Capturing, Modelling, Analyzing and providing Feedback in Martial Arts with Artificial Intelligence to support Psychomotor Learning Activities

Autor: Alberto Casas Ortiz

Directora: Olga C. Santos

UNED

ETS de Ingeniería Informática

Universidad Nacional de Educación a Distancia (UNED)
Escuela Técnica Superior de Ingeniería Informática

Trabajo de Fin de Máster del Máster Universitario
en Inteligencia Artificial Avanzada

Junio 2020
Madrid, España

ACKNOWLEDGEMENTS

I would like to express my gratitude to the director of this Master's Thesis, for her guidance and valuable suggestions during the planning and development of this research work.

I would also like to thank my Master of American Kenpo Karate and my training partners for all the knowledge that I have obtained from them.

My grateful thanks are also extended to my girlfriend for supporting me during this research project and helping me testing the applications.

Finally, I would also like to extend my special thanks to my father, my mother, my grandmother and my sister for their support and encouragement throughout this research project.

I would like to dedicate this work to my grandfather. Wherever you are, this is for you.

ABSTRACT

This Master's Thesis explores how Artificial Intelligence (IA) can assist in the learning of psychomotor activities, specifically the learning of a martial art, through the development of an AI-based application that executes in an Android device. Martial arts are an interesting domain for this because it encompasses most of the characteristics that can be found in other psychomotor activities. Different methods for capturing, modelling and analyzing human motion, as well as providing feedback to the user have been reviewed. In addition, another bibliographical review of 27 publications has been carried out to evaluate till what extend these methods have been already applied to martial arts. For this research work, inertial methods have been selected for capturing motion. In particular, the inertial sensors of an Android device have been used for capturing the execution of a set of movements of American Kenpo Karate from 20 volunteers. The captured data was then modeled, by segmenting and labelling the movements, and smoothing the time series using Exponentially Weighted Moving Averages. The resulting dataset, formed by 240 movements, was then used for training and comparing three neural network-based classifiers: FC-ANN, 1D-CNN and LSTM. Neural networks were selected because of their ability of learn complex functions and the fact that some neural network architectures have been created specifically for analyze time series. Further, the weights learned by a neural network can be transferred to other domains through the technique known as transfer learning. Obtained results suggest that LSTM is the type of neural network that can better classify the movements studied, obtaining an accuracy of 1.0 in the training set, and an accuracy of 0.94 in the testing set. For demonstrating that those methods can be applied, an AI-based real-time Android application has been developed. This application employs the studied methods, as well as a feedback strategy created using the results of a questionnaire carried out with the purpose of identifying the issues that online learning of a psychomotor activity can entails. The application has then been tested, generating a good impression in the users. Following the open science philosophy, all contributions are shared in the GitHub repository.

Keywords: Martial Arts, Human Motion Capture, Human Motion Modeling, Human motion Analysis, Sequence Analysis, Feedback Strategies, American Kenpo Karate

RESUMEN

Este Trabajo de Fin de Máster explora las maneras en las que la Inteligencia Artificial (IA) puede asistir en el aprendizaje de actividades motoras, concretamente en el aprendizaje de un arte marcial, a través del desarrollo de una aplicación Android que utilice técnicas de IA. Las artes marciales son un dominio interesante para esto dado que engloban muchas de las características que se pueden encontrar en una actividad psicomotora. Diferentes métodos para capturar, modelar y analizar movimiento, así como proveer feedback a los usuarios, han sido revisados aquí. Además, se ha realizado otra revisión bibliográfica de 27 publicaciones para evaluar hasta donde se han llegado a aplicar estos métodos en artes marciales. En este trabajo de investigación, se han elegido métodos inerciales para capturar movimiento. En particular, los sensores inerciales de un dispositivo Android han sido usados para capturar la ejecución de un set de movimientos propio del Kenpo Karate Americano, hecha por 20 voluntarios. Los datos capturados han sido modelados, separando y etiquetando cada movimiento, y las secuencias temporales capturadas han sido suavizadas usando el algoritmo conocido como Exponentially Weighted Moving Averages. El dataset resultante, formado por 240 movimientos, ha sido usado para entrenar y comparar tres clasificadores basados en el uso de redes neuronales: FC-ANN, 1D-CNN y LSTM. Se han elegido las redes neuronales porque tienen una gran capacidad para aprender funciones complejas, y por el hecho de que algunas arquitecturas de redes neuronales han sido creadas específicamente para analizar secuencias temporales. Además, los pesos aprendidos por una red neuronal se pueden transferir a otros dominios mediante la técnica conocida como transfer learning. Los resultados obtenidos sugieren que LSTM es el tipo de red neuronal que mejor clasifica los movimientos estudiados, obteniendo un número de aciertos de 1.0 en el set de entrenamiento, e igual a 0.94 en el de validación. Para demostrar que estos métodos pueden ser aplicados, se ha desarrollado una aplicación Android que funciona en tiempo real y utiliza los algoritmos de IA mencionados como base. Se ha desarrollado también una estrategia para proveer feedback al usuario, utilizando los resultados de un cuestionario que se ha llevado a cabo con el propósito de identificar los problemas que puede conllevar el aprendizaje online de una actividad psicomotora. Esta aplicación ha sido probada, generando una buena impresión en los usuarios. Siguiendo la filosofía open science (ciencia abierta), todas las contribuciones han sido compartidas en GitHub.

Palabras Clave: Artes Marciales, Captura de Movimiento Humano, Modelado de Movimiento Humano, Análisis de movimiento Humano, Análisis de Secuencias, Estrategias de Feedback, Kenpo Karate Americano

I GLOSSARY

3D Skeleton	A 3D representation of the joints of the human body.
Accelerometer	Device used for measure acceleration in the three spatial axes.
Aikido	Japanese martial art created by the grand master Morihei Ueshiba that focus in the optimal use of forces during a confrontation. This martial art is based in scientific principles and have a great traditional component.
American Kenpo Karate	American martial art founded by the senior grand master Ed. K. Parker that focus in self-defense. This is a modern martial art based in scientific principles, that also have a traditional component.
Android	An operative system for smartphones.
Android Device	A device that is running the android operative system.
Artificial Intelligence (AI)	Discipline that studies how to provide characteristics of an intelligent being to machines.
Auditory feedback	Feedback obtained through the sense of hearing.
Augmented Feedback	Feedback given to a subject by an external source. This external source can have many forms such as an external observer or a device.
Augmented Reality	Discipline that studies how to give a user the sensation of a virtual component being present in the real world.
Blocking Set I	A set of defensive movements (blocks) taught in the curriculum of American Kenpo Karate.
Dojo	School or gym dedicated to the practice of a martial art.
Extrinsic Feedback	Feedback given to a subject by his/her own internal senses.
Form	Alternative name for kata in English used in some martial arts like American Kenpo Karate.
Gamification	A set of techniques used for giving characteristics of a videogame to something that is not a videogame.
Gyroscope	Device used for measure rotation in the three spatial axes.
Haptic feedback	Feedback obtained through the sense of touch.
Jeet Kune Do	American martial art founded by the famous martial artist Bruce Lee, focusing in getting rid of the traditional part of a martial art and keeping the functional.
Kata	An established sequence of movements used in some martial arts that simulates a combat against one or multiple opponents. Normally, the student practices a kata alone, against imaginary opponents.
Keras	API that ease the use of some machine learning backends such as Tensorflow.
Labanotation	A system used for recording motion that uses abstract symbols for representing direction and parts of the body.
Labanotation Score	A score used in Labanotation for recording motion.

Magnetometer	Device used for measuring the influence of the magnetic field of the earth in the three spatial axes.
Martial Art	Combat systems practiced with different purposes such as self-defense or physical training. Martial Arts usually have a great traditional or philosophical component.
Motion	Progression of an object from one point to another.
Motion analysis	A technique used for analyzing captured and modelled data with the purpose of extracting or inferring new information.
Motion capture technique	A technique used for capturing movement and storing it, normally in a digital form.
Motion modelling	A technique used for modelling motion captured data, with the purpose of making it more understandable, reusable or preparing it for analysis.
Motor Skill	A skill that involves motor activity.
Movement	Change of position over time. In this document we refer to movement as a movement generated by a part of the body.
Multimodal feedback	Feedback obtained from more than one modality such as visual, haptic or auditory.
Psychomotor activity	Physical activity that involves the use of cognitive functions.
Set	In American Kenpo Karate, a recompilation of movements from a specific topic (Blocks, strikes, kicks…)
Tensorflow	An open-source library of machine learning.
Virtual Reality	Discipline that studies how to give a user the sensation of being inside of a virtual environment.
Virtual sickness	Malaise experimented by people when using a virtual reality device due to perceiving a sensation of motion that is not real.
Visual feedback	Feedback obtained through the sense of sight.
Weapon	Object or device used with the purpose of harm others or self-defense. A weapon can be a part of the body, like a punch.

II FIGURES

III TABLES

CONTENTS

1 INTRODUCTION

1.1 Motivation: About Martial Arts and the Influence of Science

Martial arts are systems of combat, with or without weapons, that arose thousands of years ago in the five continents and have prevailed and evolved till our time. The expansion of these systems in Asian countries led to a strong fusion of martial arts with philosophy, politics and religion, and consequently, martial arts have now a big role in Asian countries [1]. Although the main purpose of martial arts was military and self-defense, the particular characteristics of martial arts have attracted researchers from diverse disciplines such as health sciences [2], psychology [3], business [4], videogames [5] or robotics [6].

Advances in science have influenced over the years the evolution of martial arts. We can find teachings of scientific principles like inertia or gravity in traditional martial arts like Aikido [7], which implies that students can learn basic principles from different fields such as physics or biomechanics [8], [9] in an intuitive and practical way. There are even martial arts built over scientific principles like American Kenpo Karate [10], which explicitly teach how to use inertia, gravity, torque or mass in their techniques.

The learning of a martial art is an arduous task that takes a lot of effort, time and involves learning principles and theories from different fields. People could believe that in a world of fire weapons, where people can safely wander around in many countries, martial arts are no necessary anymore for regular people. Although that is a true fact, there are a lot of reasons to practice and learn a martial art. There are a lot of studies about how a martial art not only involves self-defense, but several social [3], psychological [11] and health [2] benefits. Some studies even theorize about implanting martial arts in schools, and some countries have already implanted it [12], [13].

Traditionally, martial arts are taught person to person orally [1]. This could be because the learning of a martial art involves the learning of 3D movements, and that motion cannot be easily portrayed in 2D motionless images or texts. In Figure 1 we can see an example of how an aikido technique is captured on an image, using lines to show the flow of the movement [7]. The lack of dimension and motion makes difficult to learn and imitate those movements.

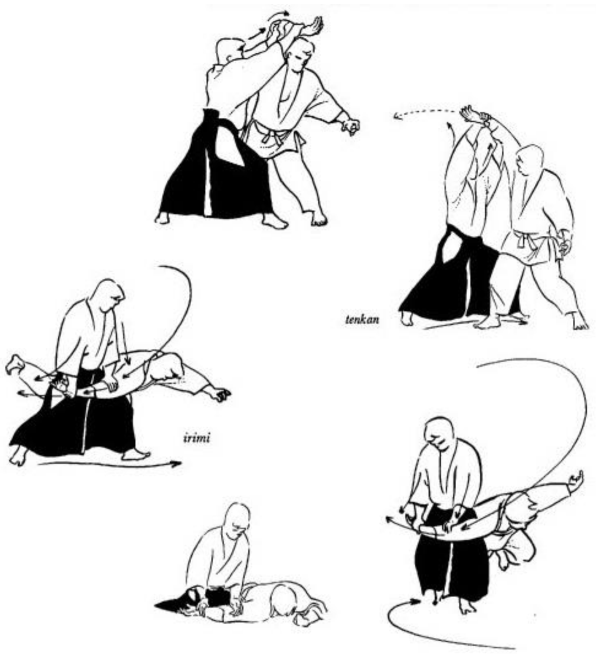

Figure 1. Explanation of an aikido technique. Image obtained from [7].

Even with oral tradition being the main way of transmitting the knowledge of a martial art, there are some martial arts books with great cultural value that teaches techniques, warfare and fighting methods, using or not weapons, in different parts of the world. In China, we can find "Jixiao Xinshu" (1560) [14], which covers different Chinese warfare techniques, written by general Qi Jiguang. In Japan, we can find The Book of Five Rings (1645) [15], which teaches the style of the legendary samurai Miyamoto Musashi. In Spain there are a lot of books about fencing, including "El Palo y el Sable" (1851) [16], which shows a set of techniques explained by the Spanish Captain Balbino Cortés. In the United States, the famous actor and martial artist Bruce Lee wrote Tao of Jeet Kune Do (1975) [17], explaining his fighting method. However, ancient books or manuscripts about martial arts or fighting techniques are uncommon due to the secrecy maintained by many schools or countries. This is something that has changed in the last decades, when many martial artists have written innumerable martial arts books.

Computer sciences highly influenced in the last years the way in which we learn a martial art. The apparition of streaming and videoconference services in the last decades have allowed long-distance teaching. This has demonstrated to be useful in situations of quarantine, like the situation caused in 2020 by the novel coronavirus SARS-CoV-2, when a lot of gyms, dojos and training centers had to close, and many teachers were forced to teach and train their students online. Trainers and students have encountered new problems derived from these services (some of which are identified in a questionnaire distributed in this Master's Thesis, see below) like the fact that the videos are still in 2D. Technologies for motion capture and modelling can be a good solution for storing and showing 3D movements digitally [18], [19].

In the same way as advances in science and computer sciences have influenced in the way we learn a martial art, advances in Artificial Intelligence (AI) could led to a new evolution of martial arts, enhancing the experience of students and teachers in training, specially to provide a personalized and adapted support based on user modelling [20] which could also address the limitations of current psychomotor teaching and learning online that were identified in the questionnaire. In fact, previous works already suggest that AI techniques can be used to support the modeling of human motion during psychomotor learning [21] as well as to provide multisensorial feedback (i.e. vibrotactile [22]) in complex motor skills such as those required when learning martial arts movements.

With the purpose of identifying new problems that could have arose from taking martial arts to online training, a questionnaire has been prepared containing different questions about positive and negative aspects of online teaching and learning of psychomotor activities. A total of 29 persons answered this questionnaire. Two of them were teachers, who among other issues, reported that when teaching their psychomotor activities online, they miss a direct interaction with their students and cannot properly observe how their students make the movements or ask for help. In turn, the outcomes from the 27 students suggest that the main problem of online learning resides in the apparition of technical problems related with the internet connection and the lack of space that some students have in their homes, making difficult the execution of some movements. The lack of personal interaction in psychomotor activities that requires interaction between two or more persons, and the lack of communication and feedback from a teacher are two problems that can be directly related with the online learning of a martial art. On the other hand, some advantages have been identified as well, like the possibility of training and learning from home, the possibility of taking the classes at different times and the use of videos. AI techniques could help to avoid some of those problems by providing personalized guidelines to the students that take into account the physical space available, simulating opponents, providing multisensory feedback which can include AI enriched videos of the user execution, fostering some of the advantages identified in the questionnaire like easing training at home. The details of the results can be seen in Appendix IV.

Further, the affective state of a student during psychomotor learning is important since it can affect the learning process [23]. Advances in technology could help to identify the student's affective state during learning and to give an affective response personalized according to the personality traits of the student. A situation like the quarantine caused by SARS-CoV-2 could affect the affective state of students and teachers, so in the questionnaire a set of questions regarding the affective state during online teaching/learning have also been included. Outcomes suggest that students tend to feel more bored, less motivated and less satisfied in online classes. Affective computing techniques supported by AI could help to provide an affective response to the users, by creating training plans, congratulating the user or managing the execution of errors in a way that the user does not feel frustrated or bored. Again, the results of this questionnaire can be seen in Appendix IV.

Thus, in this document, **different techniques of motion capture, modelling, analysis and feedback that could help assisting in the learning and teaching of a martial art,** are reviewed. In addition, a **prototype, powered by AI techniques, that can assist in the learning of a set of movements** is presented and evaluated**.**

1.2 Objective, Hypothesis and Methods

The main objective of this Master's Thesis is to ***research the possibilities of AI techniques to assist in the learning of martial arts movements.*** The research hypothesis that I will demonstrate is that it is possible to *recognize, model and analyze human motion, as well as give feedback to the user through the development of an AI-based application that executes in an Android device.*

An **Android device** has been chosen because it offers several advantages: 1) it is easy to develop and publish an app for Android, 2) Android devices are highly accessible by people all around the world, 3) Android devices have a set of useful sensors embedded into them that could facilitate the capture of the movements, such as accelerometer and gyroscope, and 4) there are a lot of libraries and documentation freely available online.

Following the state of the art in personalized psychomotor learning [24], this application has to be able to execute four different phases that cover the sensing and modeling of the movement as well as the design and delivery of the feedback. With this context in mind, and taking into account the review of the state of the art carried out, in this work we have defined the following phases, which are explained in more detail in Section 2.1: i) **capturing** the executed movements, ii) creating a **model** to represent the captured movements if it is necessary, iii) **analyzing** the characteristics of the motion and iv) giving indications or **feedback** to the user.

Due to the constraints of the Master's Thesis (30 ETCS = 750 hours) and taking into account that the most common Android device is a smartphone (and by default, smartphones already include inertial sensors to measure the movement of the device, such as those discussed in Section 2.2.2), we decided to focus the research and developments of this work on the **movements of one body limb, in particular the arm**, as it is possible to place this Android device (i.e., the smartphone) on the forearm as a wearable so that any movement performed is captured by the inertial sensors of the device, and at the same time, this placement allows the user to have direct view of the smartphone screen to access to visual information before and after performing the movements, as well as received auditory and vibrotactile feedback during the execution.

The movements selected are part of **the American Kenpo Karate's Blocking Set I**, which will be explained with more detail in Section 3.1.1, and consist in blocking defensive movements. In order to train the AI algorithms implemented in the system to model and analyze the set of movements, we have carried out a **user study following the ethical guidelines of UNED** (as well as the indications of the Master), which have allowed to collect data of the movements of 20 voluntaries with different levels of experience. Appendix III compiles the documentation that supported the user study, including the informed consent signed by the participants of this research and the questionnaire fulfilled with their demographic and personal information. The quarantine situation due to the SARS-CoV-2 prevented the collection of data from more users, as dojos closed in mid-March. The data collected will be explained in Section 3.2 and Section 3.3.

1.3 Possible Applications

An AI-based system with the characteristics described above could be useful in different ways. The main goal of the application would be assisting the learning of a martial art. Sometimes students have trouble when **training alone at home** the things they have learned in class and this system could incentive them to train through gamification techniques, notifications and reminders. The results and analytics of the training could even be sent to the teacher, allowing him/her to follow the progress of the students.

This AI-based system could also be useful in environments where there is only one teacher and a lot of students, so the teacher cannot be aware in detail of the movements of all students simultaneously and give them feedback. In this case, the system could be **a complement to the teacher** and help in his/her supervision of the students, as it can give feedback to each student independently but also point to the teacher those students that require his/her attention for additional human support.

The fact that the system is continuously recording the movement and giving feedback to the user could also help in the **prevention of injuries** through the detection of wrongly executed movements that could attempt against the biomechanics of the body.

In addition, **online learning** of a martial art could be possible (and more accurate) through the use of this AI-based system, since the system can also provide feedback to the teacher, and not only to the student, so both can better understand which movements have been executed and their characteristics, which were one of the main limitations identified in the aforementioned questionnaire. In fact, as mentioned before, this could be useful in situations like the current quarantine period during the outbreak of the novel coronavirus SARS-CoV-2, which is affecting all kinds of activities worldwide.

Of course, a system with these characteristics could not only be helpful when assisting in the learning of martial arts, but also assisting in the learning of any psychomotor activity such as dance, sports or playing instrument. So, a model able to be transferable between different martial arts, and to other psychomotor activities is preferred.

The research work carried out for this Master's Thesis has been divided in four phases, so they can be used on isolation for other specific purposes. For example, the motion capture phase could be applied in gesture recognition, the modelling phase in motion storage and cultural conservation, the techniques used in the analysis phase could be applied in other applications that involves the analysis, recognition or even generation of temporal sequences, and the interactive and feedback phase could be applied in interactive video games.

1.4 Contributions

The research carried out in this Master's Thesis has focused on studying the different AI based techniques for capturing, modeling and analyzing motion, as well as providing feedback. In addition, an application (which takes advantage of the inertial sensors of an Android device) has been developed for assisting the learning of a set of movements of a martial art. The captured movements have been labeled and segmented to create a dataset that is to be used with the AI algorithms. The data in this dataset have been then modeled and prepared using sequence modeling techniques, specifically Exponentially Weighted Moving Averages, for its later analysis using Neural Networks. Three different types of Neural Networks have been tested and compared (Fully-Connected Neural Network, 1-Dimensional Convolutional Neural Network, and LSTM Neural Network), and the best classifier obtained has been used in the development of the application. The application, called KSAS (Kenpo Set Assisting System), is able to guide a user through the execution of a set of movements, giving feedback when a movement has been executed wrongly or correctly.

In line with the open science philosophy, the source codes of the application for capturing data, the scripts used for modeling the data and training the networks, and the KSAS application are publicly available on GitHub[1][2][3].

Some of the ideas derived from this research work have been exposed in an event that took place in the installations of the UNED as part of the "Semana de la Ciencia y la Innovación 2019 / Week of Science and Innovation 2019", which is held annually in November in Madrid (Spain). In this last edition, the student carrying out this Master's Thesis and the director of it were the keynotes of the talk entitled "Ciencia y Datos en las Artes Marciales / Science and Data in Martial Arts"[4]. Part of the motion capture process was carried out during this event, in a participatory session that was organized after the two keynotes. Figure 2 shows the student of this Master's Thesis speaking in the event, and Figure 3 shows a participant recording his movements with the Master student supervision.

Figure 2. Student of this Master's Thesis speaking in the mentioned event.

