09/2023) (cit. on p. 4).
- [16] Reza Sadjadi, Neha Godbole, Richard Lewis, David Balkwill, and Florian Eichler. *Gait and Balance Outcomes in Adrenomyeloneuropathy (AMN)(P5.2-101)*. 2019 (cit. on p. 5).
- [17] Switzerland Kistler Instrumente AG Winterthur. *3D force plate*. URL: <https://www.kistler.com/US/en/3d-force-plate/C00000090> (visited on 09/2023) (cit. on p. 5).
- [18] Ewa Rudnicka, Paulina Napierała, Agnieszka Podfigurna, Błażej Męczekalski, Roman Smolarczyk, and Monika Grymowicz. «The World Health Organization (WHO) approach to healthy ageing». In: *Maturitas* 139 (2020), pp. 6–11 (cit. on p. 7).
- [19] Giuseppe Francesco Papalia, Rocco Papalia, Lorenzo Alirio Diaz Balzani, Guglielmo Torre, Biagio Zampogna, Sebastiano Vasta, Chiara Fossati, Anna Maria Alifano, and Vincenzo Denaro. «The effects of physical exercise on balance and prevention of falls in older people: A systematic review and meta-analysis». In: *Journal of clinical medicine* 9.8 (2020), p. 2595 (cit. on p. 7).

- [20] Le-Cong Wang, Ming-Zhu Ye, Jian Xiong, Xiao-Qian Wang, Jia-Wei Wu, and Guo-Hua Zheng. «Optimal exercise parameters of tai chi for balance performance in older adults: A meta-analysis». In: *Journal of the American Geriatrics Society* 69.7 (2021), pp. 2000–2010 (cit. on pp. 7, 8, 10).
- [21] Dongling Zhong et al. «Tai Chi for improving balance and reducing falls: An overview of 14 systematic reviews». In: *Annals of physical and rehabilitation medicine* 63.6 (2020), pp. 505–517 (cit. on pp. 7, 8, 10).
- [22] Peter M Wayne et al. «A systems biology approach to studying Tai Chi, physiological complexity and healthy aging: design and rationale of a pragmatic randomized controlled trial». In: *Contemporary clinical trials* 34.1 (2013), pp. 21–34 (cit. on p. 8).
- [23] Ching Lan, Jin-Shin Lai, and Ssu-Yuan Chen. «Tai Chi Chuan: an ancient wisdom on exercise and health promotion». In: *Sports medicine* 32 (2002), pp. 217–224 (cit. on pp. 8, 10).
- [24] Guo-Yan Yang et al. «Tai Chi for health and well-being: a bibliometric analysis of published clinical studies between 2010 and 2020». In: *Complementary therapies in medicine* 60 (2021), p. 102748 (cit. on p. 9).
- [25] Guo-Yan Yang et al. «Evidence base of clinical studies on Tai Chi: a bibliometric analysis». In: *PloS one* 10.3 (2015), e0120655 (cit. on p. 10).
- [26] Patricia Huston and Bruce McFarlane. «Health benefits of tai chi: what is the evidence?» In: *Canadian Family Physician* 62.11 (2016), pp. 881–890 (cit. on p. 10).
- [27] Alice MK Wong and Ching Lan. «Tai Chi and balance control». In: *Tai Chi Chuan* 52 (2008), pp. 115–123 (cit. on p. 10).
- [28] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. «Deep 3D human pose estimation: A review». In: *Computer Vision and Image Understanding* 210 (2021), p. 103225 (cit. on p. 11).
- [29] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. «Microsoft coco: Common objects in context». In: *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*. Springer. 2014, pp. 740–755 (cit. on p. 11).
- [30] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. «2D Human Pose Estimation: New Benchmark and State of the Art Analysis». In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2014 (cit. on pp. 11, 15).

- [31] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. «Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments». In: *IEEE transactions on pattern analysis and machine intelligence* 36.7 (2013), pp. 1325–1339 (cit. on p. 11).
- [32] Leonid Sigal, Alexandru O Balan, and Michael J Black. «HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion». In: *International journal of computer vision* 87.1-2 (2010), pp. 4–27 (cit. on p. 11).
- [33] Yann Desmarais, Denis Mottet, Pierre Slangen, and Philippe Montesinos. «A review of 3D human pose estimation algorithms for markerless motion capture». In: *Computer Vision and Image Understanding* 212 (2021), p. 103275 (cit. on pp. 11–13).