[1] Motion Recorder repository: https://github.com/AlbertoCasasOrtiz/Motion-Recorded-Android
[2] Martial Arts Movements Classifier: https://github.com/AlbertoCasasOrtiz/Martial-Arts-Movements-Classifier
[3] KSAS repository: https://github.com/AlbertoCasasOrtiz/KSAS
[4] Event link: http://www.madrimasd.org/semanaciencia2019/actividad/ciencia-y-datos-en-las-artes-marciales

Figure 3. Student of this Master's thesis (right) supervising and guiding the capture of a movement of a voluntary (left) during the Week of Science and Innovation 2019 activity.

Part of the work carried out in this Master's Thesis, specifically the questionnaire regarding the emotional state during online classes of a psychomotor activity (Appendix IV), is used in the INT^{2}AFF project (INTelligent INTra-subject development approach to improve actions in AFFect-aware adaptive educational systems) funded by the Spanish Ministry of Science, Innovation and Universities under Grant PGC2018-102279-B-I00. More details about the applicability of this research to this project can be found in Section 6.3.

As a preliminary study to this Master's Thesis, a paper regarding the use of Labanotation for modeling human motion, focusing in martial arts was written in collaboration with Martha H. Eddy[5], an expert in Laban Movement Analysis and submitted to the e 27th ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2019, SCIE indexed). Reviews received liked the approach proposed but considered that it was immature for publication. We are now preparing a paper with the outcomes of this research to be submitted to a JCR indexed journal.

The knowledge necessary to develop this Master's Thesis has been obtained mainly from the subjects taken in the master. The subject Robotics gave me a better understanding of the different sensors used for capturing the motion (i.e., accelerometer, gyroscope and magnetometer). In the subject Computer Vision convolutional neural networks were studied. This kind of neural network has been applied in this research project as well. The subject Bioinspired Neural Methods gave a better understanding of the underlying theories that allowed neural networks to be invented. The subject Data Mining gave a better insight on how to organize the dataset used and how to apply and compare different classifiers, including neural networks. The knowledge obtained in the different subjects coursed in the Master has been complemented with the Deep Learning Specialization, imparted by Andrew NG in Coursera [6]. In this specialization, the use of different types of neural networks, the different ways of approaching a problem, the use of libraries such as Keras or Tensorflow and the ways in which a classifier can be optimized for obtaining better results is explained.

[5] Martha Eddy: http://www.movingoncenter.org/events/laban-with-martha-eddy/

[6] Specialization webpage: https://www.coursera.org/specializations/deep-learning

1.5 Structure of the Document

This document is organized as follows:

- Section 2 reviews the different approaches, techniques and systems that aboard the four phases mentioned before. Then, a set of researches focusing in martial arts that involves any of the four phases are reviewed.
- Section 3 explains the structure of the AI-system, called KSAS, that has been developed in this research, and the methods used for capturing, modelling and analyzing human motion, as well as the techniques used for giving feedback to the user. It also presents the results obtained when applying different AI methods (in particular, neural networks).
- In Section 4 the results of this research are discussed.
- Section 5 discloses with the conclusions drawn from this study and the development of the AI-based system, showing how the hypothesis posed is demonstrated and the research objective accomplished.
- Section 6, exposes some future lines of work that can be derived from this Master's thesis and how they could be carried out.
- Finally, in Section 7 the bibliography used in the elaboration of this document is compiled in order of appearance using the style of references called IEEE managed by the Mendeley[7] bibliographic system.

In addition, the document includes a glossary, a list of tables and figures, and some appendixes. Appendix I and Appendix II include two tables recompiling the information obtained after reviewing a set of publications using AI techniques for capturing, modelling or analyzing martial art movements, as well as giving feedback to the users. Appendix III explains some details about the process of motion capture, including how the captured information regarding the physical activity of the user has been organized, and the documents used such as the informed consent to be signed by the participants, and a questionnaire to collect information about their physical activity. Appendix IV contains the results of a questionnaire carried out with the purpose of obtaining information about the advantages and disadvantages that students and teachers have identified when teaching/learning psychomotor activities online, as well as their emotional state during online classes.

[7] Mendeley webpage: https://www.mendeley.com/

2 STATE OF THE ART

In the last decades, a great number of researchers have started to focus on the use of computer science for the elaboration of systems that are able to assist in the teaching of a martial art (Section 2.7). This ambitious task involves the combination of different techniques, disciplines, sensors, algorithms and devices that are reviewed in this section.

Although this research is mainly focused in martial arts, many of the papers, applications and systems that are reviewed are applicable in the development not only of a system for assisting the learning of a martial art, but may be also applied to other psychomotor activities like dance, recognition of gestures, or even to other fields such as robotics or videogames.

2.1 Phases of a Systems for Assisting the Learning of a Martial Art

Reviewing the literature, different studies focusing on some tasks that could correspond to the different phases that an AI-based system to support the learning of martial arts should achieve have been identified.

The way in which we could aboard these different tasks or phases may depend highly on the way in which we have defined the previous phase, so some methods in subsequent phases could be more suitable than others. The four identified phases (already anticipated in Section 1.2) are:

- **Motion Capture:** The first phase consists in finding and defining a way of capturing the movements that are going to be learnt.
- **Motion Modelling:** The second phase involves different ways of modelling and describing the captured data in a way that it could be easily processed, analyzed, visualized or stored, depending on the goal sought.
- **Motion Analysis/Processing:** The third phase implies different ways of analyzing, classifying, predicting or processing the recorded data in a way that more information can be obtained. The data used in this phase could be the raw captured data from the first phase or the modeled data from the second phase.
- **Give feedback or indications to the user:** The last phase depends of the information that have been extracted from the data and consists in giving that information to the user in a way that it is useful for him/her. For example, when executing a sequence of movements in the wrong sequence, the system must be able to indicate which movements have been executed in the wrong order. This would be very useful if the feedback is given in real time while the user is learning the sequence of movements.

The interaction between those four phases and the user can be seen in Figure 4. The cycle starts with some indications by the system to perform a movement and ends with some feedback after the performance by the user, or the request for performing another movement.

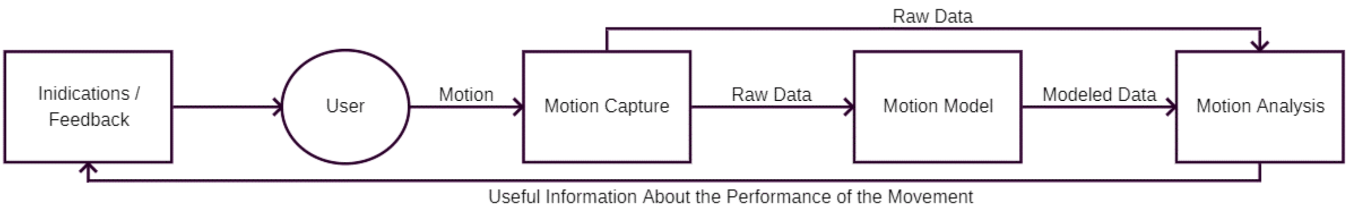

Figure 4. Flow of information between the identified phases for assisting the learning of a martial art and the user.

Since the development of psychomotor systems is still in an early stage, many of the studies we have found in the literature and reviewed focus exclusively in a subset of these phases. In the following sections, the focus is put in reviewing literature regarding human motion capture (Section 2.2), human motion modeling (Section 2.3), human motion analysis (Section 2.4) and ways to give feedback or indications to the user about the movement performed (Section 2.5). After that, an analysis of the characteristics of learning of a martial art is reported (Section 2.6), Next, some martial arts applications that include some of the techniques reviewed previously are presented (Section 2.7). Finally, there is a summary of the state of the art (Section 2.8) which can be complemented with the Tables 19 and 20 included in the Appendixes I and II

2.2 Capturing Human Motion

The first problem that must be faced when developing a psychomotor learning system is finding a way to record and sense the different characteristics of the motion such as velocity, acceleration or relative position of target parts of the body. There are attempts of creating devices to record motion in humans and animals even since the 19th century[8] [18]. The fields that have benefited the most of human motion capture are animation, cinema, videogames [18] and health sciences [25].

Reviewing the literature, I have noticed that existing classifications of human motion capture approaches do not include some of the methods I have found, either because some of the techniques did not exist when those classifications were created, because some of the methods are unsuitable for recording motion from the full body, or because some methods have been put apart for newer methods. So, I have provided my own classification that includes those missing. I have classified the main existing approaches for capturing human motion into the following:

- **Optical methods:** These methods involve the use of different types of cameras and optical sensors for capturing a sequence of 2D and 3D images of the body.
- **Inertial methods:** These methods consist in the use of inertial measurement units such as accelerometers or gyroscopes, which allow to obtain acceleration, rotation, and infer velocity and position.
- **Muscular methods:** These methods consist in the use of devices, such as electromyograms, able to directly measure muscular activity.
- **Other methods:** I have included here methods that have been put apart for newer optical and inertial methods or methods that are still in research and development.
    - **Magnetic methods:** Magnetic methods are based in the use of magnetic fields and sensors. Nowadays, magnetic methods have been put apart by cheaper methods such as inertial methods or newer optical methods.
    - **Mechanic methods:** Mechanic methods imply the use of an exoskeleton formed by a set of potentiometers. Due to some disadvantages regarding the restriction of movements when using mechanical mocap suits, these methods were quickly put apart by optical and magnetic methods.

[8] In 1879, the pioneer Eadweard Muybridge invented the zoopraxiscope. After capturing a sequence of images of a moving target, this device was able to project the sequence giving the sensation of motion. Muybridge used this technique in his studies of motion. Since then, there are more attempts to capture motion, like the chronophotographic fixed-plate camera of Etienne-Jules Marey able to capture sequential images of motion on a plate, or the famous rotoscope patented by Max Fleischer in 1915 and used in the study of motion for creating animations. The rotoscope was later replaced by more modern computer animation and motion capture techniques.

- **Sound-based methods:** These methods are based on the use of sound or ultrasonic transmitters and receptors and are still in research and development, but advances in newer optical methods and inertial methods are eclipsing the research of sound-based methods.

In the following sections, the mentioned methods and the ways in which motion can be captured using them are explained.

2.2.1 Optical Methods

Optical methods were the first motion capture methods that appeared and are broadly used in animation, cinema and videogames [18]. These methods are also the most commonly used in sport applications [26]. These systems can be formed by a set of cameras and a set of markers located in a motion capture suit (mocap suit), depth-cameras that are able to measure the depth of a scene, or regular cameras detecting human poses and shapes. The different approaches are described in the next subsections.

2.2.1.1 Vision with Set of Cameras and Markers

These sets are formed by a group of cameras, positioned around a place where the motion will be executed, and markers placed in different parts of the body [18], [19]. The markers are normally placed in a special mocap suit [18]. There are many commercial systems based on this method [27], but the most popular for sport applications are Vicon[9] and Qualysis[10] [26].

There are two kinds of systems depending on the characteristics of the markers [18]:

- **Systems with passive markers:** The camera emit light (normally infrared) using a LED and captures the light reflected by the markers.
- **Systems with active markers:** The marker emit light using a LED that is then captured by the cameras.

Once the information of all the markers is gathered by the set of cameras, it is combined to estimate the 3D position of each marker [18]. For estimating the position of a marker, a minimum of 2 cameras is needed [18]. To enhance the accuracy or measure more bodies (which implies that the cameras have to detect more markers), the number of cameras needs to be increased [18]. Real applications in sports normally use between 4 and 24 cameras using capture frequencies between 100Hz and 500Hz [26].

These systems have the advantage of being highly accurate and having a high capture rate [18], [26]–[28]. Motion of more than one person can be captured, the number of markers can be increases as needed, and the inferred data points can be used to generate a 3D skeleton that represents the body of each person [18].

Although these systems are highly accurate and broadly used, there are some limitations. The first limitation is their high cost if compared to other systems [18], [19]. It is also difficult and time consuming to create an adequate setup for the cameras and the users since it has to be decided where to locate the cameras and the markers [26]. Even sometimes special installations must be built or accommodated when using these systems [19]. Besides, the own body or other users can occlude the markers, and thus more post-processing in needed to infer the position [18], [26]. The lightning of the environment can produce noise in the captured data, and it is something that has to be controlled [18]. Another problem is that the motion has to be executed in a limited space, and post-processing techniques to infer rotation are needed, which limits the

[9] Vicon webpage: https://www.vicon.com/
[10] Qualisys webpage: https://www.qualisys.com/

ability of these systems to be used in real-time applications [18], [19], [26]. Real-time analysis is restricted to stick figures [18], and the post-processing for getting final results can be time consuming [19], [27].

Figure 5 shows an optical system for capturing human motion based in a set of cameras that emit infrared light. This light is then reflected by the markers in the mocap suit and captured by the cameras. It can be seen the cameras emitting infrared light positioned in the top of the image, an actress wearing a mocap suit in the middle, and a screen with a 3D representation of the body and the cameras in the left top sector of the imagen.

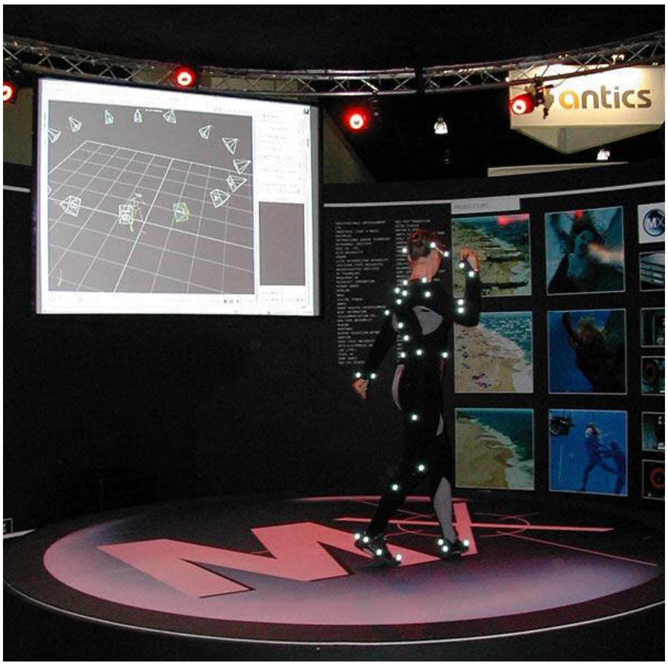

Figure 5. Set of cameras and Mocap suit. Image obtained from Wikimedia Commons[11].

2.2.1.2 Markerless Vision with RGB-D Cameras

RGB-D cameras or Depth cameras are cameras that captures not only color, but also depth. There are many techniques for extracting depth information using depth cameras for different applications [29]. Existing techniques can be divided into two categories [30]:

- **Passive 3D Vision:** These techniques only use the information of the environment for generating the depth data. The depth information can be inferred from the environment, using spatial information or even the focus of the camera. The problem of these techniques is that they do not perform well in environments with poor lightning or objects without textures, and more sophisticated software techniques are required.
- **Active 3D Vision:** These techniques use an additional source of light (normally lasers or light in the infrared spectrum) for generating patterns. Those patterns are emitted to the environment, and its reflection is captured by the sensors. Then, the differences in time or space of the pattern when arriving to the sensor are processed to obtain the depth data. These techniques solve the lighting problem of passive methods since they

[11] Image source: https://upload.wikimedia.org/wikipedia/commons/7/73/MotionCapture.jpg

generate their own references in the environment but require more resources and specialized hardware.

Regarding the extraction of depth information, the real-time methods used in the most known commercial applications are the following [29], [31]:

- **Structured light or pattern projection:** The camera projects a known 2D pattern, normally using infrared light. Then the reflection of the pattern in the environment is captured and the depth is computed through the spatial distortion of the pattern in the captured image [29]. The main advantage of this method is its simplicity and high resolution and accuracy, but specific applications may require specific processing [29]. This method is used in Kinect V1[12] [32] and Asus Xtion[13] [31].
- **Time of flight:** The camera emits an infrared wave continuously, modulated over time and captures its reflection. The depth is captured using the wavelength as reference, calculating the time that the wave takes to return to the sensor after reflecting on an object [29]. This method is used in Kinect V2[14] [31]–[33].
- **Passive Stereo Vision's triangulation:** There are two cameras which capture two imagens from different perspectives. Then, corresponding points between the images are found and triangulation techniques are used to estimate the depth of the points [29]. This method is more similar as how human vision works, but it could be difficult to find corresponding points due to lightning issues, texture issues or even perspective issues [29]. This method is used in Bumblebee XB3[15] [33].

In any case, the captured image consists in a set of density points, so a denser information is obtained and not only the position of some markers. The most known and used depth-camera device is Kinect, used in videogames industry [26], [34]. Using depth cameras has some advantages over the use of a set of cameras with markers on sport applications. The main one is that only one device is needed, and markers are no necessary anymore [26].

However, there are some problems regarding the low resolution of the depth images and the capture rate, which could limit the use to capture fast motion or focus on some body parts [26]. Furthermore, the body parts or objects can be occluded by other objects and the rear of the body cannot be captured. A solution for this is the use of multiple depth cameras, but that would imply to take out the advantage of using only one device [35]–[37].

With depth cameras, a 3D image formed by depth-pixels can be translated into 3D points (since the width and height of the image, and the calculated depth are known) and then it could be used to infer a 3D skeleton [34]–[36], [38], [39]. Figure 6 shows an example of a depth camera.

[12] Discontinued. Its successor is Kinect Azure DK: https://azure.microsoft.com/en-us/services/kinect-dk/
[13] Xtion webpage: https://www.asus.com/3D-Sensor/Xtion/
[14] Discontinued. Its successor is Kinect Azure KD: https://azure.microsoft.com/en-us/services/kinect-dk/
[15] Bumblebee XB3 webpage: https://www.flir.es/support/products/bumblebee-xb3-firewire/

Figure 6. Kinect Camera. Image obtained from Wikimedia Commons[16].

2.2.1.3 Markerless Vision with Regular RGB Cameras

Another approach when trying to capture human motion is to try to infer the pose or a 3D skeleton of the human body using regular RGB-cameras. This method has gained the interest of a great number of researchers since it solves many of the problems when using cameras with markers and depth cameras, like the high monetary cost [40]. Since when using regular RGB cameras only 2D information is obtained, more sophisticated algorithms, including AI techniques, are needed to infer the 3D information. These methods are reviewed in more detail in Section 2.4.

This kind of method can use one RGB camera or a set of RGB cameras. The information about the poses and the motion is obtained by using different techniques, like algorithmic methods that use some knowledge about the scene obtained a priori [41]–[47], or machine learning methods [42], [48]–[51], that use trained models such as Convolutional Neural Networks. The main disadvantage of algorithmic methods is that a priori knowledge needs to be obtained before applying it, so it may not be generalizable to scenes where that knowledge cannot be obtained. The main disadvantage of machine learning methods is that a dataset is needed for training the models and getting a good performance can be a time-consuming task.

The use of RGB cameras overcome the disadvantage of having to use specialized and expensive hardware (as those described in previous subsections), since RGB cameras are easy to get and cheaper than other optical methods [40]. Furthermore, RGB cameras can be really small and be placed anywhere [52]. Some RGB cameras, as those used in video surveillance, can also be remotely controlled by rotating it for changing the field of view and obtaining a 360º view.

A curious approach to this problem can be found in [51], where the authors propose to use a drone that is continuously orbiting around the user, getting a 3D skeleton of its movements. This allows the drone to take images of the user from different perspectives and allows the user to move freely since his/her movements do not have to be executed in a bounded space due to the drone following him/her. Drones can also be used in conjunction with Head-Mounted Displays (HMDs) to provide the user a mirrored vision of him/herself executing the movement in augmented reality environments [53].

A commercial method for markerless vision with regular RGB cameras that became well known due to its high sales was EyeToy for PlayStation 2[17] [40]. Nowadays it is even easier to use these methods since most of the people carry a smartphone, which have embedded RGB cameras [50].

---

[16] Image source: https://upload.wikimedia.org/wikipedia/commons/f/fe/KinectSensor.png

[17] Discontinued. Its successor was PlayStation Eye: https://en.wikipedia.org/wiki/PlayStation_Eye. It was also discontinued and the current successor is a depth camera: https://www.playstation.com/en-us/explore/accessories/vr-accessories/playstation-camera/

In addition, RGB cameras can also be used for recording motion that will be captured using other methods for data labelling or synchronization purposes [54], [55]. An example of an RGB camera for Arduino[18] boards can be seen in Figure 7.

Figure 7. Example of RGB Camera. CMOS OV7670 module [56].

2.2.2 Inertial Methods

Inertial methods are based in the use of Inertial Measure Units (IMUs). These methods have been broadly studied and used in robotics [57], [58], and it is beginning to break into the motion capture field with commercial options such as the wearables and mocap suits from Xsens [59]–[61]. These methods are less common in sport applications due to the fact that are relatively new and a still fresh research field [26].

The basic IMU has an accelerometer to measure acceleration, and a gyroscope to measure rotation and compensate errors derived from orientation [57]. Integrating information from acceleration and rotation over time, the position can be estimated [57], [58].

The advantage of these methods is that external references are not required, and can estimate acceleration, velocity, position and rotation directly from the motion [57]. These devices are low energy consuming and can be really small [62], [63], so they can be used to capture the motion of different parts of the body in the form of a wearable [60], [62], [64]–[66]. It is easy to create wireless wearables using Bluetooth modules and batteries [67], [68].

The combination of multiple sensors in a network allows the creation of aforementioned mocap suits that can capture motion of the full body [59], [61], [68]–[70]. This allows the estimation of 3D skeletons and poses, as optical methods. In this case, the inertial sensors take the position of the markers used in the optical methods. A great advantage against the use of optical methods is that the motion captured with inertial mocap suits does not have to be executed in a determined range of view, but in the range of the wireless signal receiver that is often broader. Also, inertial sensors do not have the problem of occlusion that optical methods have. The devices can also store the data for later processing and analysis.

The main inconvenience of inertial methods is that the integration of the data for estimating position can produce an accumulative drift error [57]. Kalman filters are the standard solution for estimating the correction of this error [58]. There are many researchers focusing in correcting

[18] Arduino webpage: https://www.arduino.cc/

the drift error with Kalman filters based methods [60], [71]. Nonetheless, even with these inconveniences, it can perform similarly as optical methods [59], [72].

Some IMUs can have a magnetometer, which allows the gyroscope to have a fixed point of reference (north), so it can estimate the orientation with better accuracy using a well-known point of reference. But electromagnetic noise could affect its performance [57], [73], [74].

These methods can be combined with other methods and sensors such as Global Navigation Satellite System (GNNS) [59], [69] to get positional information of the subject, or Optical Sensors [75]–[78] and ultrasonic sensors [79] to get better estimation of 3D skeletons.

An example of an IMU module for Arduino boards can be seen in Figure 8.

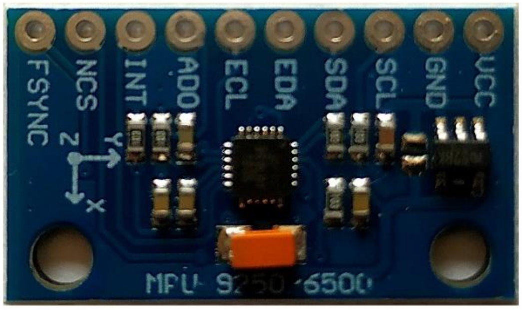

Figure 8. Example of IMU. MPU-9250 Module [80].

2.2.3 Muscular Methods

I have decided to name these methods muscular methods since they measure electrical muscular activity, using electromyograms and mechanomyograms [81]. In this way, the activation of a muscle can be detected, and thus, characteristics of a movement can be determined, such as muscular response time [82] or even the gesture that a subject is executing [81].

These methods are used for measuring muscular activation and response [82]–[84] and recognizing Laban Qualities (Section 2.3.2) [81], [85], [86] and gestures [81].

2.2.4 Other Methods

As commented above, other methods include those that are less used in sport applications [26], either because the methods have been put apart by newer methods, or because the methods are still under research. They are magnetic methods, mechanic methods and sound-based methods.

2.2.4.1 Magnetic Methods

Magnetic methods are based on the use of a magnetic transmitter, which generates an electromagnetic field in which the user will move. The user then has a set of electromagnetic sensors on his/her body. These sensors infer their own relative position with respect to the transmitter [18], [19], [87].

The main advantage of these methods is that the sensors can record position and rotation without post-processing, which is useful in real-time applications [18], [19], [27]. Besides, the sensors are not occluded by objects or other subjects and many subjects can perform at the same time [18], [19].

The main disadvantage of these methods is that metallic or electronic elements in the environment can produce interferences or noise in the measures [18], [19], and the accuracy decreases as the distance to the transmitter is increased [27]. Furthermore, these systems capture at lower sampler rate than other systems [18]. In addition, the area in which the subject can perform is limited by the electromagnetic field, and the equipment can limit the user's movements [18], [19]. The advances in wireless technology are a solution for this last problem [88].

These methods are less common in sport applications [26]. This could be because cheaper technologies like depth cameras or inertial sensors have arose. Another factor could be that many sport applications require the user to move freely in a broad space, but using magnetic methods, the space in which the subject can move is limited by the electromagnetic field [19].

2.2.4.2 Mechanic Methods

Mechanic methods are based in the use of exoskeletons that measure directly the angles of the body joints using potentiometers [18]. These systems are suitable for real-time applications since they can infer the poses of the body directly from the potentiometers. The global translation of the user in the environment is difficult to estimate since only the angles of a set of joints are measured [18]. For example, if a user is walking 10 meters to the right, it is easy to estimate the angles of the joints of his/her body, but it is difficult to estimate how much distance has him/her walked only from those angles. There would not be many differences between a user that is actually walking and a user pretending to walk while staying in the same place.

The main advantage of these systems consist in the fact that they do not rely on external sensors or information sources, so occlusion or interferences cannot occur [18]. In turn, their main disadvantage is that it can restrict the subject movements [18], [26]. In addition, these systems can break easily and captures data at a lower rate [18]. These disadvantages makes mechanical methods unpractical for sport applications [26], so other alternatives should be used.

2.2.4.3 Sound-based Methods

Sound-based methods are methods that use different frequencies of sound waves and elements for estimating human motion, joint angles, poses or gesture. These methods are broadly used in robotics for robot positioning [58]. I have classified these methods into two:

- **Using transmitters and receivers** [89]–[95]**:** These methods use a set of ultrasonic transmitters located as wearables in the body, and receivers that are able to estimate the position of the transmitters, normally by triangulation. These methods are used in health sciences as a cheaper and easier alternative to optical methods.
- **Using echo** [96]–[99]**:** These methods emit a sound wave and receive the reflection of the wave in different surfaces. Using time of flight techniques, similar to techniques used in time-of-flight depth cameras, the depth can be estimated. These methods have the advantage that wearables or markers are not needed anymore, but the detection of shapes and poses is limited. Echo based methods works similar as echo-localization works in bats and dolphins.

These methods are no commonly used sport applications [26], but are a good and cheap alternative to other methods in gesture recognition [89], [97]–[99].

2.3 Modelling Human Motion

After capturing the data, it is usually necessary to model it. A good modelling could provide a better and standardized representation of the data, easing their reuse, analysis, visualization and storage [100]. A good model could even allow to generate and synthetize new motion [101], [102].

As it has been shown before, when capturing human motion, the focus can be put in different parts: 1) the full body (e.g. using mocap suits), 2) portions of the body (e.g. using depth cameras), 3) characteristics of the motion (e.g. using accelerometers) or 4) electrical activity of the body (e.g. using electromyograms). Each of those options needs a different way of modelling. As explained next, this makes a difference in the processing.

On the one hand, when capturing motion of the full body or a part of the body, the information captured corresponds to how the joints of the body are moving with respect to each other. This information is captured frame by frame and can be represented as a 2D silhouette, a 3D skeleton or a 3D representation of the body part [18].

On the other hand, when capturing characteristics of the motion or electrical activity of the body, the corresponding signal is transformed into a set of data values that are obtained by sampling at a determined rate. This generates a temporal series of measures that can be modelled using mathematical methods or machine learning methods [103].

2.3.1 Levels of Abstraction for Modeling Human Motion

When explaining human motion modelling, I am going to follow the ideas from [104], where an action is formed by a sequence of poses, and a sequence of poses is formed by poses. The idea from [104] comes from making an analogy with NLP, where the syntax of a phrase is formed by a sequence of morphemes and morphemes are formed by sequences of phonemes. In the same way, actions are formed by sequences of silhouettes, skeletons or signals, and those sequences are formed by sequences of silhouettes, skeletons or signals at a given time. I have added some information about time series to those ideas and created Figure 9, where I have identified four layers or levels of abstraction when modelling human motion:

- The **first level** corresponds to the meaningless raw captured data, which can be either the body representations of the joints, or the time series of the signals collected such as acceleration, rotation, position of the pixels in an image, etc.
- The **second level** represents data captured in a given instant of time. If the captured data in an instant of time is expressed as a set of configurations of the joints of the human body, a configuration of the human body is obtained, otherwise it corresponds to captured signals at a given time. A set of signals captured at a given time can also be combined or processed to obtain new characteristics of the motion. For instance, if the gravity in the x axis is subtracted to the acceleration in the x axis (both measured in the same instant of time), a new feature (which is the lineal acceleration in the x axis) can be obtained.
- The **third level** represents the motion, that can be expressed as a succession of poses and data over time. If a succession of human body configurations is created, a representation of the motion of the human body is obtained. If a succession of signals or electrical activity of the body at a given time is created, a time series of how those signals change over time is obtained. In this case, we have a set of values measured in the time window. Thus, a combination of signals (or values of the same signal) in a time interval can produce new characteristics of motion, which can be computed either in the time or frequency domains.

- The **fourth level** represents actions. Actions are meaningful pieces of motion or characteristics of motion like walk, run or play tennis. Giving a meaning to each piece of motion and obtaining information from it is the main purpose of many researchers. An action can then be defined by a meaningful succession of human body configurations, a meaningful succession of signals, or a meaningful succession of a combination of human body configurations and signals.

The main difference between the second and the third level is that in the second level the data captured in an instant of time is modelled (For example, a 2D silhouette is extracted from an image), and in the third level, the sequence of data over time is modelled (For example, a sequence of 2D silhouettes, which conforms a video, and thus, motion). The data captured in an instant of time can be modeled in isolation (A 2D silhouette can be obtained from an image), but some task require sequences of data to be modelled (The generation of temperature images indicating the amount of movement that takes place in a scene over time requires the full sequence of images).

In the following sections, I am going to review how to model the second, third and fourth levels of abstraction for the captured raw data.

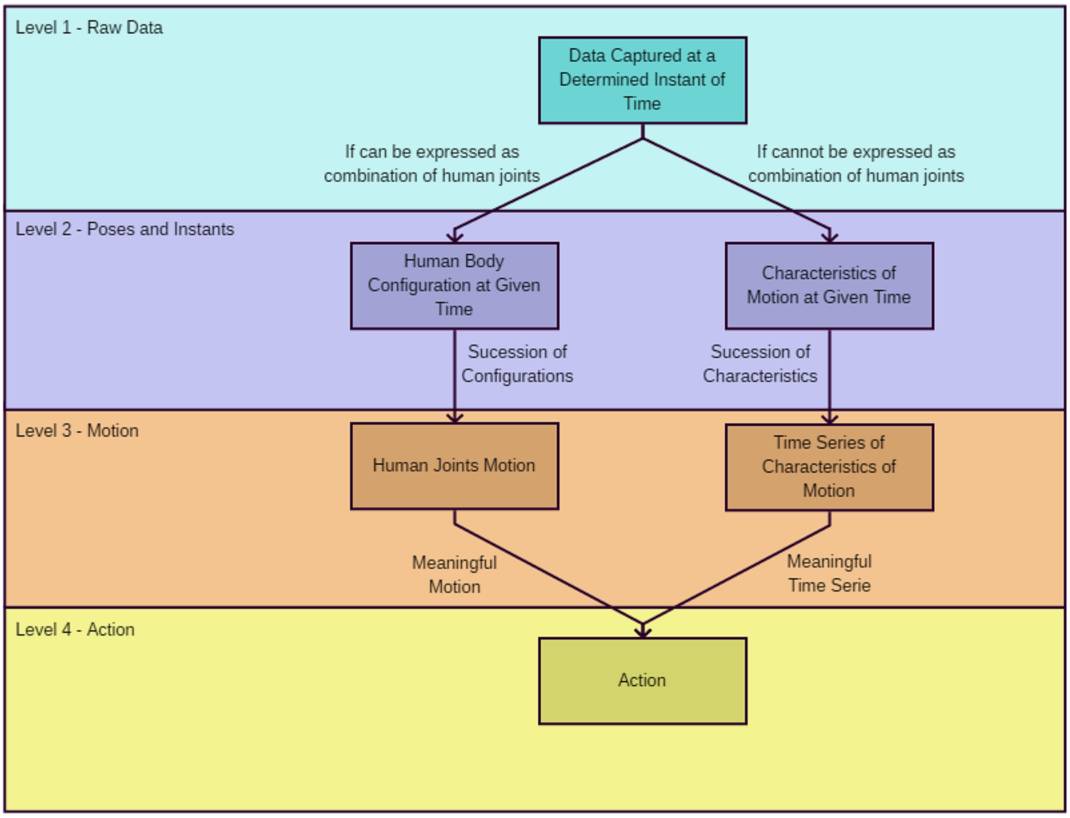

Figure 9. Levels of abstraction when modelling human motion.

2.3.1.1 Level 1 – Raw Data

Level 1 corresponds with the meaningless raw captured data using the aforementioned motion captured methods at a determined sample rate. The data at this level has no meaning, and it is difficult to understand the data at sight. It needs to be modeled, so the underlying information can be accessed with ease.

2.3.1.2 Level 2 – Poses and Characteristics of Motion at a Given Time

If the data of the level 1 at a determined instant of time can be expressed as a combination of joints of the human body, a configuration of the human body is obtained. This configuration is what is called a pose and can be represented as a 2D/3D silhouette or as a 3D skeleton.

Otherwise, if the data obtained in level 1 at a determined instant of time cannot be expressed as a combination of joints of the human body, a set of characteristics of the motion is obtained. These characteristics can vary highly depending of the goal of the research, and can be formed by acceleration, velocity, electric activity, heart rate or temperature amongst others.

When modelling Level 2, we are giving a meaning to each one of the values obtained at a given time instant. For example, one of the values captured at a given time in Level 1 could correspond to the speed of the motion, another could correspond to the angle of the elbow joint of the body, and so on.

2.3.1.2.1 Modelling Poses

A pose is a configuration of the joints of the human body in an instant of time. There are different ways of modeling poses, mainly depending of the purpose of the application or the system used for capturing data. As introduced above, they are 2D/3D silhouettes and 3D Skeletons.

2.3.1.2.1.1 2D Silhouettes

A 2D silhouette is the graphical representation of the body in 2D, thus, is the 2D projection of a 3D body [104]. This representation can be obtained from a 2D image captured using a RGB camera, by subtracting the background [43] or directly recognizing the pose [42], [44]. The main problem of this is that normally the 2D silhouettes are obtained from 2D images, so spatiotemporal information is lost. This can be solved by capturing silhouettes from multiple points of view, for example, using multiple cameras. In this case, the pose is formed by multiple silhouettes [104].

Different parts of the body can be determined in the silhouette using different techniques like image segmentation [104]. A 2D stick figure can be obtained for representing the body joints, obtaining a 3D skeleton from the 2D pose, which could solve the problem of the lost spatiotemporal information. This is a complex task that is still a research interest [47], [48], [105].

An example of a 2D silhouette obtained by background subtraction can be seen in Figure 10.

Figure 10. Example of a 2D silhouette. Image obtained from [43].

2.3.1.2.1.2 3D Silhouettes

A 3D silhouette is the graphical representation of the body in 3D. A 3D silhouette is different from a 3D skeleton (see next) because 3D silhouettes do not have joints defined. A 3D Skeleton can be obtained from a 3D silhouette and a 3D silhouette can be obtained from a 3D skeleton if a model of the body has been obtained a priori, for example, using a 3D scanner.

This representation can be obtained either by the use of the depth pixels obtained by a depth camera [106], and or by the use of a set of RGB cameras capturing the body from different points of view [107]. In both cases the obtained model consists in a cloud of points in a 3D space representing the silhouette.

An example of a 3D silhouette obtained by using multiple depth cameras can be seen in Figure 11.

Figure 11. Example of 3D silhouette. Image obtained from [35].

2.3.1.2.1.3 3D Skeleton

A 3D skeleton consists in a set of joints, representing joints of human body and links between them. It is thus, a 3D digital representation of the different joints of the human body [18], [19], [108]. An example of a 3D skeleton can be seen in Figure 12. The idea comes directly from anatomy and biomechanics [109]. The joints of the human body that are normally captured and modelled are synovial joints, of which are six types, as shown in Figure 13.

It is the most common technique used when capturing the full body using mocap suits or depth cameras, and the main advantage is that the pose can be seen from any perspective using 3D modelling software such as Blender [110].

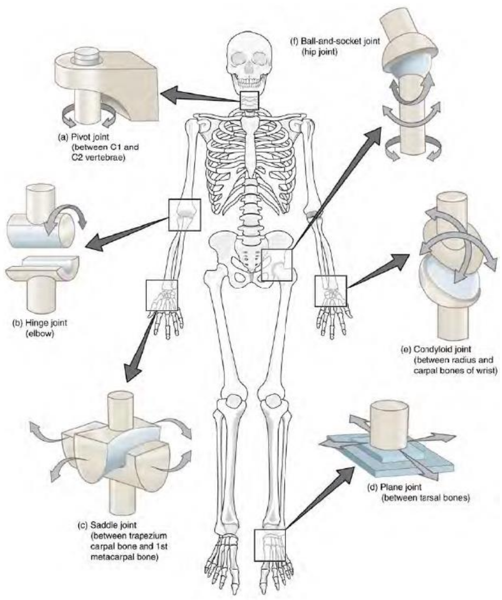

Figure 12. Types of synovial joints of human body. Image obtained from [109].

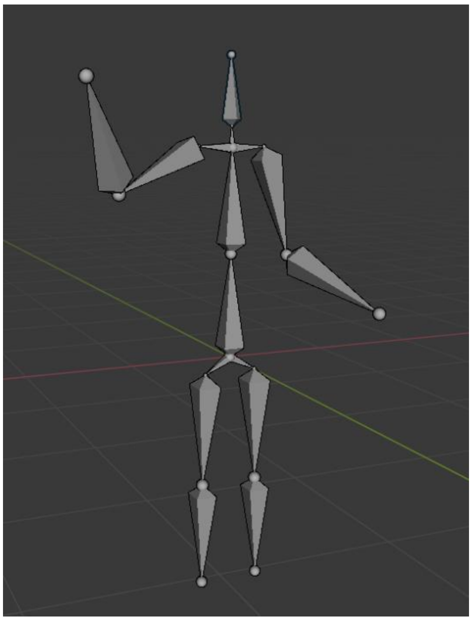

Figure 13. 3D Skeleton formed by 11 joints generated with blender [110].

2.3.1.2.2 Modelling Characteristics of Motion

When modelling characteristics of motion, it is needed to determine which characteristic correspond with which captured data. It can be easily done in some motion capture techniques such as inertial sensors, where we already know that the captured information corresponds with acceleration, but it can be more complicated when using other kind of capture techniques such as sets of cameras with markers, where we have to determine which marker corresponds with each part of the body.

2.3.1.3 Level 3 – Motion and Time Series

Once we have defined and modelled poses and characteristics of motion at a given time, the succession over time of poses results in a representation of human motion over time, and the succession over time of characteristics of the motion results in the temporal evolution of those characteristics.

It is important to note that in both cases, time series are obtained, so human motion and the temporal evolution of its characteristics can be processed using the same techniques. Modelling a time series implies to create a model that can represent the trends or values in the series, cleaning noise and smoothing the data. The ways in which a time series can be modelled have been broadly studied in socio-economic fields with the purpose of visualizing trends or forecast and predict values [103].

In the following sections, some mathematical models for modelling time series are reviewed, as well as machine learning techniques that could be useful for this task.

2.3.1.3.1 Mathematical Models

The main goal of modelling time series is to extract the underlying information and structure of the series [103]. There are several methods for modelling time series:

- **Regressive models** [111]**:** The output values of the model are a combination, linear or non-linear, of input values. In these models, they normally need to optimize some coefficients to fit the model to the data.

- **Autoregressive models** [103]: The output values of the model are a combination, linear or non-linear, of values obtained from previous time steps in the series. As in regressive models, they normally need to optimize some coefficients to fit the model to the data.
- **Moving Averages models** [103]**:** The output values of the model are averages of values obtained from previous time steps in the series.
- **Nonlinear Moving Averages models** [103]**:** The output values of the model are averages of the result of applying a non-linear function to the values obtained from previous time steps in the series.

As it can be seen, there are several methods for mathematically modelling time series, which can also be combined for creating new methods such as ARMA (Autoregressive Moving Averages) [95] or ARIMA (Autoregressive Integrated Moving Averages) [95].

2.3.1.3.2 Machine Learning Models

This kind of models use different algorithms and techniques that are able to learn the underlying characteristics and organization of a dataset [112]. The model then, is learned directly from the input data and with the purpose of solving a certain task, with usually is to classify the input data into a set of classes. The learning of the model is done by tuning an adjusting a set of values or weights, that define the learned characteristics. There are many types of machine learning classifiers, but in the following section, some of the most representatives according to [112] are reviewed.

2.3.1.3.2.1 Decision Trees

Decision trees represent functions that, given an input, returns an output value corresponding with the class at which the input belongs [112]. The information is organized as a tree, with a set of nodes that processes each input and returns a value to the next node. Each node can either have a set of branches that connect the node with a deeper node in the branch, or be a final node, that returns the final class assigned to the input. The simplest type of decision tree is the Boolean tree, in which each node can redirect the data to one node if it meets a condition, or to another node if not. A decision tree that represents the structure of a dataset and that can be used for inferring the class of a given example can be built by means of a recursive algorithm following the strategy known as divide-and-conquer. The advantage of using decision trees is that the learned use to be simple and can be understood by a human. However, if more data or classes are added to the dataset, a new full tree must be created.

2.3.1.3.2.2 Support Vector Machines

Support Vector Machines (SVM) are able to create an optimal "separator" or discriminator between the different classes that are classified [112]. It can linearly separate the examples of the dataset in different classes by embedding them in different hyperplanes by using a kernel, which is useful for datasets in which the classes are not linearly separable. SVM learns by adjusting a set of weights that can be part of a linear regression model (Section 2.3.1.3.1). The advantage of using support vector machines is that the discriminator between two classes is optimal, but if more than two classes need to be classified, an SVM must be trained for classifying each class against the others. This technique is known as one-vs-all.

2.3.1.3.2.3 Neural Networks

Neural Networks (ANN) are formed by a set of units, called neurons, that are organized in interconnected layers. Each unit of a neural network performs a mathematical function over a set of weights contained in the unit, and the input data from previous layers (or the same layer in recurrent units) There are many types of neural network depending on how the units are interconnected and on the function executed by the unit.

The weights contained in the units, after being adjusted, store the learned information about the dataset. Those weights are adjusted by iterating over the dataset using optimizer functions [112], [113] such as Back Propagation [114] or Adam Optimization [115]. Open source libraries such as Tensorflow[19] or Keras[20] facilitate the development of models based on neural networks and their use in different platforms such as web applications or Android applications.

The use of Neural Networks entails a set of advantages that make them suitable for solving problems for different fields. The main advantage of neural networks is its plasticity when solving problems that entails includes different types of input data (text, images, time sequences…), since different types of neural networks have been created with that purpose [113]. For example, Convolutional Neural Networks were created inspired in how human vision works, and thus, it has been successfully applied in computer vision problems, and Recursive Neural Networks have been created for processing time sequences, and they have demonstrated their potential in natural language processing problems [113].

An advantage of neural networks that has been gaining popularity in the last years is transfer learning [113]. Transfer learning consist in take an already trained model that solves one problem and adapt it to other problems by learning some iterations over a new dataset. This allows to share the generalizable underlying information learned by a model and apply it to another problems. Then, the neural network just needs to learn specific information about the new dataset. This not only reduces the time spent in training the network, but also the time spent in tuning the hyperparameters of the network and creating an optimal network architecture, since somebody has already done it.

The main disadvantage of neural networks consists in the fact that the learned model can be highly complex, and difficult to understand by a human being [113]. So, once trained, a neural network must be seen as a "black box" that performs the requested task. For validating that the learnt information and model can be generalized to external data (i.e., data that is not present in the training set), test and validation sets can be used [112], [113].