- [34] Tewodros Legesse Munea, Yalew Zelalem Jembre, Halefom Tekle Weldegebriel, Longbiao Chen, Chenxi Huang, and Chenhui Yang. «The progress of human pose estimation: A survey and taxonomy of models applied in 2D human pose estimation». In: *IEEE Access* 8 (2020), pp. 133330–133348 (cit. on p. 11).
- [35] Jan Stenum, Kendra M Cherry-Allen, Connor O Pyles, Rachel D Reetzke, Michael F Vignos, and Ryan T Roemmich. «Applications of pose estimation in human health and performance across the lifespan». In: *Sensors* 21.21 (2021), p. 7315 (cit. on p. 11).
- [36] Alexander Toshev and Christian Szegedy. «DeepPose: Human pose estimation via deep neural networks». In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1653–1660 (cit. on pp. 12, 14).
- [37] Sam Johnson and Mark Everingham. «Clustered pose and nonlinear appearance models for human pose estimation.» In: *bmvc*. Vol. 2. 4. Aberystwyth, UK. 2010, p. 5 (cit. on p. 14).
- [38] Eldar Insafutdinov, Leonid Pishchulin, Bjoern Andres, Mykhaylo Andriluka, and Bernt Schiele. «DeepCUT: A deeper, stronger, and faster multi-person pose estimation model». In: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VI 14*. Springer. 2016, pp. 34–50 (cit. on p. 14).
- [39] Leonid Pishchulin, Eldar Insafutdinov, Siyu Tang, Bjoern Andres, Mykhaylo Andriluka, Peter V Gehler, and Bernt Schiele. «DeepCUT: Joint subset partition and labeling for multi person pose estimation». In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 4929–4937 (cit. on p. 14).

- [40] Eléonore G. Duroyon. «Using deep learning-based video analysis techniques to estimate clinical scores for stroke rehabilitation». Master's Degree thesis. Swiss Federal Institute of Technology (ETH) Zurich, June 2022 (cit. on pp. 14, 15).
- [41] Alexandre Fabian A Saint, Abd El Rahman Shabayek, Djamila Aouada, Björn Ottersten, Kseniya Cherenkova, and Gleb Gusev. «Towards automatic human body model fitting to a 3d scan». In: *Proceedings of 3DBODY. TECH 2017-8th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, Montreal QC, Canada, 11-12 Oct. 2017*. Hometrica Consulting. 2017, pp. 274–280 (cit. on p. 14).
- [42] Longhui Wei, Xiaobin Liu, Jianing Li, and Shiliang Zhang. «VP-ReID: Vehicle and person re-identification system». In: *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*. 2018, pp. 501–504 (cit. on p. 14).
- [43] Man Liang and Yingrui Hu. «Application of Human Body Posture Recognition Technology in Robot Platform for Nursing Empty-Nesters». In: *2020 6th International Conference on Control, Automation and Robotics (ICCAR)*. 2020, pp. 91–95. DOI: 10.1109/ICCAR49639.2020.9108070 (cit. on p. 14).
- [44] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. «RMPE: Regional multi-person pose estimation». In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2334–2343 (cit. on p. 15).
- [45] Heilym Ramirez, Sergio A Velastin, Ernesto Fabregas, Ignacio Meza, Dimitrios Makris, and Gonzalo Farias. «Fall detection using human skeleton features». In: (2021) (cit. on p. 15).
- [46] Hongtao Zheng and Yan Liu. «Lightweight fall detection algorithm based on AlphaPose optimization model and ST-GCN». In: *Mathematical Problems in Engineering 2022 (2022)* (cit. on p. 15).
- [47] Jingqi MA, Huan LEI, and Minyi CHEN. «Fall behavior detection algorithm for the elderly based on AlphaPose optimization model». In: *Journal of Computer Applications* 42.1 (2022), p. 294 (cit. on p. 15).
- [48] Xiaodong Zhao, Fanxing Hou, Jingfang Su, and Lane Davis. «An Alphapose-Based Pedestrian Fall Detection Algorithm». In: *International Conference on Adaptive and Intelligent Systems*. Springer. 2022, pp. 650–660 (cit. on p. 15).
- [49] Anitha Rani Inturi, VM Manikandan, and Vignesh Garrapally. «A novel vision-based fall detection scheme using keypoints of human skeleton with long short-term memory network». In: *Arabian Journal for Science and Engineering* 48.2 (2023), pp. 1143–1155 (cit. on p. 15).