Since in this kind of system, analysis and modelling goes hand to hand, the specific models that can be used for analyzing sequences (Fully Connected Artificial Neural Networks, 1-Dimensional Convolutional Neural Networks, and Long Short-Term Memory Neural Networks) are described in Section 2.4.

2.3.1.4 Level 4 – Actions

The most difficult task when modelling human motion is to model actions. An action can be defined as a portion of motion, with a meaning or a purpose [104]. The easiest way to model an action is to label a succession of poses, or signals, over time. More sophisticated ways of modelling an action establishing relations between actions and inferring more information is still under research. Some of the most common approaches are mentioned bellow, either for recognizing or classifying actions, or for synthetizing new motion:

- **Modelling actions as grammars**: As it can be seen in [104], an action can be modelled through the creation of grammars. This method involves the extraction of the poses from the captured motion and then, delete the duplicated poses. This allows to have a set of phonemes that will conform the grammar. Then, the grammar is created by defining a set of rules. This grammar allows to detect and identify actions. In [104], this method is applied to videos and extracted 2D silhouettes.

[19] Tensorflow webpage: https://www.tensorflow.org/
[20] Keras webpage: https://keras.io/

- **Modelling actions as graphs**: Using a set of motion data as a basis, the idea of [101] consists on the use of directed graphs to encode the relations between different motion sets. In motion graphs, the edges are motion sequences and the nodes are common poses between two or more motion sequences. Once the graph is obtained, it is possible to walk through it to search for different sets of movements or actions. These graphs were created with the purpose of generating motion sequences using 3D skeletons.
- **Modelling actions as Hidden Markov Models**: The use of Hidden Markov Models (HMMs) for modeling human activity in 2D videos has been proposed many times [41], [116], [117]. HMMs allow to model temporal sequences of frames that represent actions. These models can be used later for classifying or recognizing actions. We will talk more about HMMs in Section 2.4.2.3.
- **Modelling actions with CNN**: Different kinds of Convolutional Neural Networks (CNNs) have been proposed for human action modelling and recognition [48], [118]–[120]. As said before when talking about time series, even with the neural network internally modelling the motion, it is difficult to know which features has the neural network learned. The main purpose of using these methods is not to model the actions, but to recognize, analyze, or classify it. We will talk more about CNNs in Section 2.4.2.2.
- **Modelling actions with RNN**: Recursive Neural Networks (RNN) are a kind of neural network broadly used for learning sequence models, especially in natural language processing [113]. These ideas can be applied for learning motion sequences shown in [102], [118], learning the probabilities of a pose appearing after another pose, and thus, learning the probabilities of a sequence of poses to appear. The method described in [102] was created with the purpose of generating motion sequences using 3D skeletons. We will talk more about RNNs in Section 2.4.
- **Modelling actions as templates** [121]: Using a set of recorded actions, a template that will be formed by the characteristics of those actions can be generated. Then, new performed actions can be identified by comparing them with the stored templates. The use of templates can be combined with the use of HMMs, obtaining templates defined as an HMM. The comparison can be done by using Dynamic Time Wrapping algorithm (Section 2.4.2.1) [122] , or the Viterbi algorithm if the template is defined as an HMM (Section 2.4.2.3).
- **Labelling sequences as actions:** This is the most straightforward approach and consists on directly select a portion of a sequence (either formed by signals or joints) and label it, indicating with which action the sequence corresponds. For making it possible, it is needed some correspondence between the action and the capture data. An easy solution for this would be to record the motion capture process using an RGB camera and then synchronize the captured data and the video, so the actions can be easily labeled by observing them in the video.

As it can be seen, there are several proposed ways for modeling actions. This is an active field of research in human motion modeling and there is no consensus yet in a standardized way of modeling. The fact that some techniques for recognizing or classifying actions have an implicit learned model of the motion, may have led to many researches to focus on the use of those techniques, even when we do not really know the characteristics learned by the models.

2.3.2 Notation Systems: Labanotation

In the past, as introduced in Section 1, texts and drawings were used when trying to portray motion on paper. This had some problems like the lack of dimensionality and motion. Movement notation systems arose as a way of representing dimensionality, using symbols and motion through the representation of those symbols over time. Those systems have been very popular in dance, and many systems such as Benesh notation or Morris notation were created with the purpose of notating dance [123].

The system that we are going to review here is known as Kinetography Laban or Labanotation [123], published by the famous dancer and movement theorist Rudolf Laban in 1928. This method allows to record motion in a precise and detailed way that have demonstrated to be useful for notating not only dance, but any kind of motion [123]. Its use for modeling psychomotor activities and particularly martial arts, has already been proposed [124].

The idea of Labanotation is very similar as how we notate music [123], and consists in the use of a vertical staff and a set of symbols. The staff represents the flow time and is divided in various columns, each one for a part of the body. The symbols represent the position towards which the part of the body moves, and its elongation over the will indicates the duration of the movement. Figure 14 shows an example of a labanotation score.

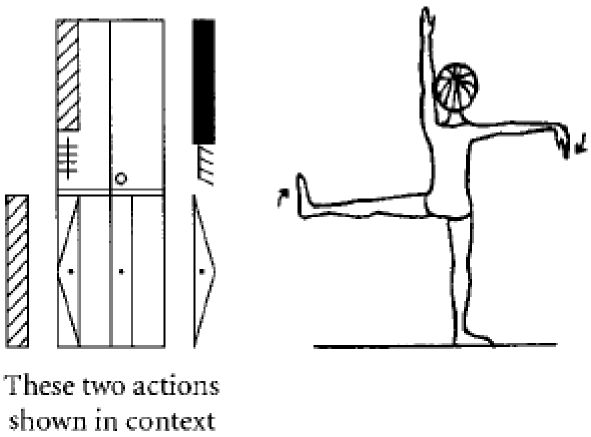

Figure 14. Example of Labanotation score and its execution. Obtained from [123].

There are some commercial applications available, in which Labanotation scores can be played as animations [125]. This notation system can be easily represented in a computer through the use of XML tags [125], and there are even attempts to translate motion captured through motion capture techniques into Labanotation [126], [127]. It can be applied to any field that studies human motion such as Robotics [128].

2.4 Analyzing Human Motion

For obtaining useful information and categorize the captured and modeled motion, we need to analyze it. There are several tasks that can be done when analyzing human motion such as extracting new characteristics of the motion from the signals, recognizing movements or gestures, or even synthetize new sequences of motion. In this section, the term characteristics refer not only to the extracted characteristics from the signals, but to the signals as well.

In the following sections, different techniques for getting new characteristics from the data, classifying and recognizing actions and gestures, or synthetizing new motion sequences are reviewed.

2.4.1 Getting Characteristics from Motion

One of the most important tasks when analyzing motion is to extract new characteristics that could help to increase the accuracy of the developed algorithms. These characteristics can be obtained by analyzing the captured data and combining it. In sports and martial arts, these researches normally have the purpose of comparing performance between different disciplines, and/or users [82], [129], [130], or studying the effect of different training techniques [84]. The new obtained characteristics can be part of the motion model and can be used as input for other algorithms or analysis techniques.

The methods used for estimating this kind of characteristics are normally statistical or mathematical methods, applied over the domains of frequency and time of the obtained time series [131]. A detailed review of the different methods used is not done here, since the equations, functions or algorithm used may vary highly depending of the motion capture method used, further, this Master's Thesis is focused in exploring the use of AI techniques for analysis, especially neural networks. For example, the speed of a kick can be obtained in different ways, either by calculating distances between different joints of a 3D skeleton at a determined sample rate, by calculating the distance of pixels in 2D poses at a determined sample rate, or by integrating acceleration obtained by an inertial sensor at a determined sample rate.

A new set of characteristics of motion can be obtained by using AI techniques, like Neural Networks [119]. The main advantage of learning a set of characteristics using Neural Networks is that the characteristics obtained have been inferred and selected by the network from the input characteristics specifically for solving a problem, or for being used as an input of other kinds of algorithm [113], [119], which could enhance the performance of the algorithms and the results obtained when solving the problem. As mentioned before, the problem of using this method is that the learned set of characteristics is usually difficult to understand by a human. Once the network has learned new characteristics of motion, transfer learning can be applied for using the learned underlying information of the model when learning characteristics of another dataset or domain [113].

2.4.2 Recognizing Actions and Gestures

One of the most important and difficult tasks when analyzing human motion is to recognize actions and gestures. Since this is still a fresh research field, many techniques have appeared in the last decades. In the following sections we are going to review some of those techniques.

2.4.2.1 Dynamic Time Warping

Sequence alignment is a well-known task in fields such as bioinformatics or speech recognition. The ideas and algorithm of those fields can be adapted to the task of recognizing actions and gestures since our captured data consist in a set of time sequences [108].

A well-known algorithm for aligning a measure of the similarity between two time series is Dynamic Time Warping (DTW). This method is based in the ideas of dynamic programming and can measure the similarity between two data sequences that may vary in speed [108]. DTW has been broadly used in speech recognition and audio processing.

As seen in [38], if we can obtain a model of the actions or gestures that we can recognize, we can use DTW for comparing an executed movement with the stored models, so we can get the best match as an estimation of the executed movement.

2.4.2.2 Artificial Neural Networks

This section complements the information about extracting new characteristics of motion (Section 2.3.1.3.2) and modeling actions (Section 2.3.1.4) using Neural Networks that was showed in previous sections by explaining the most common types of neural networks. Since in this research Neural Networks are going to be used for analysis because of the advantages previously mentioned, this section is focusing solely on them. Artificial Neural Networks have been broadly used in the last decades due to their plasticity when learning characteristics from any kind of dataset [113]. These networks consist in a set of layers of neurons and links between them. The neurons can perform different mathematical operations like logistic regression or convolution [113].

The most common kind of Neural Networks used in human motion analysis are Fully Connected Artificial Neural Networks (FC-ANN), Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). This is due to the fact that these two kind of neural networks perform very well when recognizing patterns and extracting characteristics from images and time series [113]. We are going to review those three approaches:

- **Fully connected Artificial Neural Networks (FC-ANNs)** [113]: FC-ANNs, also known as multilayer perceptron, are the most commonly used kind of neural network. The neuron of a regular FC-ANN usually performs a dot product between an input vector and a vector of weights of the neuron. Then, a bias value is added to the result of the dot product, and an activation function is applied to it.
- **Convolutional Neural Networks (CNNs)** [113]**:** CNNs are broadly used in computer vision and speech recognition. The neuron of a CNN consists in a matrix of weights that is convoluted over different sections of the input data. The fact that the convolution is applied using a matrix allows to capture characteristics and information shared by nearby values of the input. CNNs are useful for extracting features from motion [118], for directly recognize human actions [119], or even for inferring a 3D skeleton from a 2D pose [48].
- **Recursive Neural Networks (RNNs)** [113]**:** RNNs are broadly used in natural language processing since they were designed for sequence analysis. This kind of neural networks have the particularity of storing memory, which allows to learn characteristics from a portion of a sequence that can be useful when processing later portions of the same sequence. RNNs apply the same set of weights recursively over the inputs, and can be used for recognizing actions and gestures [118], [132], and even for generating new motion (Section 2.4.3) [102].

As mentioned before, the weights of a neural network are learned by means of an optimizer algorithm such Adam [115] or Backpropagation [114], and transfer learning can be used for applying the underlying characteristics of one problem into another [113].

2.4.2.3 Hidden Markov Models

Hidden Markov Models (HMMs) were introduced in Section 2.3.1.4 as a method for modelling actions. This section complements that information, and here we present how HMMs can be used for tasks such as classifying or recognizing movements.

HMMs are state machines in which the real states of the system are not directly observable, we can only observe a set of outputs or emissions generated by those states. In HMMs the transitions between states are probabilities. HMMs assume the Markov property, i.e., the next state only depends on the current state, and no on the past states [133]. HMMs have demonstrated to be useful in different fields such as speech recognition, bioinformatics or finances [133].

Given a set of observations, we can lean the underlying states and the transition probabilities using the Baum-Welch algorithm [133]. If we already have the learned the model, or if it has been given a priori, we can use the Viterbi algorithm [133] for extracting the most probable sequence of states given a set of observations. DTW can also be used for comparing models created by HMMs.

In human motion analysis, we can train a HMM for each action we want to recognize [108], [116], [117]. Once we have a model for each action, it is easy to determine which action has been executed by a new subject by executing the Viterbi algorithm over the sequence of captured poses for each trained HMM and selecting the one that gives the highest probabilities [41].

2.4.3 Generating Motion

Obtaining realistic human motion is an important task in videogames and animation [18]. In previous sections we reviewed the disadvantages of using motion capture techniques, like the fact that some of those techniques are expensive, and the setup of the devices and environments can be time consuming. Furthermore, modelling tasks for inferring human skeletons can be sometimes tedious. Automatic synthesis of human realistic motion could be a great solution to this. This can be achieved by the use of motion graphs (Section 2.3.4) [101] or Recursive Neural Networks (Section 2.4.2.2) [102]. Generated motion could also be useful for generating new examples for a dataset when more data is needed and there is no an easier way to obtain more examples [113].

2.5 Giving Feedback to the User

When the word feedback is used in this document, it is referring to extrinsic feedback, i.e., the feedback given by an external source, like a screen, a speaker or the teacher, and uses senses of the body such as sight or touch. The other kind of feedback is the internal feedback, given by the own body [134], which uses other senses of the body like balance [109].

An AI-based system able to assist in the learning of a psychomotor activity must be able to give feedback or indications to the user to help him/her to improve the executed movements. These indications must be provided using human senses. The most known senses that humans have, and allows humans to behave in their environment, are taste, smell, touch, sight and hearing [109]. Since taste and smell are senses that requires chemical stimulus and are more related with nutrition than with spatial orientation [109], the focus is put into sight (visual feedback), touch (haptic feedback), hearing (auditory feedback) and combinations of them (multimodal feedback) [134]. The feedback can be provided to the user either during the execution of the motor task (concurrent feedback) or after it (terminal feedback) [134].

When learning a psychomotor activity with a teacher, students are continuously receiving visual, haptic and auditory feedback from the teacher. For example, when the teacher executes an example of the movement that the student has to learn, he/she is providing visual feedback; when the teacher is explaining verbally to the student how to execute those movements, he/she is providing auditory feedback; and when the teacher is correcting the movement of the student, guiding his/her limbs to the correct position, he/she is providing haptic feedback.

2.5.1 Visual Feedback

Visual feedback is the modality of feedback that takes advantage of sense of sight [134]. Visual feedback can use different kind of visual displays, including not only regular screens, but projectors or head-mounted displays (HMDs) [134].

- **Using Screens and projectors** [134]: A model of the movements can be displayed in screens and projectors, so the student can imitate the movements performed by a virtual teacher. The student can also interact with the system and select in which signals of the movement focus.
- **Augmented Reality (AR) and Virtual Reality (VR)** [134]: Using AR and VR techniques allows more interaction between the system and the user by creating immersive virtual

environments in which the user will perform his/her activities (VR), or by embedding virtual characters or elements into the real world (AR).

   - **Head-Mounted Displays (HMD)** [135]**:** The most common technique for both, AR and VR, is the use of head-mounted displays, that have become broadly available and relatively cheap thanks to the videogame industry. Commercial options include Oculus Rift[21] (VR) and Microsoft Hololens[22] (AR).
   - **Spatially Immersive Displays** [135]**:** A spatially immersive display consist in the use of a set of projectors or screens surrounding the subject that allows an immersive experience by showing the virtual environment directly around the subject. A well-known system with these characteristics is CAVE [136].

The use of AR and VR techniques has an important disadvantage, the virtual sickness [137], which consist in the presence of a set of symptoms related to the fact that the user is perceiving a sensation of motion inside the virtual environment that is not really happening.

The provided feedback can vary highly between applications and can consist in the use of flashes [138], virtual teachers [139], or even virtual mirrors [53] where the user can see him/herself. Visual feedback is the most used in motor learning [134] and have demonstrated its efficacy in several studies [134].

2.5.2 Auditory Feedback

Auditory feedback is the modality of feedback that takes advantage of the sense of hearing [134]. This kind of feedback can be provided either by generating natural language or by sounds that indicates when a movement has been executed properly or wrongly, or even by transforming motion characteristics into determined sounds:

- **Verbal feedback** [140], [141]: The device indicates the corrections or compliments the user using directly verbal feedback, as a teacher would do.
- **Alarms** [134]: An alarm sounds when a part of the body reaches a threshold. Going beyond this threshold means that the movement has not been executed properly. The alarm can also be triggered after the movement, indicating if it has been executed properly or not. For example, an alarm can sound if the elbow is stretched more than 45 degrees.
- **Sonification** [134], [142], [143]: The magnitudes and variables of the movement or the environment such as strength or speed are represented using sound, changing pitch and volume. For example, the pressure put over a piano key is represented by modulating the volume of the sound.

Thanks to the advances in Natural Language Processing (NLP), the user can also interact with the system by talking with it [53], [140].

2.5.3 Haptic Feedback

Haptic feedback is the modality of feedback that takes advantage of the sense of touch [134]. This kind of feedback is the less investigated in motor learning [134] and can be provided through vibrotactile devices or skin stretching devices.

- **Vibrotactile** [144]–[148]: The feedback is given through vibrations sensed by the skin.
- **Skin Stretching** [145]: The feedback is given by stretching the skin.

[21] Oculus Rift webpage: https://www.oculus.com/
[22] Microsoft Hololens webpage: https://www.microsoft.com/en-us/hololens

Any of the three types of haptic feedback could be used for indicating towards where a part of the body should be moved or indicating which part of the body requires a correction. Haptic feedback is also used in haptic render with the purpose of sensing virtual objects and can also be provided by robots [134]. Its use for martial arts has been already proposed elsewhere [22]. The fact that haptic feedback can be given to the full body by means of a “feedback suit” makes this kind of feedback promising [149], [150].

2.6 Characteristics of Learning in Martial Arts

When practicing and learning a martial art, the student normally does not do it alone. Other kind of agents participate in the learning, and can interact with the student, either by observing him/her and providing feedback, or by interacting with him/her physically. Three kind of agents participate in the learning of a martial art [10]:

- **Student:** The first agent is the own student, that will execute the movements.
- **Opponent:** The second agent is the opponent, who will interact directly with the student, forcing him/her to defend him/herself, providing feedback, or helping him/her to feel the movements in a realistic situation. The opponent can be a partner, the teacher or even an imaginary opponent when performing a form (kata).
- **External Observer:** This is an external agent that can evaluate and observe the movements executed by the student and the opponent. If the external observer provides feedback, normally is an authority figure or expert like the teacher, but it can be also an advanced partner.

Martial arts learning has some characteristics that make it perfect for creating an AI-based system for assisting the learning of psychomotor activities. The learning of a martial art implies different kinds of learning:

- **Vicarious learning** [151]: The student learns by imitating the movements performed by a model. For example, when the student imitates the teacher when performing a form (kata).
- **Repetition learning** [152]: The student learns and enhance his/her movements by repeating them. For example, when the student has learnt a kata and he/she have to enhance its execution. This kind of learning requires self-observation.
- **Reinforcement learning** [153]: The student learns by associating some preferred actions with rewards or undesirable actions with penalties. For example, when in a combat the student gains a point for hitting an adversary that has his/her guard opened, or losses a point when being hit by opening his/her own guard. The student will learn to hit when the guard of the opponent is opened and to close his/her own guard.
- **Observational learning** [154]**:** The student learns by watching or observing others executing the movements. It is not necessary to imitate those movements for learning them. For example, when a student observes the movement of an advanced students. And he/she can later execute those movements.

So, the learning of a martial art implies learning by different means, which are used too in the learning of any kind of psychomotor activity. AI, then, can assist the learning of a martial art by taking the role of the teacher, a partner or an opponent, and by giving feedback and helping the student to supervise his/her own movements.

2.7 Martial Arts Applications

In this section, a set of martial arts-based publications that involves at least one of the phases mentioned in Section 2.1 are reviewed to show which of the motion capture, modelling, analysis techniques or feedback approaches are being applied, and which phases are covered by those systems.

A total of 27 publications are reviewed (n=27) and the focus is put in the motion capture methods used, the modelling technique applied, the analysis algorithm employed, and the feedback approaches proposed. The publications have been selected from Google Scholar, ACM Digital Library, Elsevier, Springer and PubMed using as search terms the general name “Martial Arts” or specific types such as "Tai Chi” or “Tae Kwon Do”. The results can be seen in Appendix I and Appendix II and will be explained in a more detailed way in the next subsections.

As shown in Figure 15, the focus has been put mainly in recent publications for the last 6 years, but some earlier publications that were relevant have been included as well.

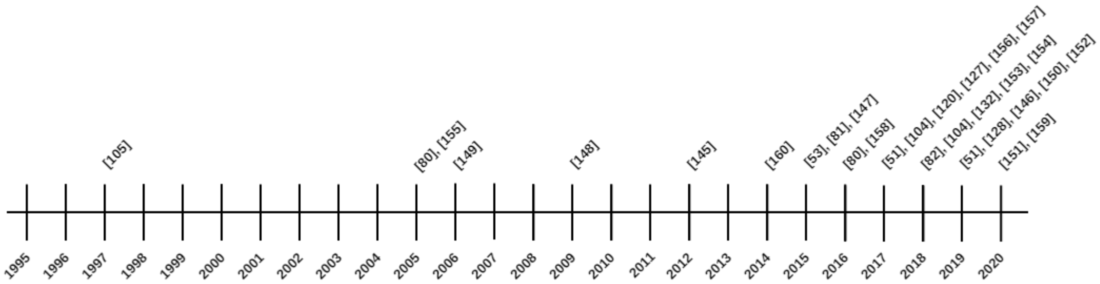

Figure 15. Distribution over time of reviewed publications.

Some of the publications reviewed may employ more than one technique in each phase, or directly do not employ any technique because they do not explore that phase, so the number of techniques explored may not be equal to the number of publications reviewed.

2.7.1 Martial Arts Studied in Martial Arts Publications

In the selected publications, 12 martial arts have been studied, with a total of 28 apparitions. Three of the publications have not specified which martial art is being studied, since their application is suitable for any martial art. In Table 1, it is shown the apparition frequency of each martial art in the studies. It must be noted that some publications involve the study of more than one martial art, so the total number of martial arts studied does not correspond with the number of publications.

As shown in the table, the most popular martial art in the selected documents is Karate (28,57% of apparitions), followed by Tai Chi (14,29% of apparitions) and Tae Kwon Do (10,71% of apparitions). The three of them encompasses more than half of the apparition of martial arts (53,57% of apparitions). This follows the popularity of Karate, being one of the most known martial arts and invited to the Japanese Olympics which are now to be held in 2021; the popularity of Tai Chi in research studies, being a martial art known by its health benefits and its easiness to deal with human motion analysis as it consists in individually performing postures; and the popularity of Tae Kwon Do, which is an Olympic sport since the year 2000.

Table 1. Frequency of martial arts in the reviewed publications.

Martial Art	Absolute Frequency	Relative Frequency (%)
Boxing[23]	2	7,14%
Karate	8	28,57%
Kendo	1	3,57%
Kick Boxing	1	3,57%
Krav Maga	1	3,57%
Kung Fu	1	3,57%
Muay Thai	2	7,14%
Seni Silat	2	7,14%
Shorinji Kempo	2	7,14%
Tae Kwon Do	3	10,71%
Tai Chi	4	14,29%
Tankendo	1	3,57%
Total	**28**	**100,00%**

2.7.2 Motion Capture in Martial Arts Applications

Table 2 shows the apparition frequency of motion capture methods in the reviewed publications. It can be seen how none of the methods mentioned in Section 2.2.4, which are less used either because obsolesce or either because stills under research, appear in the table. Auxiliary methods with purposes of labelling data or synchronizing streams have not been represented in this table since the purpose is not the obtention of human motion. Nonetheless, the full list of methods is compiled in the Appendix I.

Optical methods encompass the 87,50% of the used methods. It can be seen how the use of a set of cameras with reflective markers (using either a mocap suit or markers in determined parts of the body) is the most used method (50,00% of apparitions). Passive markers have not been used in the reviewed publications. Depth cameras are the following most used method (18,75% of apparitions). In Appendix I it can be seen how all apparitions of depth cameras utilizes Kinect sensors (either Kinect v1, which employs a structured light method, or Kinect v2, which employs a time-of-flight method), except for one that uses a Bumblebee-II sensor, which employs a stereo passive method. RGB sensors appeared with the same frequency as depth sensors, and it can be seen how two of six apparitions employed multiple RGB cameras, and the remaining four only one camera. One of the four that employed only one camera uses it in a drone that is orbiting around the user.

Muscular methods appeared three times in the reviewed publications. Those publications focus in the study of muscular response with the purpose of comparing the execution proficiency of users or groups.

Inertial methods only appear one time in the reviewed publications, and it is used in the obtention of skeletal information of the body using a mocap suit conformed of inertial sensors.

[23] Boxing is not traditionally considered a martial art. However, it has been included in this study since it is a contact sport and performs psychomotor activity related to martial arts.

Table 2. Frequency of apparition of motion capture methods in the reviewed publications.

Motion Capture Method	Absolute Frequency	Absolute Frequency Grouped	Relative Frequency (%)	Relative Frequency Grouped (%)
Set of Cameras with reflective markers	16	28	50,00%	87,50%
Depth Camera	6		18,75%	
RGB Camera	4		12,50%	
Multiple RGB Cameras	2		6,25%	
EMG	3	3	9,38%	9,38%
Inertial	1	1	3,13%	3,13%
Total	**32**	**32**	**100,00%**	**100,00%**

2.7.3 Motion Modelling in Martial Arts Applications

As shown in Table 3, the predominant technique for modeling human motion is the use of series of 3D skeletons over time (71,88% of apparitions). It has appeared in publications that capture motion using a set of cameras and markers, using depth cameras, using RGB cameras and in the only publication that use inertial sensors. This is due to the fact that the most natural way of modeling a human body is by representing its joints in space, as shown in Section 2.3.1.2.1. Publications using EMG for measuring muscular activity model it as a data series of sampled electrical activity over time (9,38% of apparitions).

When the capture of the human body has been done by using RGB cameras, the most common technique for modeling the human body is 2D silhouette (12,50% of apparitions). The generated silhouette can be later used for generating a 3D skeleton.

The use of 3D silhouettes appeared just in two publications. One of them using a stereo camera and obtaining a cloud of points of the silhouette [106], and the other uses multiple RGB cameras [107].

Table 3. Frequency of apparition of motion modelling methods in the reviewed publications.

Motion analysis method	Absolute Frequency	Relative Frequency (%)	Action analysis techniques	Absolute Frequency	Relative Frequency (%)
3D Skeleton	23	71,88%	**Random forest**	1	25,00%
3D silhouette	2	6,25%	**RNN**	1	25,00%
2D silhouette	4	12,50%	**Templates**	1	25,00%
Data series	3	9,38%	**Motion graphs**	1	25,00%
Total	**32**	**100,00%**	**Total**	**4**	**100,00%**

Regarding action modeling, only four publications model actions explicitly. This could be due to the fact that other publications just focus in one movement or action, so they do not need to model it since they already know which action has been executed, or because some publications directly compare the data of the motion with stored data without worrying about the executed action. The four methods used for modeling actions are: Random forest, RNN, Motion templates and Motion graphs.

2.7.4 Motion Analysis in Martial Arts Applications

Table 4 shows the frequency of use of motion analysis methods. The analyzed publications mainly employ mathematical or statistical methods with the purpose of extracting characteristics of the movement for comparison or classification purposes (64% of apparitions).

Machine Learning techniques have been used only three times (12% of apparitions) and consist in RNN for forecasting of motion and Random Forest Classification and SVM for classification purposes.

Since the obtained data consist in time series as explained in Section 2.3.1.3, filtering techniques such as Kalman Filters, threshold values or Moving averages has been used, alongside with DTW for comparing time series.

Table 4. Frequency of apparition of motion analysis methods in the reviewed publications.

Motion analysis method	Absolute Frequency	Relative Frequency (%)
Mathematical/Statistical	16	64,00%
ML	3	12,00%
Filtering	3	12,00%
DTW	3	12,00%
Total	**25**	**100,00%**

Even when the reviewed publications are recent, machine learning techniques have not broadly entered in martial arts analysis, and only one of the reviewed publications employs Artificial Neural Networks with the purpose of forecasting poses.

2.7.5 Feedback in Martial Arts Applications

Reviewing the publications, it can be seen in Table 5 how visual feedback is still the most used of the three modalities of feedback (88,24% of apparitions). It can be seen how advances in AR and VR techniques have influence in immersive visual feedback (60% of apparitions). The most used visual feedback technique is the use of Head-Mounted Displays (HMD) in VR (35,29% of apparitions), from which 5 are VR HMD devices and one of them is an AR HMD device (Microsoft Hololens). CAVE immersive environments are also used for immersive visual feedback (20% of apparitions). The use of screens and projectors remains dominant in this field (35,29% of apparitions). One of the publications that employs screens uses AR by embedding the player into the game.

Auditory feedback has only been used two times (11,76% of apparitions), of which one of them employs verbal feedback to provide indications to the user, and the other employs sounds to indicate that the user has hit with a virtual character.

Haptic feedback has not been found between the reviewed publications.

Table 5. Frequency of apparition of motion capture methods in the reviewed publications.

Motion Feedback Method	Absolute Frequency	Absolute Frequency Grouped	Relative Frequency (%)	Relative Frequency Grouped (%)
Visual - Screen	6	15	35,29%	88,24%
Visual - HMD	6		35,29%	
Visual - CAVE	3		17,65%	
Auditory - Sound	1	2	5,88%	11,76%
Auditory - Verbal	1		5,88%	
Total	**17**	**17**	**100,00%**	**100,00%**

2.7.6 Other Characteristics of Martial Arts Applications

As shown in Appendix II, there are some characteristics about these publications that can draw our attention. A martial art is an activity in which normally practitioners interact with each other, and this interaction takes an important part in the learning of the martial art. However, only 3 of the 27 reviewed publications (i.e., [52], [155], [156]) consider this possibility, the rest of publications focus in the study of an isolated user executing particular movements or sets of movements.

Some martial arts also involve the use of different kind of weapons, regardless, only 2 of the 27 reviewed publications focus in the study of martial arts that use a weapon (i.e., [52], [155]), and both weapons are wooden swords, even when there are hundreds of different weapons used in martial arts.

An application able to assist in the learning of a martial art should be a real-time application, able to provide feedback to the user so he/she can correct his/her errors or improve his/her movements. Of the 27 reviewed applications, just 12 are real-time applications and can be used when assisting the learning of a martial art ([83], [106], [162], [163], [138], [139], [141], [157]–[161]).

It is important to note that none of the reviewed publications has employed an Android device neither as a motion capture system, nor as a processing device.

2.8 Summary

In the previous sections, a review of the bibliography concerning motion capture, modelling and analyzing, as well as feedback methods has been carried out. Then, the different characteristics of a martial art have been presented and a review of the bibliography of publications related with studying martial arts or assisting the learning of a martial art has been carried out.

As shown, there are a lot of opportunities with AI-based techniques and methods, and the use of martial art for testing could be a great solution, since a martial arts implies interaction between users and the environment with a predefined set of movements that can be clearly identified, the use of objects such as weapons, and the characteristics of the learning of a martial art make it perfect to extrapolate the results of the studies to other psychomotor activities.

The review of publications included here and related with martial arts is the most revealing since it can be seen which are the trends in this field, and complements the revision of the bibliography carried out previously in [124]. It is important to note how AI techniques still have not entered in this field, which means a great opportunity for researchers. Most of the publications focus just on studying a set of movements or propose techniques for capture, modeling or analyzing motion, or even for giving feedback, but just a few of them are aiming to really assist in the learning of a martial art. The use of inertial sensors and RGB cameras in martial arts applications is still young, and the use of Android devices, such as smartphones, could facilitate it by providing a set of frameworks, libraries and hardware that can be easily manipulated. As shown, there is a lack of studies focusing on the use of weapons or in the interaction between users. Thus, it could also be a great opportunity to research the capture, modelling and analysis of motion that involves objects, weapons, the environment or multiple users.

In the following sections, some of the methods and techniques of the state of the art reviewed here are investigated to see if they are appropriate to the research problem at hand (i.e., ***assist in the learning of martial arts movements***). For capturing motion, the inertial sensors of an Android device are to be considered. As seen in previous sections, inertial sensors are appropriate to overcome some disadvantages of other methods like occlusion, complex configurations, cost, or mobility. Since inertial sensors capture sequences of information (acceleration over time, rotation over time…), the data should be modelled and preprocessed using a technique for modeling time series. Specifically, in this research project, the potential of Exponentially Weighted Moving Averages for smoothing the curves and deleting noise is studied. The advantage of this method is that it has a parameter that allows to control the amount of smoothness applied. Once the data has been preprocessed, the potential of three different types of neural networks for classifying and recognizing the movements is studied. As aforementioned, Neural Networks have a series of advantages such as the ability of learn an implicit model and self-organize the learned characteristics that makes them suitable for analyzing complex motion patterns. Further, there are some types of neural networks that have been specifically designed for analyzing time series such as 1-Dimensional Convolutional Neural Networks and Recurrent Neural Networks, and the fact that transfer learning can be used for transferring and applying a learned model to other martial arts, other psychomotor activities, or even other completely different fields, makes neural networks a great candidate for assistant the learning a martial art. Finally, the benefits of auditory and haptic feedback are going to be researched using the characteristics of an Android device (speaker and vibrator). Visual feedback, even being the most effective of the three feedback modalities reviewed, is not considered since due to the experimental setting defined, the screen of the Android device is not accessible during the whole execution of the movements, as the arm moves to different positions. Moreover, the idea is to study other kinds of feedback that can enrich the response to the user though other sensorial channels.

3 KSAS, AN APPLICATION FOR LEARNING A MARTIAL ART

With the purpose of demonstrating the research hypothesis of this Master's Thesis, which is that it is possible to use AI techniques to recognize, model and analyze human motion, as well as give feedback to the user with an Android device, I have developed an AI-based prototype for assisting the learning of a set of movements from American Kenpo Karate[24] called KSAS (Kenpo Set Assisting System). Those movements are part of the American Kenpo Karate's blocking set I. The full source code of KSAS is publicly available on GitHub[25].

In the following sections, the movements used from the American Kenpo Karate's blocking set I are shown. Following, the process of motion capture is explained, alongside with some information about the users that participated in the capture to train the neural networks used to model and analyze the movements. Then, the modelling and analysis techniques, and the feedback strategies employed are explained. Finally, the structure of the full application is shown.

3.1 American Kenpo Karate

American Kenpo Karate [10] is a martial art developed by the martial artist Edmund K. Parker in the 20th century. The development of this martial art was highly influenced by physics and biomechanics, and many scientific principles such as inertia, mass or potential have been included directly in its curriculum, which has evolved till our days.

The curriculum of American Kenpo Karate, as in many other martial arts, is divided into belt degrees. There are ten belt degrees between white and black, and in each belt a set of techniques, forms (katas), and sets of movements are studied. Part of the first set taught in American Kenpo Karate, Blocking Set I, will be the set[26] assisted by the version of KSAS developed for the Master's Thesis.

3.1.1 Blocking Set I

Blocking Set I is the first set learned by an American Kenpo Karate student. This is because the teachings of first belt degree of American Kenpo Karate (i.e., white belt) puts the emphasis in defensive movements.

This set is formed by six blocking movements: Upward block, Inward Block, Outward Extended Block, Downward Block, Rear Elbow Block and Push Down Block. Since most of the participants in this research had never practice American Kenpo Karate before, the Push Down Block has been suppressed for making the data capture sessions more agile for beginners since prior to execute this movement, the arm already returns to the original position, and because it is the only block that involves the use of the open hand and not the punch, which also could confuse beginners. Then, the five collected movements can be seen in Figure 16.

---

[24] Even when the name of this martial art is American Kenpo Karate, it is a different martial art from Kenpo and from Karate, even when the three of them are related. It is a modern martial art developed in US in the 50s.

[25] KSAS source code: https://github.com/AlbertoCasasOrtiz/KSAS

[26] Note that set is a group of specific movements in Kenpo Karate. See the glossary for more information.

Figure 16. American Kenpo Karate's Blocking Set I.

3.2 Information About Users

KSAS must be able to assist the learning of Blocking Set I, so the application must be able to learn the characteristics of the set before. Thus, the first step in the development of KSAS is to capture the necessary movements. For this purpose, twenty healthy volunteers have been recruited (n = 20). The motion capture process has followed the guidelines of the Ethics Committee of the UNED to carry out the study, as well as the indications of the Master. The signed consent and the questionnaire that the participants had to fulfill are annexed in Appendix III. The participants were asked to learn the Blocking Set I (if they do not already know it) and execute it meanwhile it is captured by a motion capture device.

Before starting the motion capture process, all the participants were asked to give their signed consent allowing the use of their data for this thesis and for subsequent researches. The participants were asked to fulfill a questionnaire regarding some demographic information and their experience in martial arts as well. All the participants were adults aged between 21 and 65 years (Table 6). The mean age is 34.6, with a standard deviation of 12.22. Four of the participants were women, and the remaining sixteen were men (Table 7). The participants were asked which was their dominant hand. 16 subjects were right-handed, and 4 subjects were left-handed (Table 8).

Eight of the participants are currently practicing a martial art. Six of them practice American Kenpo Karate, one of them practices Aikido, and one of them practices Boxing. Eight of the participants have practiced at least one martial art in the past and only one of the participants who is now practicing a martial art have practiced a different martial art in the past. This gives a total of 16 participants with some experience in martial arts and 4 participants without experience practicing martial arts. The martial arts practiced currently and in the past by the participants can be seen in Tables 9 and 10 respectively.

The participants' data were collected in scientific events about martial arts, in martial arts dojos and in private sessions. The data of eight of the participants were collected during the event that took place in the installations of UNED as part of the "Semana de la Ciencia y la Innovación 2019 / Week of Science and Innovation 2019", which is held annually in November in Madrid (Spain). In this last edition, the student carrying out this Master's Thesis and the director of it were the keynotes of the talk entitled "Ciencia y Datos en las Artes Marciales / Science and Data in Martial

Arts" [27]. The data of six of the participants were collected in Kenpo Karate Studio, the gym where the Master's Thesis student practices American Kenpo Karate. The data of the rest of the participants were captured in private sessions.

For storing the information about the participants obtained in the questionnaire, a local MySQL relational database has been developed. This allows to obtain information in tables using SQL queries. The structure of the database can be seen in Appendix III.

Table 6. Age range of the users.

Age range	Absolute Frequency
<18	0
18-24	2
25-34	11
35-44	2
45-54	3
>65	2
Total	**20**

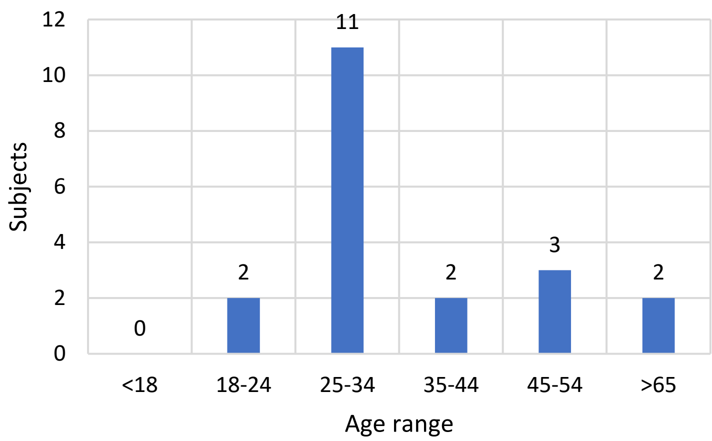

Figure 17. Age range of the users.

Table 7. Gender of the users.

Gender	Absolute Frequency
Woman	4
Man	16
Total	**20**

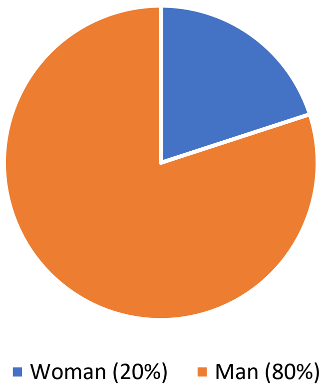

Figure 18. Gender of the users.

[27] Event website: http://www.madrimasd.org/semanaciencia2019/actividad/ciencia-y-datos-en-las-artes-marciales

Table 8. Dominant hand of the users.

Dominant hand	Absolute Frequency
Left-handed	6
Right-handed	14
Total	**20**

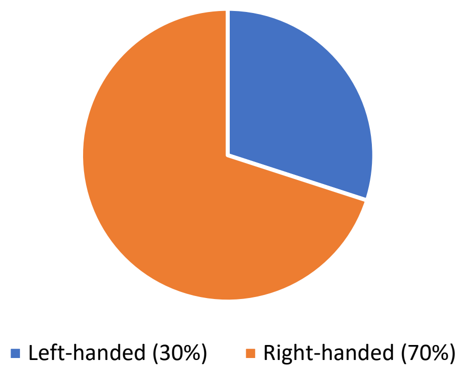

Figure 19. Dominant hand of the users.

Table 9. Martial Arts currently practiced by the users.

Martial Art	Absolute frequency
Aikido	1
American Kenpo Karate	6
Boxing	1
Total	**8**

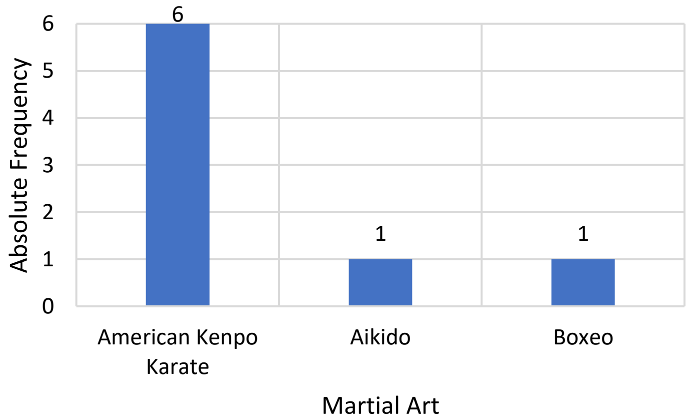

Figure 20. Martial Arts currently practiced by the users.

Table 10. Martial arts practiced in the past by the users.

Martial arts	Absolute Frequency
American Kenpo Karate	2
Boxing	2
Judo	2
Kung Fu	2
Kali	1
Karate	1
Silat	1
Tae Kwon Do	1
Total	**12**

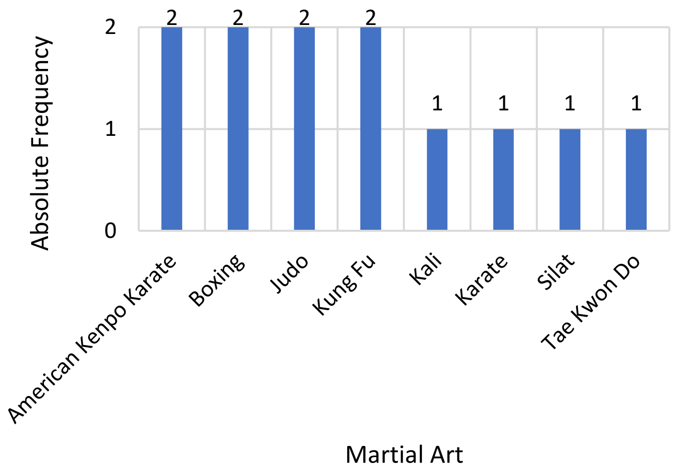

Figure 21. Martial arts practiced in the past by the users.

3.3 Capturing the Movements

As mentioned in the introduction, an Android device that already has accelerometer, gyroscope and magnetometer is used for motion capture. An Android device has been selected because of its simplicity, the availability of libraries for controlling the sensors and the ease of code and deploy new applications. Furthermore, android devices are broadly available and relatively cheap. The android device used for motion capture is a smartphone Xiaomi Mi A2.