- [50] Yuan Tian, Ning Li, and Bo Mao. «Video Behavior Detection based on Optimized Alphapose in Electricity Facility Management». In: *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. 2021, pp. 364–367 (cit. on p. 15).
- [51] Jan Docekal, Jakub Rozlivek, Jiri Matas, and Matej Hoffmann. «Human keypoint detection for close proximity human-robot interaction». In: *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. IEEE. 2022, pp. 450–457 (cit. on p. 15).
- [52] Xinyu Lv, Na Ta, Tao Chen, Jing Zhao, Haicheng Wei, et al. «Analysis of Gait Characteristics of Patients with Knee Arthritis Based on Human Posture Estimation». In: *BioMed Research International 2022 (2022)* (cit. on p. 15).
- [53] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. «Realtime multi-person 2d pose estimation using part affinity fields». In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7291–7299 (cit. on p. 15).
- [54] Aditya Viswakumar, Venkateswaran Rajagopalan, Tathagata Ray, and Chandu Parimi. «Human Gait Analysis Using OpenPose». In: *2019 Fifth International Conference on Image Information Processing (ICIIP)*. 2019, pp. 310–314. DOI: 10.1109/ICIIP47207.2019.8985781 (cit. on p. 16).
- [55] Erika D’Antonio, Juri Taborri, Eduardo Palermo, Stefano Rossi, and Fabrizio Patanè. «A markerless system for gait analysis based on OpenPose library». In: *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. 2020, pp. 1–6. DOI: 10.1109/I2MTC43012.2020.9128918 (cit. on p. 16).
- [56] Weiming Chen, Zijie Jiang, Hailin Guo, and Xiaoyang Ni. «Fall detection based on key points of human-skeleton using openpose». In: *Symmetry* 12.5 (2020), p. 744 (cit. on p. 16).
- [57] Chuan-Bi Lin, Ziqian Dong, Wei-Kai Kuan, and Yung-Fa Huang. «A framework for fall detection based on OpenPose skeleton and LSTM/GRU models». In: *Applied Sciences* 11.1 (2020), p. 329 (cit. on p. 16).
- [58] Farzan Majeed Noori, Benedikte Wallace, Md Zia Uddin, and Jim Torresen. «A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network». In: *Scandinavian conference on image analysis*. Springer. 2019, pp. 299–310 (cit. on p. 16).
- [59] Woojoo Kim, Jaeho Sung, Daniel Saakes, Chunxi Huang, and Shuping Xiong. «Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose)». In: *International Journal of Industrial Ergonomics* 84 (2021), p. 103164 (cit. on p. 16).

- [60] Kehan Chen. «Sitting posture recognition based on openpose». In: *IOP Conference Series: Materials Science and Engineering*. Vol. 677. 3. IOP Publishing. 2019, p. 032057 (cit. on p. 16).
- [61] Brighton Li, James Williamson, Nicole Kelp, Taylor Dick, and Antonio P. L. Bo. «Towards balance assessment using Openpose». In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. 2021, pp. 7605–7608. DOI: 10.1109/EMBC46164.2021.9631001 (cit. on p. 16).
- [62] Fritz Webering, Holger Blume, and Issam Allaham. «Markerless camera-based vertical jump height measurement using OpenPose». In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 3868–3874 (cit. on p. 16).
- [63] Yung-Che Li, Ching-Tang Chang, Chin-Chang Cheng, and Yu-Len Huang. «Baseball Swing Pose Estimation Using OpenPose». In: *2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*. 2021, pp. 6–9. DOI: 10.1109/RAAI52226.2021.9507807 (cit. on p. 16).
- [64] Camillo Lugaresi et al. «Mediapipe: A framework for building perception pipelines». In: *arXiv preprint arXiv:1906.08172* (2019) (cit. on pp. 16, 27–29).
- [65] Camillo Lugaresi et al. «Mediapipe: A framework for perceiving and processing reality». In: *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*. Vol. 2019. 2019 (cit. on p. 16).
- [66] Jen-Li Chung, Lee-Yeng Ong, and Meng-Chew Leow. «Comparative analysis of skeleton-based human pose estimation». In: *Future Internet* 14.12 (2022), p. 380 (cit. on p. 16).
- [67] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. «Blazepose: On-device real-time body pose tracking». In: *arXiv preprint arXiv:2006.10204* (2020) (cit. on pp. 16, 30).