The device is put as a wearable in the forearm of the user using an accessory band similar as the bands used by runners and athletes (Figure 22). Before starting the motion capture process, all the participants have been asked to execute warming exercises. Then, the Blocking Set I was explained to the participants and they could practice it a few times. Finally, the Android device was put in the forearm of the participants, and they were asked to execute the Blocking Set I, mirroring the simultaneous execution of the set by the researcher. The motion of both arms was captured twice per each participant: the first one for testing that everything was working properly, and the second one for storing the captured data.

Figure 22. Device put as wearable in the arm.

3.3.1 An Application for Capture Motion

For capturing the motion of the participants using the sensors of an android device, an application called Motion Capturer has been developed. The development of this application started with the study of how the sensors and the Android libraries work, as well as understand the captured data. This application is publicly available on GitHub [28]. The GUI of the app is very simple (Figure 23) and consists of two buttons for starting and stopping the motion capture, and a text field for inserting the ID of the participant.

[28] Motion Recorder Android source code: https://github.com/AlbertoCasasOrtiz/Motion-Recorded-Android

Once the start button has been clicked, a message appears in the screen indicating that the motion capture process has started. The application then starts to read data from the sensors and once the stop button has been clicked, stores it into a CSV file (Figure 24) in the external storage of the Android device. This CSV file is formed by eighteen columns containing sequences of rows whose length may vary depending of the time that the motion capture process has been working. The sample rate of the sensors has been established at SENSOR_DELAY_GAME (20 milliseconds). The information of each column in the CSV, obtained directly from the device, is described as follows [29]:

- **Accelerometer_x:** Acceleration in the x axis (m/s^2).
- **Accelerometer_y:** Acceleration in the y axis (m/s^2).
- **Accelerometer_z:** Acceleration in the z axis (m/s^2).
- **Gravity_x:** Gravity force in the x axis (m/s^2).
- **Gravity_y:** Gravity force in the y axis (m/s^2).
- **Gravity_z:** Gravity force in the z axis (m/s^2).
- **Gyros_x:** Rate of rotation around x axis (rad/s).
- **Gyros_y:** Rate of rotation around y axis (rad/s).
- **Gyros_z:** Rate of rotation around z axis (rad/s).
- **Lin_accel_x:** Acceleration in the x axis without gravity (m/s^2).
- **Lin_accel_y:** Acceleration in the y axis without gravity (m/s^2).
- **Lin_accel_z:** Acceleration in the z axis without gravity (m/s^2).
- **Game_rot_vector_x:** Rotation vector component along the x axis (No unit).
- **Game_rot_vector_y:** Rotation vector component along the y axis (No unit).
- **Game_rot_vector_z:** Rotation vector component along the z axis (No unit).
- **Magn_field_x:** Geomagnetic field strength along x axis (μT).
- **Magn_field_y:** Geomagnetic field strength along y axis (μT).
- **Magn_field_z:** Geomagnetic field strength along z axis (μT).

[29] Information about sensors: https://developer.android.com/guide/topics/sensors/sensors_motion

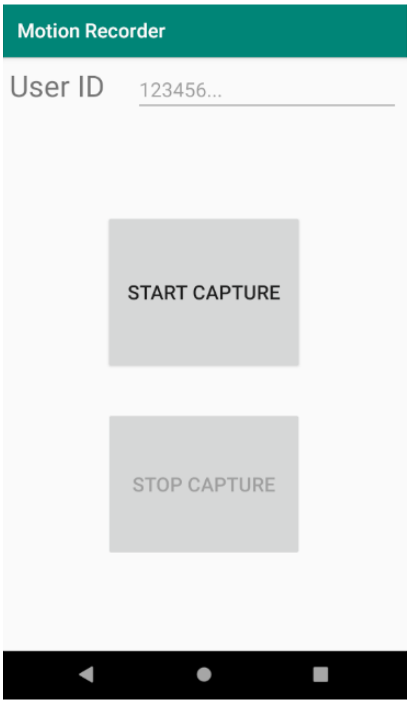

Figure 23. UI of the motion capture application.

	A	B	C	D	E	F	G	H	I	J
1	accelerometer_x	accelerometer_y	accelerometer_z	gravity_x	gravity_y	gravity_z	gyros_x	gyros_y	gyros_z	lin_accel_x
2	4.099	-3.067	8.514	4.226	-2.962	8.339	-0.001	0.014	0.145	0.156
3	3.996	-3.052	8.409	4.213	-2.974	8.341	-0.009	0.03	0.13	0.115
4	4.127	-2.914	8.361	4.198	-2.987	8.344	-0.016	0.053	0.119	0.164
5	4.257	-2.938	8.461	4.181	-3.003	8.347	-0.03	0.079	0.115	0.095
6	4.333	-3.026	8.263	4.166	-3.018	8.349	-0.042	0.097	0.09	0.152
7	4.42	-2.988	8.203	4.143	-3.032	8.355	-0.066	0.135	0.033	0.369
8	4.367	-2.995	8.421	4.112	-3.044	8.366	-0.097	0.222	-0.057	0.505
9	4.4	-3.062	8.469	4.075	-3.053	8.381	-0.131	0.306	-0.165	0.767
10	4.314	-3.045	8.37	4.048	-3.055	8.393	-0.164	0.314	-0.303	1.418
11	4.355	-3.043	8.232	4.061	-3.044	8.391	-0.189	0.19	-0.521	2.439
12	4.549	-3.072	8.361	4.142	-3.004	8.366	-0.19	-0.076	-0.854	3.468
13	4.654	-3.165	8.363	4.311	-2.907	8.314	-0.107	-0.431	-1.319	4.273
14	4.879	-3.199	8.253	4.611	-2.725	8.215	0.076	-1.085	-1.838	4.153
15	5.504	-3.139	8.289	5.102	-2.448	8.009	0.373	-2.331	-2.089	2.178
16	6.541	-3.242	8.435	5.845	-2.07	7.597	0.875	-4.152	-1.961	1.426
17	7.657	-3.4	8.801	6.913	-1.546	6.781	1.668	-6.928	-1.945	-0.358
18	8.641	-3.335	9.455	8.116	-0.933	5.425	2.648	-9.505	-1.785	-1.529
19	8.834	-2.961	10.362	9.161	-0.368	3.48	3.776	-11.558	-1.222	-2.974
20	7.369	-1.29	11.452	9.752	0.009	1.033	4.827	-12.979	-0.75	-4.262
21	7.381	1.86	10.398	9.661	0.122	-1.681	5.333	-13.83	-0.612	-4.035
22	6.689	3.946	10.001	8.754	0.093	-4.419	5.577	-14.868	-1.521	-1.629
23	6.747	2.255	9.23	7.344	0.115	-6.498	5.616	-12.919	-3.845	-2.023
24	6.359	2.265	7.317	5.655	0.075	-8.012	5.292	-11.692	-5.477	-0.875
25	5.65	2.959	4.104	3.802	-0.234	-9.037	4.952	-11.12	-5.543	2.486
26	5.758	5.021	1.114	2.215	-0.869	-9.514	4.918	-9.006	-4.997	3.305
27	7.214	6.232	5.828	1.11	-1.756	-9.584	5.123	-6.396	-4.439	0.477
28	5.356	5.076	4.849	0.319	-2.83	-9.384	5.628	-4.959	-3.703	-1.993
29	4.764	4.662	7.099	-0.449	-3.989	-8.948	6.059	-5.094	-2.924	-3.916
30	6.234	3.733	12.783	-1.336	-5.108	-8.264	6.025	-6.175	-2.397	-6.701
31	5.461	-1.063	13.262	-2.235	-6.105	-7.342	5.78	-7.145	-2.104	-5.743
32	1.547	-6.11	9.57	-2.905	-6.884	-6.351	4.836	-7.148	-2.263	-3.504
33	-1.688	-7.953	4.413	-3.159	-7.444	-5.547	3.43	-5.401	-2.365	3.396
34	-4.346	-9.775	-4.972	-2.919	-7.822	-5.145	2.154	-1.497	-2.199	7.495
35	-7.985	-14.152	-9.957	-2.635	-7.952	-5.098	0.884	1.356	-0.711	2.1
36	-7.911	-17.76	-16.857	-2.6	-7.826	-5.307	-0.415	2.471	1.414	-1.052
37	-6.357	-21.062	-22.836	-2.556	-7.669	-5.552	-0.85	2.185	1.142	-3.627
38	0.263	-20.904	-19.539	-2.536	-7.592	-5.665	-0.618	0.578	0.132	-5.296
39	4.578	-20.808	-16.604	-2.606	-7.587	-5.641	-0.234	-1.023	-0.33	-0.538
40	-0.556	-19.989	-16.22	-2.661	-7.611	-5.583	-0.087	-1.252	-0.521	2.018

Figure 24. First rows of a captured movement in a CSV file.

3.4 Modeling the Dataset

In the early stages of the research, the use of Labanotation (Section 2.3.2) for modelling the data was proposed[30], but Labanotation requires to know the relative position of the modeled parts of the body to indicate the direction towards it is moving, and obtaining the relative position from inertial devices, combining information from accelerometer, gyroscope and magnetometer, can be a really complex task that involves numerical integration methods, which generate accumulative error that cannot be easily avoided. Since the data obtained from the Android device already comes as a time sequence, it is easy to model the data as it is and then take advantage of it using time sequence analysis methods.

To prove that the hypothesis of this Master's Thesis is correct, and that AI techniques (specifically Artificial Neural Networks as mentioned in Section 2.8) can assist in the learning of a martial art, three classifiers based on the use of artificial neural networks (ANN), convolutional neural networks (CNN) and recursive neural networks (RNN) are used to test which of them is more suitable for this problem. Since the data is analyzed by means of Artificial Neural Networks, the model of the data must be done with the purpose of making the data more understandable, organized, and ease its analysis using the mentioned classifiers. It is important to remember that those methods learn characteristics and models of their input data, but those internal models are extremely difficult to understand or visualize, and that is beyond the scope of this research. However, those methods offer a great plasticity when classifying complex numerical sequences, especially Recursive Neural Networks (RNN). Further, once a good model has been learned and an appropriate set of hyperparameters has been selected, transfer learning techniques can be applied for learning classifiers able to solve new tasks. The following subsections **describe the research process followed to identify the most appropriate technique to model the data**, which consist in: i) label and identify each one of the movements, ii) normalize the data (optional), iii) smooth the time sequences, and iv) pad the sequences.

The methods selected and employed for label the movements, normalizing the data, smoothing the curves and padding the sequences are described in the following sections.

3.4.1 Labelling and Segmenting the Movements

The application for capturing the data takes care of the labelling of each column of the CSV, and since a CSV file is generated per each arm, we can separate the left-handed executions from the right-handed executions. The motion of 20 participants was captured per each arm, so a total of 40 CSV files containing a set of eighteen time series (one per sensor signal) was obtained.

Each set of time series represents characteristics of an execution of the Blocking Set I. After representing the data as a line chart, I noticed that in the representation of the three spatial axes measured by the gyroscope it can be seen each movement isolated from the others since there is a little pause between movements that sets the values measures by the gyroscope to 0 (Figure 25). Since the order in which each movement has been executed is known, the information given by the gyroscope can be used as a reference for manually segmenting the time series and then labelling the movements in each piece, separating the results in different CSV files.

[30] A paper regarding the use of Labanotation for modelling martial arts movements was presented for UMAP 2019. A research of how martial arts could be modeled using this movement notation was carried out, and an expert on Labanotation participated in the study. Unfortunately, the paper was considered immature and was not accepted.

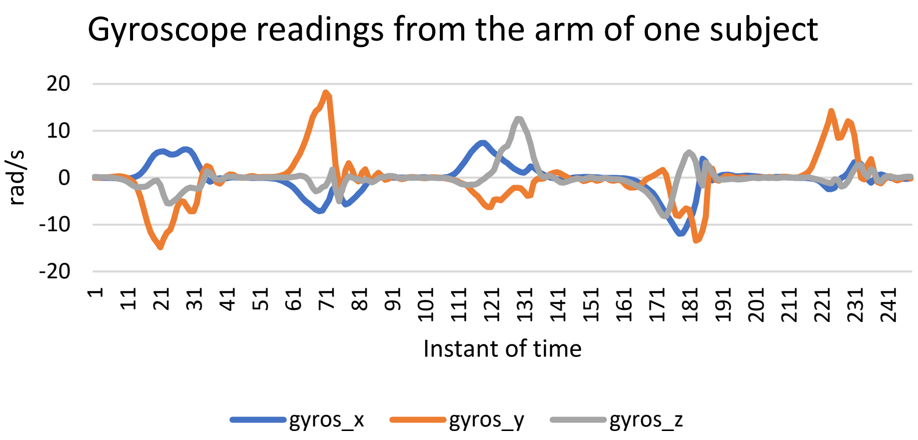

Figure 25. Gyroscope readings from the arm of one participant, showing 5 different movements.

Once the movements performed by each participant has been isolated and labeled, a set of 10 movements is obtained per each participant, five per arm. This gives a total of 200 movements (5 x 2 x 20) that will conform the dataset. Aside from the movements, KSAS should be able to recognize when a user is not executing any movement, so examples of static noise captured by the sensors have been obtained from the beginning and the end of the executions, giving a total of 240 movements (200 + 2 x 20). Each of the movements (CSV files) is then a record of the dataset, and each column (Sequence) is a variable. Since there are five different movements and static noise, the number of classes of the dataset is 6.

3.4.1.1 Visualizing the movements

Once the movements have been isolated and labelled, we can visualize them. Below, representations of all the characteristics of a recorded movements can be seen in Figure 26. As it can be seen, it is difficult to visualize those characteristics in the same chart due to the differences in values intervals and the high density of lines.

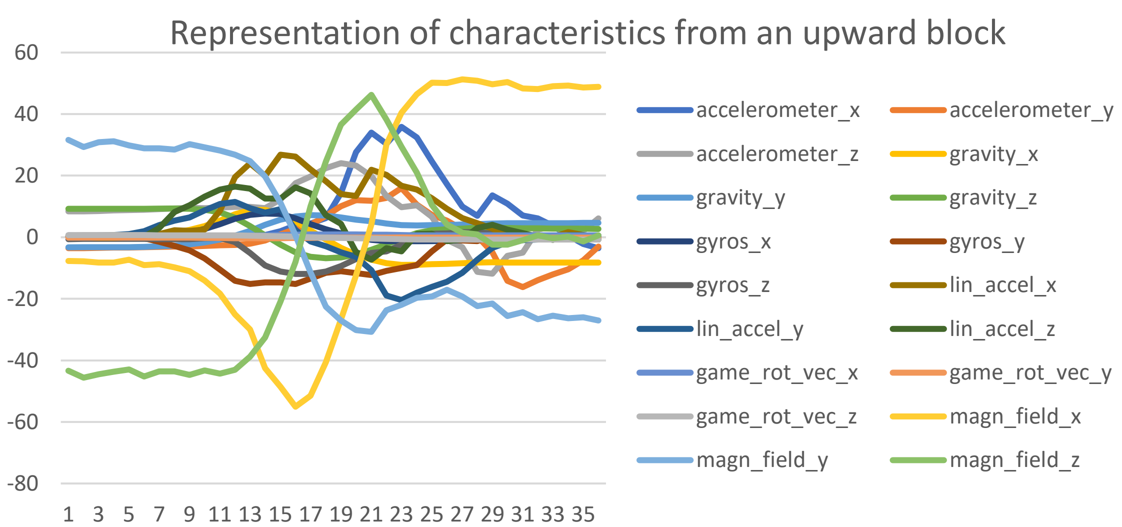

Figure 26. Representation of all the measured characteristics in the same chart.

A solution for this is to visualize each characteristic in an isolated chart. In the next pages, the characteristics of an Upward Block executed with a right hand are depicted, In particular, acceleration, gravity, rotation, linear acceleration, gamer rotation vector and magnetic field in each of the three Cartesian coordinate axes are represented in Figures 27, 28, 29, 30, 31 and 32 respectively.

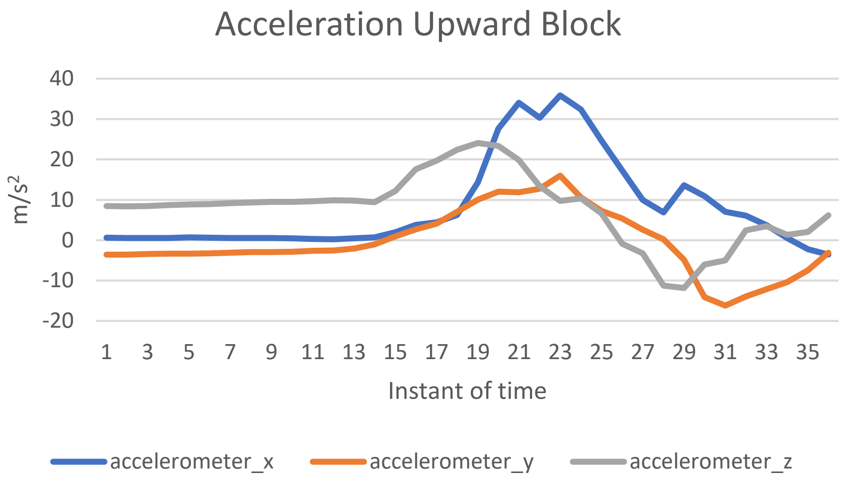

Figure 27. Acceleration measured in each axis for an Upward Block.

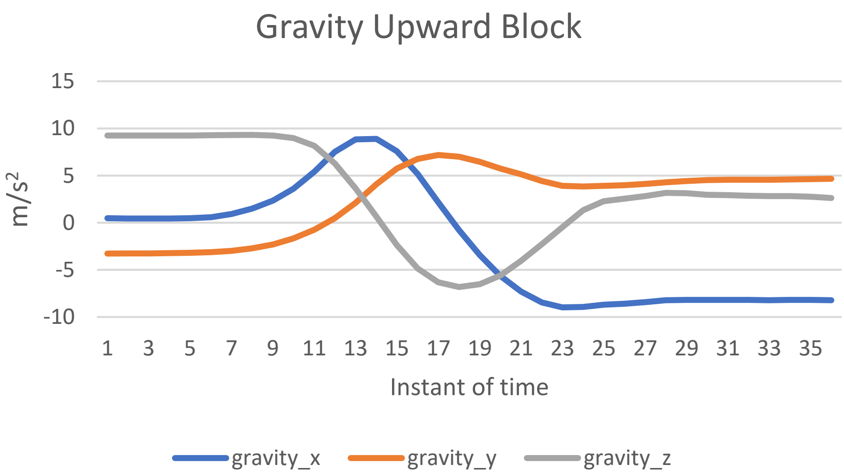

Figure 28. Gravity measured in each axis for an Upward Block.

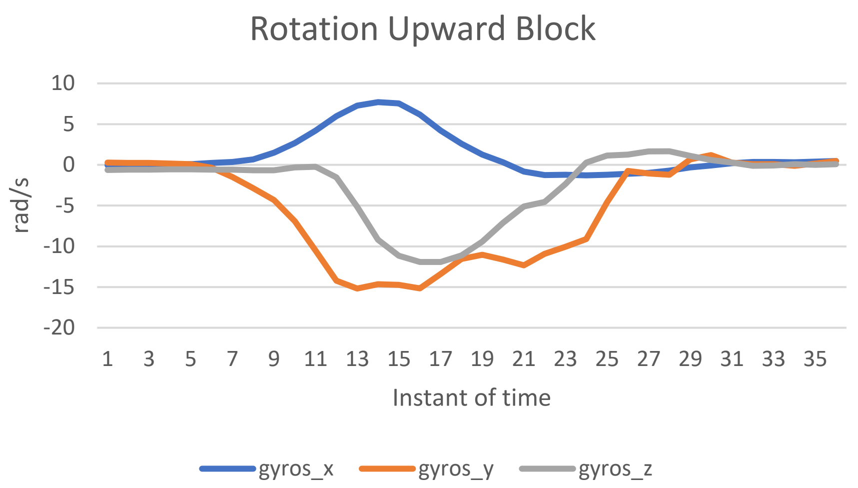

Figure 29. Rotation measured in each axis for an Upward Block.

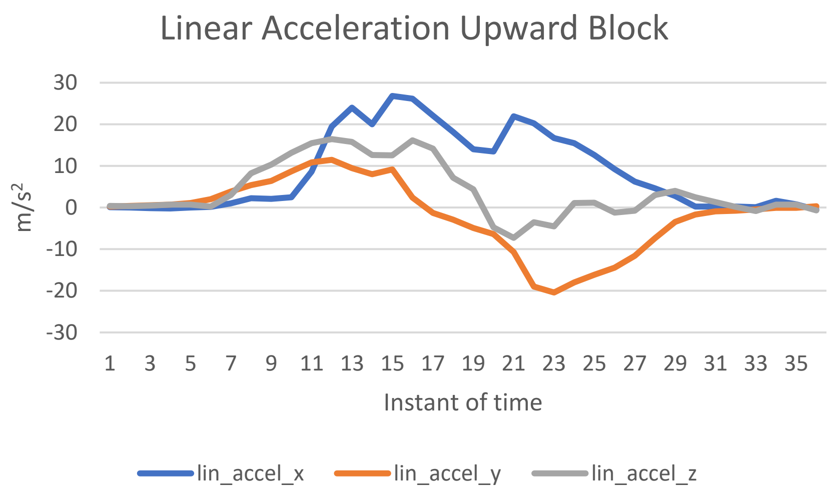

Figure 30. Linear acceleration measured in each axis for an Upward Block.

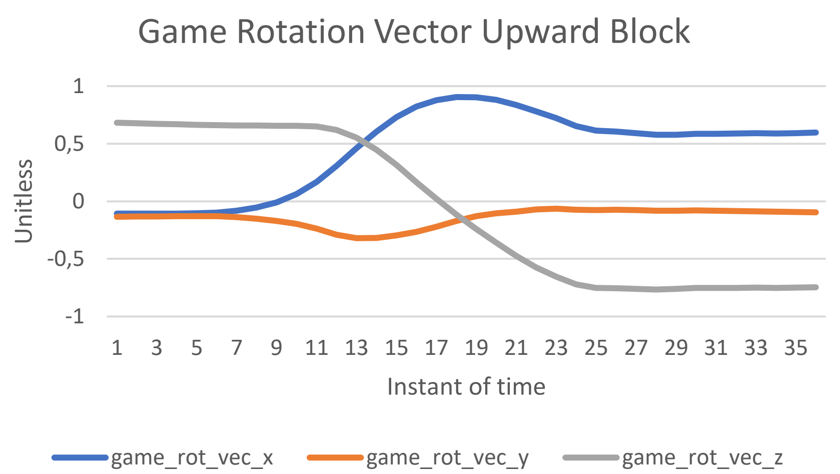

Figure 31. Game rotation vector measured in each axis for an Upward Block.

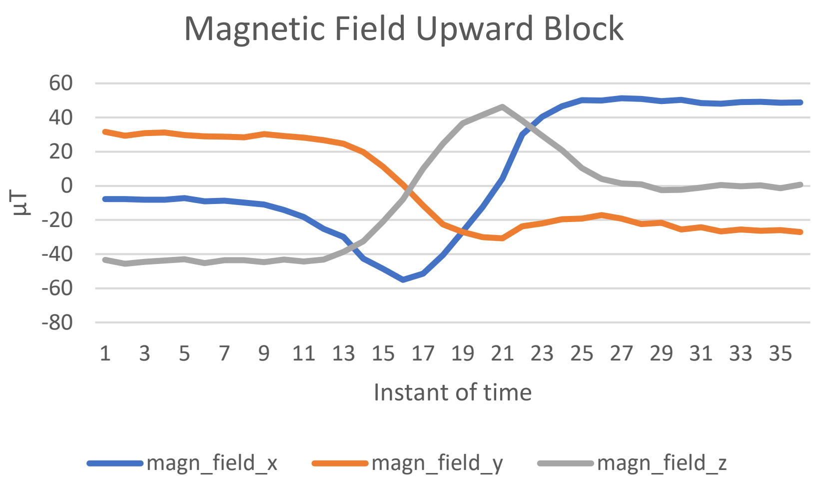

Figure 32. Magnetic field measured in each axis for an Upward Block.

It can be seen in the previous charts that there are two kinds of characteristics:

- The first kind of characteristic is the one that has a static value that changes during the movement and returns to the original value. Rotation measured by a gyroscope (Figure 29) is a good example since when the participant is not moving the values of rotation in the three axes is 0, and changes during the movement just to return to 0 when the user finish the movement. Acceleration and linear acceleration also belong to this kind of characteristic.
- The second kind of characteristic is the one that has an initial value that changes during the movement and remains in the new value. An example of this kind of characteristic is the magnetic field (Figure 32). It can be seen how the value of the axis x starts with a value near 0, changes during the movement, and find it new value near 40. Gravity and game rotation vector also belong to this kind of characteristic.

Using this information, we can then identify the different movements by comparing those characteristics. Doing it manually is a difficult task since the quantities measured in each axis may vary between different executions and participants, and the fact that the executions may come from both arms and the measures of the different axis for the same movement may change as well. I have noticed that the best characteristics for manually comparing movements is the rotation measured by the gyroscope. In Figures 33, 34, 35, 36, 37, 38, 39, 40, 41 and 42, the rotation measured for executions of the five movements by two different participants (one beginner and one expert) is shown. When tried for similar movements it produces similar patterns changing only in elongation or intensity.

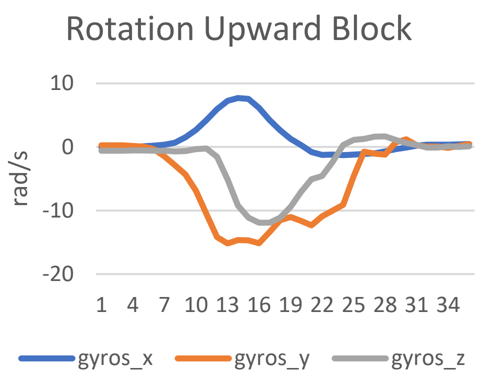

Figure 33. Rotation of an Upward Block executed by a beginner student.

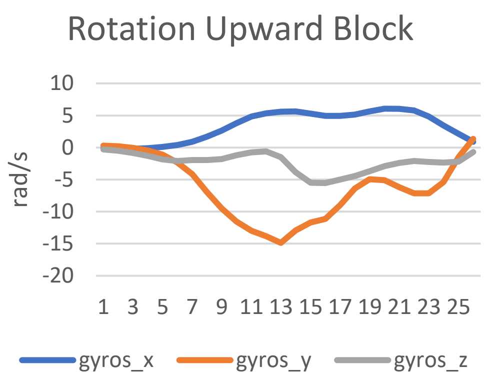

Figure 34. Rotation of an Upward Block executed by an experienced student.

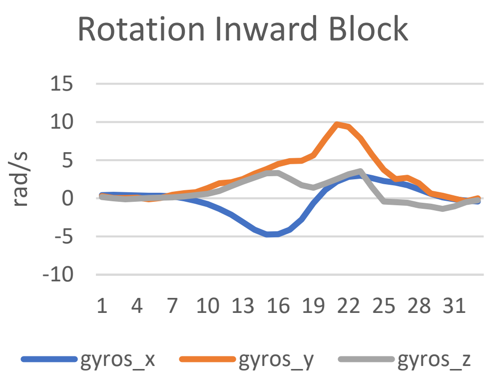

Figure 35. Rotation of an Inward Block executed by a beginner student.

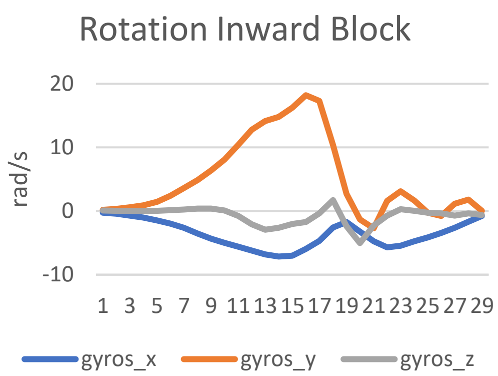

Figure 36. Rotation of an Inward Block executed by an experienced student.

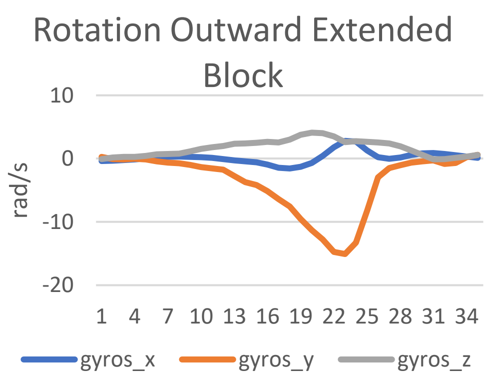

Figure 37. Rotation of an Outward Extended Block executed by a beginner student.

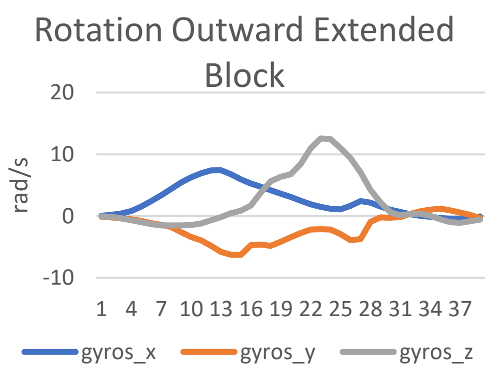

Figure 38. Rotation of an Outward Extended Block executed by an experienced student.

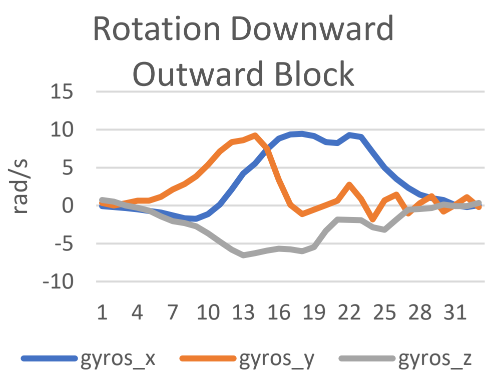

Figure 39. Rotation of a Downward Outward Block executed by a beginner student.

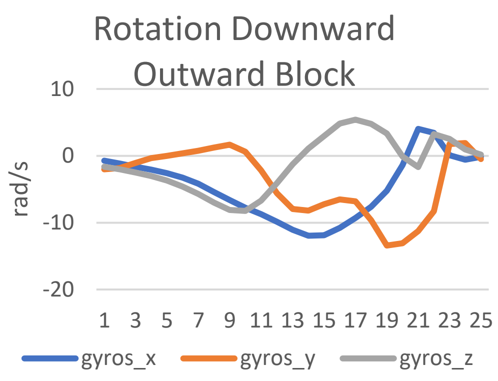

Figure 40. Rotation of a Downward Outward Block executed by an experienced student.

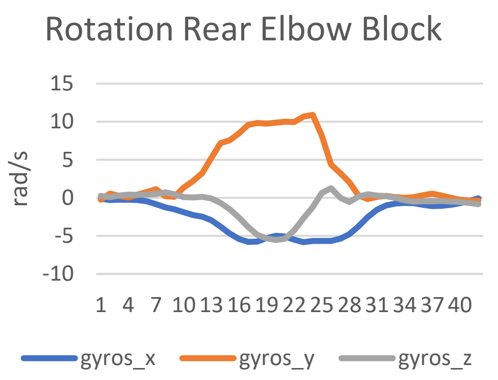

Figure 41. Rotation of a Rear Elbow Block executed by a beginner student.

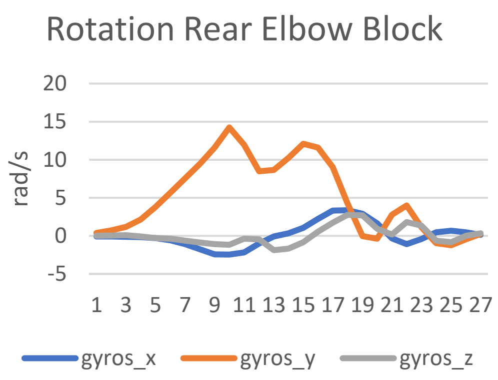

Figure 42. Rotation of a Rear Elbow Block executed by an experienced student.

Specifically, the movements at the left have been executed by a beginner that had only practiced the set for the motion capture process for this study, and the movements at the right have been executed by a person who has been practicing this set for years. It can be seen how an inward block (Figures 35 and 36), produces an increase in the y axis while maintaining the other two axes relatively static; or how the Upward block (Figures 33 and 34) produces an increase in the x axis, a decrease in the y axis, and maintains the z axis relatively static. A block that is more difficult to classify manually is the Downward Outward Block (Figures 39 and 40), since it involves a circular movement that affect the different axis, and the angle of the hand may change. Even so, it is still difficult to really recognize and compare movements manually. Thus, in this Master's Thesis we have explored if **AI techniques can help to address this problem and facilitate the identification of movements.**

3.4.2 Normalizing the Dataset

The data captured from different participants may be in different intervals, for example, if the acceleration of the arm of one participant is higher than the acceleration of another. This could lead to the system to overfit over those intervals and not generalize the learned characteristics, leading to fail when analyzing data from outside of the dataset. Normalization of each sequence could help to avoid this problem, establishing all the values between 0 and 1.

Equation 1. Normalization

$$X^{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where:

- X^{norm}: Sequence X normalized.
- X: Original raw sequence X.
- X_{min}: Minimum value in X.
- X_{max}: Maximum value in X.

The expression in Equation 1 is the one used for normalizing the data. The normalization of each sequence is done locally, e.g., the maximum and minimum values of Equation 1 are obtained from each sequence and not from the global dataset with the purpose of getting rid of differences of the interval between participants. An example of one of the sequences normalized and without normalizing can be seen in Figure 43. It can be seen in the y axis how the raw sequence is inside the interval (-21, 16), and the normalized sequence is inside the interval (0.30, 0.75).

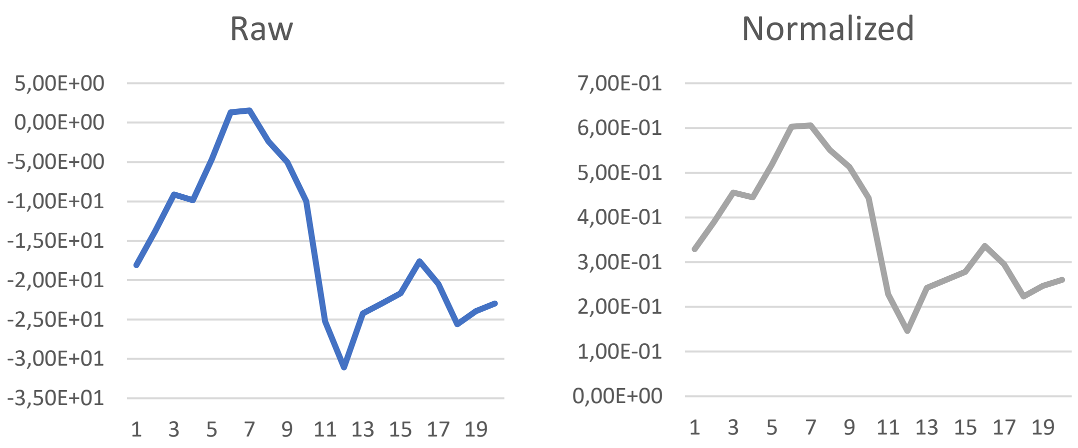

Figure 43. Comparation of normalized time series and raw time series.

3.4.3 Smoothing the Sequences

Once the movements have been isolated, labeled and normalized, the curves may be smoothened to delete noise. For this purpose, a variation of Moving Averages, known as Exponentially Weighted Moving Averages (EWMA) has been used (Equation 2).

Equation 2. Exponentially Weighted Moving Averages.

$$X_{t+1}^{smooth} = (X_t * \beta) + (1 - \beta) * X_t^{smooth}$$

Where:

- X_{t+1}^{smooth}: Value that will be calculated for a determined instant of time.
- X_t^{smooth}: Value calculated using EWMA in a previous instant of time.
- X_t: Value of the original sequence in the previous instant of time.
- β: Parameter of EWMA. Lower values give smoother curves. If β = 1, EWMA is not applied.

Since this function uses previously calculated values for calculating the following ones, it is necessary to have an initial value X_0^{smooth}. The value used in this implementation is X_0, e.g., the first value of the original sequence. As shown in Equation 2, EWMA has a parameter called beta (β) that can be adjusted for obtaining more or less smoothed curves. An example of the application of EWMA over a normalized curve using β=0.3 can be seen in Figure 44.

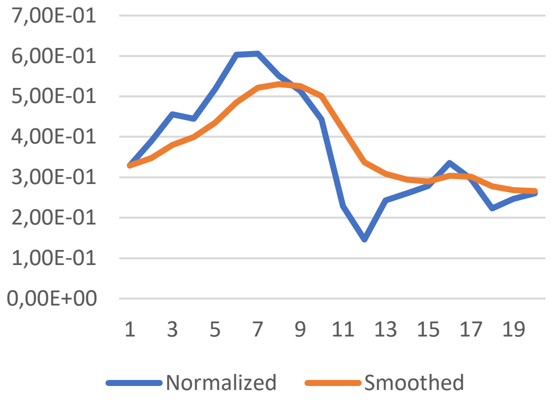

Figure 44. Example of smoothed curve using EWMA with a beta value of 0.3.

3.4.4 Padding the Sequences

The Neural Networks structures used in the analysis of the data requires the input sequences to have the same length, so the sequences need to be expanded or padded. This is an easy task that can be achieved by selecting the length of the longest series (which is 56) and extending the rest of the series by adding 0's at the end of the series to reach the maximum length. The value 0 has been selected since it is the one used by available libraries like Keras. This way, all records of the dataset can have the same length and can be properly used as input sequences for Neural Networks. In figure 45 an example of an extended series can be seen.

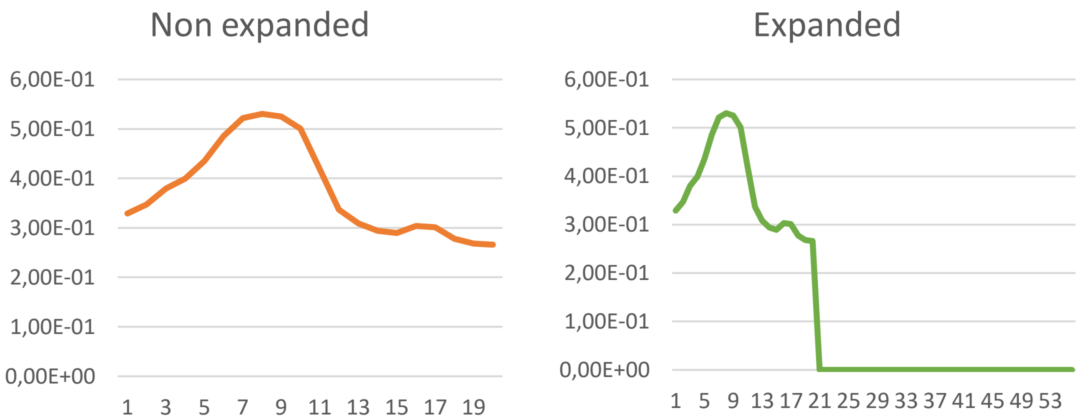

Figure 45. Example of expanded time series.

After padding the sequences, a matrix of size (240, 18, 56) is obtained. The first dimension of the matrix represents the number of elements in the dataset. The second dimension represents the number of characteristics measured, and the third dimension represents the length of each sequence.

3.4.5 Resulting Dataset

The resulting dataset is then formed by 240 examples, that can belong to 6 different classes. Each example is a vector of 18 characteristics, each one containing a sequence of 56 values, which can belong to one class, so the full dataset can be represented as a matrix of size (240, 18, 56, 1), and each example can be represented as a matrix of size (18, 56, 1).

For analyzing the data, it is useful to have the classes represented as one-hot vectors, which are vectors with a size equal to the number of classes (6). These vectors are formed by 0's, except by the position of the class, which contains a 1. So, if we have an example that belongs to the class 3, the one-hot vector that represents the class would be [0, 0, 0, 1, 0, 0]. The first index corresponds to static noise, the second index correspond to an Upward Block, the third index correspond to an Inward Block… and so on. This gives a dataset of size (240, 18, 56, 6).

Table 11. Size of elements in the dataset.

	Value
Nº Examples	240
Nº Characteristics	18
Length of each characteristic	56
Nº Classes	6
Size of full dataset	(240, 18, 56, 6)

It is convenient to separate each example from the classes when feeding a neural network with it. This gives a set of examples of size (240, 18, 56) and a set of classes of size (240, 6).

3.5 Analyzing the Dataset

Once the input data has been structured, labelled, normalized, smoothed and padded, the analysis can begin. As mentioned before, at this level, the research of the Master's Thesis has focused on comparing the analysis of the data using three different types of neural networks, implemented in Python using the Keras[31] API, that provides a simplified API of the Tensorflow[32] library. The different kind of neural networks compared in this research are the following:

- **Fully Connected Artificial Neural Networks (FC-ANN)** [113]**:** These are the regular neural networks, in which each neuron performs the dot product between the inputs (outputs of the previous layer) and a vector of weights that are learned by the network. Then, a bias value that is also learned by the network is added to the result of the dot product. An activation function may be applied as well, but it has not been applied in this implementation.
- **1D Convolutional Neural Networks (1D-CNN)** [113]**:** Each neuron of a CNN performs a convolutional operation over the input vectors, using a set of filters with a determined size (kernel size), generating a set of vectors as output. The output of a CNN consists in a set of characteristics extracted from the input data and usually is passed to an ANN for performing the classification or recognition task.

[31] Keras webpage: https://keras.io/
[32] Tensorflow webpage: https://www.tensorflow.org/

- **Long Short-Term Memory Networks (LSTM)** [113]**:** This is a type of Recursive Neural Network (RNN). Each neuron consists in a set of gates for controlling the information that enters and exits the cell, allowing to "remember" values while processing the sequence. Again, the output of the LSTM consists in a set of characteristics extracted from the input data that usually are passed to an ANN for performing the classification or recognition task.

The implementation of the three types of neural networks have been done in a Google Colab Notebook[33], and the code is also publicly available on GitHub[34].

The following two subsections provide an explanation of the layers used when explaining the network structures and developing it, and the training parameters used. Then, there are three subsections explaining the specific structure of each network (FC-ANN, 1D-CNN and LSTM), and finally, there is one subsection comparing the results of training the networks using the selected configurations of parameters, and selecting the best classifier obtained for the implementation of KSAS.

3.5.1 Layers in Keras

When developing the networks and explaining their structures, the terminology of Keras is used in this document. The different layers of Keras used in the development of the networks are explained in this section:

- **Input Layer:** This layer takes the input of the networks and gives it to the next layer.
- **Flatten Layer:** This layer takes a matrix as input and converts it into a one-dimensional vector.
- **Dropout Layer:** This layer can randomly deactivate inputs, scaling the values of the rest of inputs in a way that the sum of all inputs is not affected. This layer receives a parameter, the dropout value, that indicates the rate of deactivated neurons. The use of a dropout layer is a technique of regularization useful for avoiding overfitting.
- **Dense Layer:** This layer represents a layer of a FC-ANN.
- **Conv1D Layer**: This layer represents a layer of a 1D-CNN.
- **LSTM Layer:** This layer represents a layer of a LSTM network.
- **Activation Layer:** This layer performs an activation function over the inputs. The activation function used in the experiments performed is the SoftMax function, which is normally used for obtaining the final output of the network, giving a vector of values between 0 and 1, being the index of the highest value the predicted class.

Using the layers mentioned above, a model of the networks is created. Then, that model needs to be compiled for establishing how the model will be trained. The models are trained using an Adam optimizer over the loss function *logcosh*, with the goal of enhancing the accuracy of the model when classifying movements.

Once the model has been compiled, the model can start the optimization process, which will give as result a trained network. The optimization will be done by separating the dataset into two subsets randomly selected in each execution: 70% of the examples in the dataset conform the training set and the remaining 30 conform the test set. The training and test set have been divided randomly instead of using other available techniques such as k-folds, because it tends decreases the variance in the results. In k-folds, each training set appears at least once in the

[33] Google Colab webpage: https://colab.research.google.com/
[34] Martial Arts Movements Classifier source code: https://github.com/AlbertoCasasOrtiz/Martial-Arts-Movements-Classifier

training set, which seems fair. However, the results obtained, even when are less biased, have more variance. Selecting the training and test set randomly allows to reduce that variance in exchange of increase the bias. Anyway, in both methods, the same number of subsets can be tested and averaged for comparing the classifiers.