- [68] Ira Nath, Aditya Shaw, Aniruddha Bhadra, Diya Dey Bhowmick, Sayandeep Ghosh, and Sreeja Chatterjee. «A Novel Personal Fitness Trainer and Tracker powered by Artificial Intelligence enabled by MEDIAPIPE and OpenCV». In: *International Journal of Intelligent Systems and Applications in Engineering* 11.3 (2023), pp. 1085–1094 (cit. on p. 17).

- [69] Vedangi Agarwal, Konark Sharma, and Abha Kiran Rajpoot. «AI based Yoga Trainer - Simplifying home yoga using mediapipe and video streaming». In: *2022 3rd International Conference for Emerging Technology (INCET)*. 2022, pp. 1–5. DOI: 10.1109/INCET54531.2022.9824332 (cit. on p. 17).
- [70] Ardra Anilkumar, Athulya KT, Sarath Sajan, and Sreeja KA. «Pose estimated yoga monitoring system». In: *Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems-ICICNIS*. 2021 (cit. on p. 17).
- [71] Shubham Garg, Aman Saxena, and Richa Gupta. «Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application». In: *Journal of Ambient Intelligence and Humanized Computing* (2022), pp. 1–12 (cit. on p. 17).
- [72] Charles Andrew Q Bugarin, Juan Miguel M Lopez, Scud Gabriel M Pineda, Ma Franzeska C Sambrano, and Pocholo James M Loresco. «Machine vision-based fall detection system using mediapipe pose with iot monitoring and alarm». In: *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE. 2022, pp. 269–274 (cit. on p. 17).
- [73] Thiago Buarque de Gusmão Lafayette, Victor Hugo de Lima Kunst, Pedro Vanderlei de Sousa Melo, Paulo de Oliveira Guedes, João Marcelo Xavier Natário Teixeira, Cmthia Rodrigues de Vasconcelos, Veronica Teichrieb, and Alana Elza Fontes da Gama. «Validation of angle estimation based on body tracking data from RGB-D and RGB cameras for biomechanical assessment». In: *Sensors* 23.1 (2022), p. 3 (cit. on p. 17).
- [74] Yasutaka Uchida and Tomoko Funayama. «Possibility of Gait Analysis with MediaPipe and Its Application in Evaluating the Effects of Gait-assist Devices». In: *International Journal on Advances in Life Sciences Volume 15, Number 1 & 2, 2023* (2023) (cit. on p. 17).
- [75] Ameer Latreche, Ridha Kelaiaia, Ahmed Chemori, and Adlen Kerboua. «Reliability and validity analysis of MediaPipe-based measurement system for some human rehabilitation motions». In: *Measurement* 214 (2023), p. 112826 (cit. on p. 17).
- [76] Jingdong Wang et al. «Deep high-resolution representation learning for visual recognition». In: *IEEE transactions on pattern analysis and machine intelligence* 43.10 (2020), pp. 3349–3364 (cit. on p. 17).
- [77] Tensorflow. *MoveNet: Ultra fast and accurate pose detection model*. URL: <https://www.tensorflow.org/hub/tutorials/movenet?hl=it> (visited on 09/2023) (cit. on p. 17).

- [78] Tensorflow. *Pose estimation and classification on edge devices with MoveNet and TensorFlow Lite*. URL: <https://blog.tensorflow.org/2021/08/pose-estimation-and-classification-on-edge-devices-with-MoveNet-and-TensorFlow-Lite.html> (visited on 09/2023) (cit. on p. 17).
- [79] BeomJun Jo and SeongKi Kim. «Comparative analysis of OpenPose, PoseNet, and MoveNet models for pose estimation in mobile devices». In: *Traitement du Signal* 39.1 (2022), p. 119 (cit. on p. 18).
- [80] Sheetal Girase, Omkar Dutta, Adwait Mahadar, Atharva Ghodmare, and Mangesh Bedekar. «Identifying Incorrect Postures While Performing Sun Salutation Using MoveNet». In: *International Conference on Communication and Intelligent Systems*. Springer. 2022, pp. 575–587 (cit. on p. 18).
- [81] Zixin Cai. «PoseBuddy: workout platform with MoveNet». In: (2022) (cit. on p. 18).
- [82] Zhadra Kozhamkulova, Bibinur Kirgizbayeva, Gulbakyt Sembina, Ulmeken Smailova, Madina Suleimenova, Arailym Keneskanova, and Zhumakul Baizakova. «MoveNET Enabled Neural Network for Fast Detection of Physical Bullying in Educational Institutions». In: *International Journal of Advanced Computer Science and Applications* 14.5 (2023) (cit. on p. 18).
- [83] Sven Kreiss, Lorenzo Bertoni, and Alexandre Alahi. «Openpipaf: Composite fields for semantic keypoint detection and spatio-temporal association». In: *IEEE Transactions on Intelligent Transportation Systems* 23.8 (2021), pp. 13498–13511 (cit. on p. 18).
- [84] Younes Belkada, Lorenzo Bertoni, Romain Caristan, Taylor Mordan, and Alexandre Alahi. «Do pedestrians pay attention? eye contact detection in the wild». In: *arXiv preprint arXiv:2112.04212* (2021) (cit. on p. 18).
- [85] Maximilian Pröll. «Pedestrian Crossing Prediction in the Context of Autonomous Driving». Diploma Thesis. Faculty of Informatics, Vienna University of Technology, May 2022 (cit. on p. 18).
- [86] Marina Moseva, Ksenia Polyantseva, Mikhail Gorodnichev, and Artem Pavlikov. «Algorithm for Predicting Pedestrian Behavior on Public Roads». In: *2023 Systems of Signals Generating and Processing in the Field of on Board Communications*. IEEE. 2023, pp. 1–6 (cit. on p. 18).
- [87] Bilal Abdulrahman and Zhigang Zhu. «Real-time pedestrian pose estimation, tracking and localization for social distancing». In: *Machine vision and applications* 34.1 (2023), p. 8 (cit. on p. 18).
- [88] Andrea Valerio. «Development of Machine Learning-Based Algorithms for Assessing Tai Chi Exercise Proficiency in Older Adults». Master’s Degree thesis. Politecnico di Torino, Mar. 2020 (cit. on pp. 21, 23–27).

- [89] Peter Wayne and Mark Fuerst. *The Harvard Medical School Guide to Tai Chi*. Shambhala, 2013 (cit. on pp. 21–26).
- [90] Salena B. Prakah-Asante. «Automated Interpretable Assessment of Tai Chi Exercise Proficiency with Machine Learning, Biomechanics and Data Visualization». Bachelor’s Degree thesis. Harvard John A. Paulson School of Engineering, Mar. 2023 (cit. on p. 27).
- [91] Google. *MediaPipe Pose*. URL: <https://github.com/google/mediapipe/blob/master/docs/solutions/pose.md> (visited on 03/2023) (cit. on pp. 30, 31).
- [92] Google. *Pose landmark detection guide*. URL: https://developers.google.com/mediapipe/solutions/vision/pose_landmarker (visited on 05/2023) (cit. on p. 30).
- [93] Liliana Lo Presti and Marco La Cascia. «3D skeleton-based human action classification: A survey». In: *Pattern Recognition* 53 (2016), pp. 130–147 (cit. on p. 31).
- [94] MathWorks. *xcorr*. URL: <https://it.mathworks.com/help/matlab/ref/xcorr.html> (visited on 03/2023) (cit. on p. 33).
- [95] Chris Ding and Hanchuan Peng. «Minimum redundancy feature selection from microarray gene expression data». In: *Journal of bioinformatics and computational biology* 3.02 (2005), pp. 185–205 (cit. on p. 35).
- [96] Smazzanti et al. *mRMR. Efficient Feature Selection*. URL: <https://github.com/smazzanti/mrmr> (visited on 05/2023) (cit. on p. 35).
- [97] John W Sammon. «A nonlinear mapping for data structure analysis». In: *IEEE Transactions on computers* 100.5 (1969), pp. 401–409 (cit. on pp. 35, 36).
- [98] Pollard et al. *Sammon mapping in Python*. URL: <https://github.com/tompollard/sammon> (visited on 05/2023) (cit. on pp. 35, 36).
- [99] Chao Chen, Andy Liaw, Leo Breiman, et al. «Using random forest to learn imbalanced data». In: *University of California, Berkeley* 110.1-12 (2004), p. 24 (cit. on pp. 35, 37).
- [100] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. «Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning». In: *Journal of Machine Learning Research* 18.17 (2017), pp. 1–5. URL: <http://jmlr.org/papers/v18/16-365> (cit. on pp. 35, 37).
- [101] F. Pedregosa et al. «Scikit-learn: Machine Learning in Python». In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on pp. 36, 37).