The training set is used for training the network and learn the weights necessary for classifying the data, and the test set is used for confirming that the learned information can be generalize over data that is not present in the training set. The batch size is 40, e.g., each training iteration will use 40 randomly selected examples from the training set, and it will be trained over 4000 epochs, e.g., there will be 4000 iterations.

3.5.2 Training Parameters

The three types of neural network have been trained using the same hyperparameters (Table 12). The only exception is the training of the LSTM, that has not been trained using a learning rate of 0.0001. This decision was taken after checking that the FC-ANN and the 1D-CNN gave better results using a learning rate of 0.0001, so that value was used with LSTM obtaining excellent results. A few executions using a learning rate of 0.00001 with LSTM were carried out, but due to the long training times of the LSTM, and the poor results obtained, were discarded.

Table 12. Parameters used for training the networks.

Parameter Name	Values	Explanation
Normalization	True, False	Application or not of normalization over each sequence of each example.
β (EWMA)	0.1, 0.3, 0.5, 0.7, 0.9, 1	Value β of EWMA that indicates how much does the curve will be smoothen. Smaller values give more smoothen curves, and a value of 1 gives the original curve.
Optimizer	Adam	A recently developed (2015) optimization algorithm used for training the model and reduce the loss function.
Learning rate (Adam)	0.00001, 0.0001	Proportion of weights update in each epoch.
β1 (Adam)	0.9	Exponential decay rate for the first moment estimates.
β2 (Adam)	0.999	Exponential decay rate for the second moment estimates.
Decay (Adam)	0.0001	Modulates how the learning rate changes over time.
Loss Function	logcosh	Logarithm of the hyperbolic cosine executed over prediction error. The optimization of this parameter for obtaining high accuracy predicting the movements is the goal of training the network.

3.5.3 Fully Connected Artificial Neural Network

A fully connected artificial neural network (FC-ANN) has been trained with the purpose of correctly classifying each movement using the parameters mentioned in Section 3.5.2. The structure of the network can be seen in Figures 46 and 47. Figure 47 is the summary of the network generated with Keras, and Figure 46 has been created by the researcher of this document for a better understanding of the network structure.

The network consists in an Input Layer that takes an example of the training set as an input matrix of size (18, 56) and gives it to a Flatten Layer that converts the input vector into a one-dimensional vector of size (1008). Flatten the matrix is necessary since the input of a Dense Layer needs to be a one-dimensional vector. The flattened vector is passed to a Dense Layer of 512 neurons giving a vector of size (512) as output. A dropout layer is then applied to the output of the dense layer with a dropout value of 0.5, giving a vector of the same size. Then, the output of the Dropout Layer is passed to a Dense Layer of 6 neurons, and an Activation Layer using the activation function SoftMax is applied for obtaining the output of the network. This layer has 519,686 parameters that must be learned.

The value 512 of the first Dense Layer has been selected because it is the power of 2 nearest to the half value of the output of the previous layer, 1008. The value 6 of the second Dense layer corresponds with the number of classes of the problem.

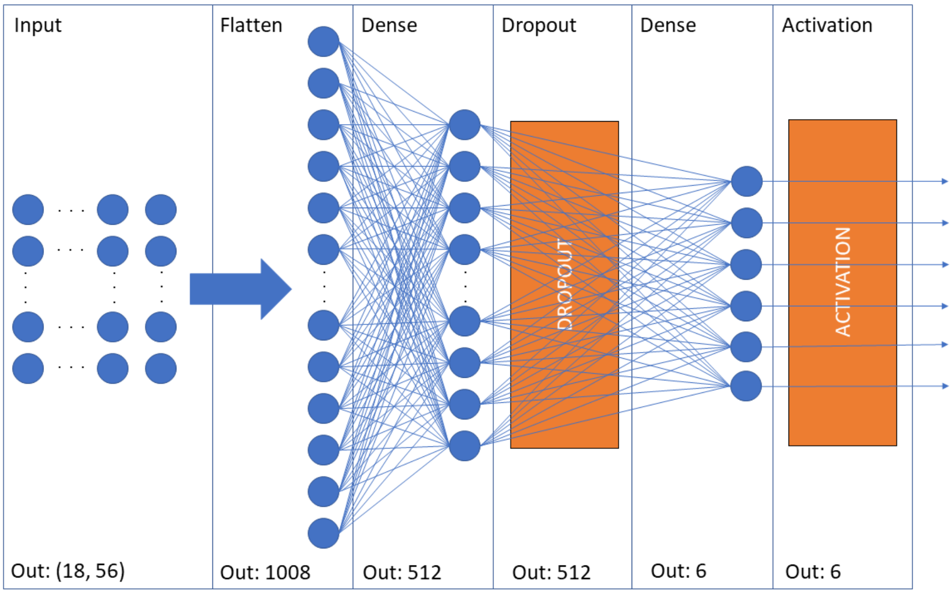

Figure 46. Structure of the FC-ANN.

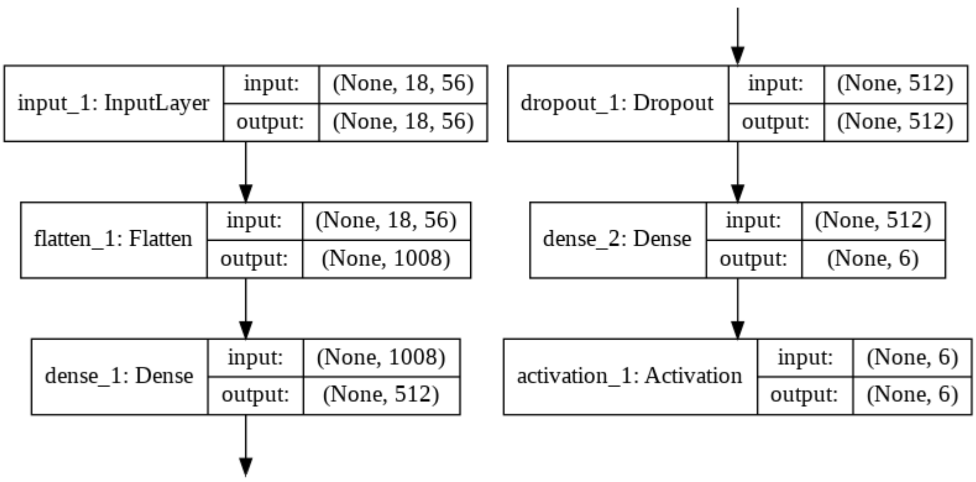

Figure 47. Model plot generated by Keras for the FC-ANN.

3.5.4 1-Dimensional Convolutional Neural Network

A 1-Dimensional Convolutional Neural Network (1D-CNN) has been trained using the parameters shown in Section 3.5.2. The structure of the network can be seen in Figures 48 and 49. Figure 49 is the summary of the network generated with Keras, and Figure 48 has been created by the researcher of this document for a better understanding of the network structure.

The network consists in an Input Layer that takes the example of the training set as an input matrix of size (18, 56). The input is directly given to a Conv1D layer formed by 32 kernels of size (5). Each of those kernels performs a convolutional operation over the input, giving as output a matrix of size (14, 32). Then, a Dropout Layer with a dropout value of 0.5 is applied, leaving a matrix of the same size (14, 32). The data output by the Dropout Layer is converted then into a 1-dimensional vector of size (448). This vector is then passed to a Dense Layer of 6 neurons, and an Activation Layer using the activation function SoftMax is applied for obtaining the output of the network. This layer has 11,686 parameters that must be learned.

The value 5 for the size of kernels have been selected for allowing to share information between neighbors' elements of the sequence, and the value 32 for the number of kernels have been selected because is the power of 2 nearer the half of 56 (the length of the sequences).

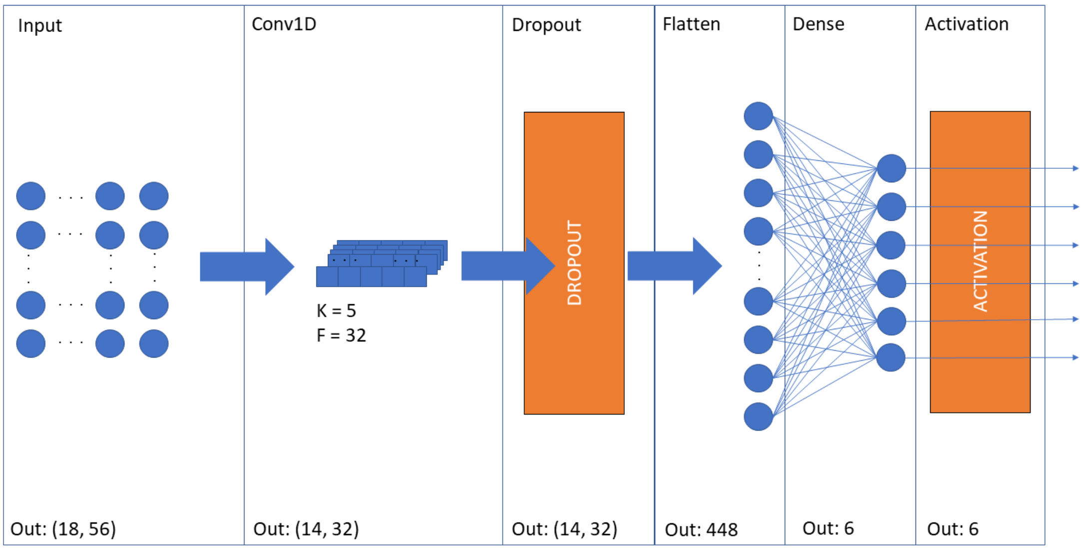

Figure 48. Structure of the 1D-CNN.

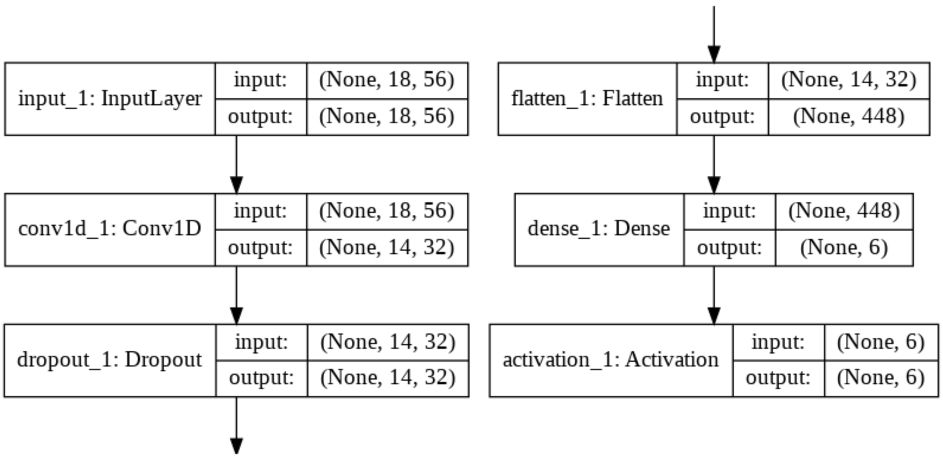

Figure 49. Model plot generated by Keras for the 1D-CNN.

3.5.5 LSTM Recursive Neural Network

A LSTM has been trained using the parameters shown in Section 3.5.2, except for the learning rate value 0.00001. The structure of the network can be seen in figures 50 and 51. Figure 51 is the summary of the network generated with Keras, and Figure 50 has been created by the researcher of this document for a better understanding of the network structure.

The network consists in an Input Layer that takes the example of the training set as an input matrix of size (18, 56). The input is given directly to a LSTM layer with 56 units. This layer receives the temporal sequences and processes them, giving as a result a matrix of the same size (18, 56). Then, a Dropout Layer with a dropout value of 0.5 is applied, leaving a matrix of the same size (18, 56). The data output by the Dropout Layer is converted then into a 1-dimensional vector of size (1008). This vector is then passed to a Dense Layer of 6 neurons, and an Activation Layer using the activation function SoftMax is applied for obtaining the output of the network. This layer has 31,366 parameters that must be learned.

The value 56 for the number of units has been chosen so the size of the vector does not change after being processed by the LSTM.

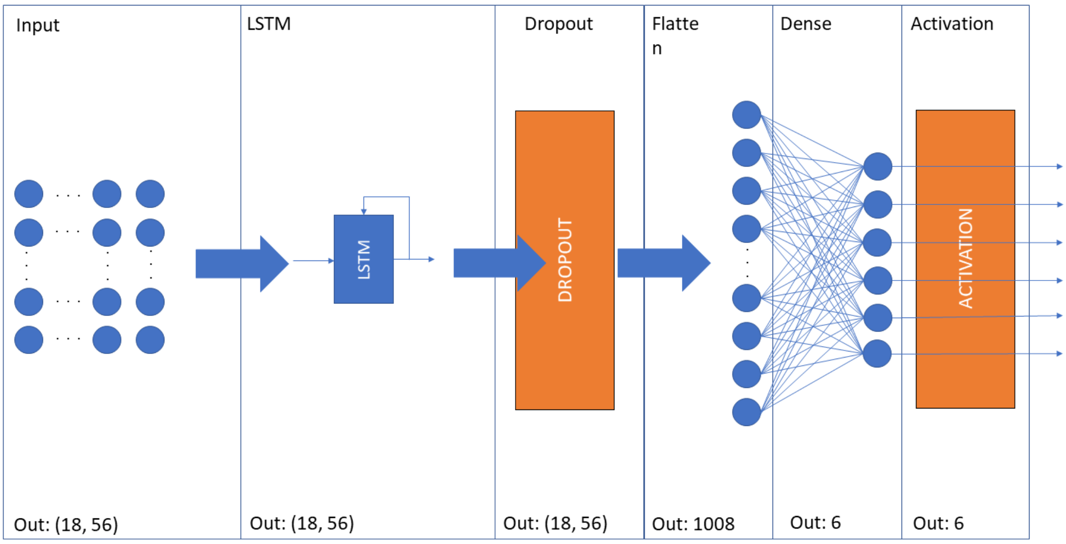

Figure 50. Structure of the LSTM Network.

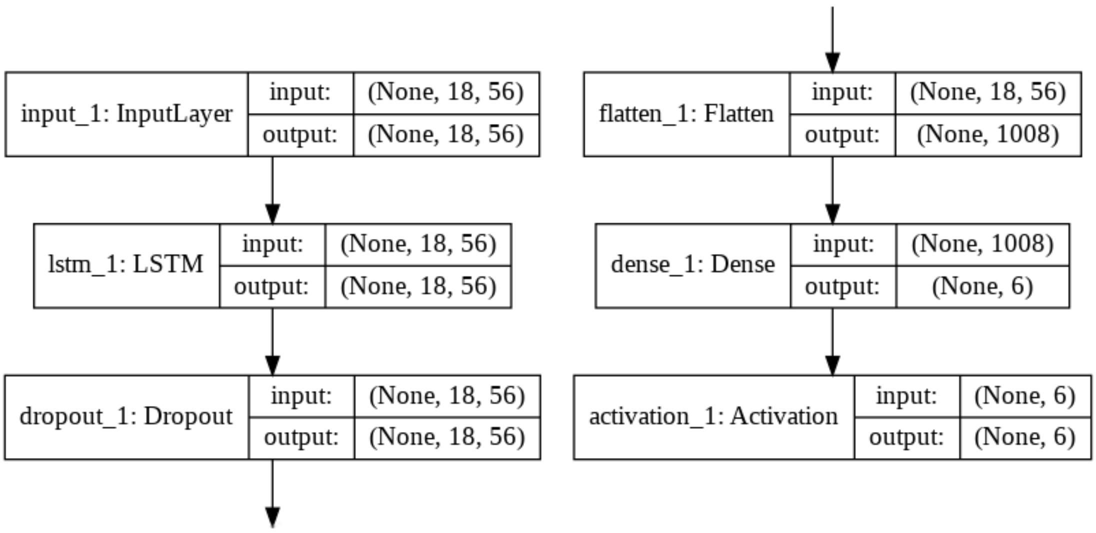

Figure 51. Model plot generated by Keras for the 1D-CNN.

3.5.6 Results After Training the Three Network Structures

For each configuration of the parameters (set of parameters), the training of the networks has been executed five times. This is because the results of a lonely execution could be affected by the random initialization of the weights, and the average of five executions could give a better insight of how the classifier behaves with a specific configuration. Then, a total of 300 executions has been carried out: 120 executions for ANN, 120 executions for CNN and 60 executions for RNN (as mentioned before, the value 0.00001 for the learning rate has not been tested with LSTM). The data of each execution have been downloaded manually from Google Colab and organized in an Excel file.

3.5.6.1 Results of FC-ANN

The data contained in each cell of Tables 13 and 14 is the average of five training executions of the FC-ANN with the given parameters. The training of the network has been executed 120 times, and it lasted 6 hours and 16 minutes, lasting 3 minutes and 6 seconds per execution.

Table 13. Average of five executions per each configuration of parameters for FC-ANN. lr = 0.00001.

FC-ANN								
Mean	**lr = 0.00001**							
	Normalization				**No Normalization**			
EWMA b	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**
0.1	0.95	0.76	0.90	171	0.95	0.88	0.93	171
0.3	0.95	0.78	0.90	190	0.96	0.88	0.94	188
0.5	0.97	0.78	0.91	204	0.96	0.85	0.93	184
0.7	0.97	0.78	0.91	206	0.96	0.84	0.93	189
0.9	0.97	0.78	0.91	205	0.96	0.86	0.93	184
1	0.97	0.79	0.92	205	0.96	0.85	0.93	187

Table 14. Average of five executions per each configuration of parameters for FC-ANN. lr = 0.0001.

FC-ANN								
Mean	**lr = 0.0001**							
	Normalization				**No Normalization**			
EWMA b	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**
0.1	0.97	0.75	0.91	185	0.97	0.84	0.93	186
0.3	0.99	0.76	0.92	190	0.97	0.84	0.93	189
0.5	0.99	0.76	0.92	187	0.97	0.84	0.93	187
0.7	0.99	0.76	0.92	188	0.97	0.85	0.93	185
0.9	0.99	0.75	0.92	187	0.97	0.84	0.93	187
1	0.99	0.75	0.92	178	0.97	0.85	0.93	184

There are subtle differences between the executions, but in Figures 52 and 53 it can be seen how the executions using normalization have an accuracy on the training set slightly higher than those that do not use normalization, and the accuracy over the test set is higher in the executions using normalization. Figures 54 and 55 show hot the accuracy over the full dataset tend to be higher too when using normalization. The differences in accuracy when applying different values of β (EWMA) are subtle too, but it can be seen how using a learning rate of 0.0001 and a β of 0.1 generates worse results when using normalization than when using other values for β. When the learning rates is 0.00001 the accuracy using normalization increases slightly as the value of β increases. It seems that the value of β does not affect the accuracy when normalization is not used.

It can be seen in Figures 52 and 53 how the intervals between the accuracy over the training and the test set is higher when using normalization than when not. It could be that it is more difficult to generalize the learning when using normalization. When normalization was applied, using a learning rate of 0.0001 gave better results, but without normalization, both values, 0.0001 and 0.00001 gave similar results.

The best model obtained using FC-ANN has 0.97 of accuracy over the training set, 0.90 of accuracy over the test set and 0.95 of accuracy over the full dataset. This model was trained with a learning rate of 0.00001 and a value of β of 0.03 without normalization. It can be seen in Figure 55 how there is a maximum value in accuracy over the full dataset using these parameters.

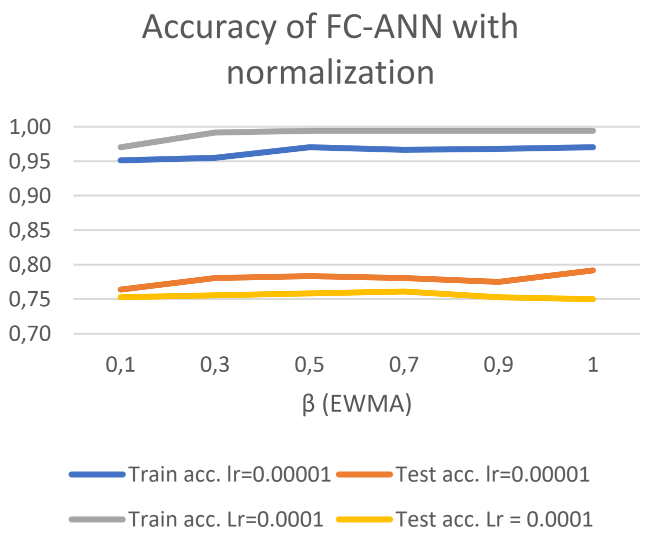

Figure 52. Accuracy of FC-ANN with normalization over training and test sets.

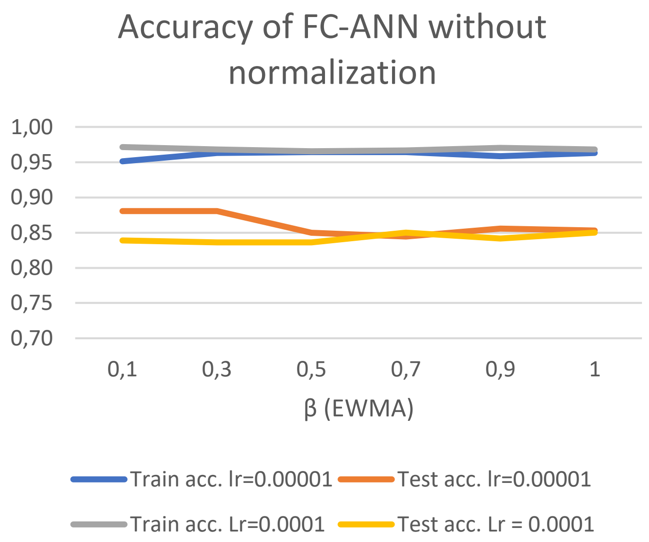

Figure 53. Accuracy of FC-ANN without normalization over training and test sets.

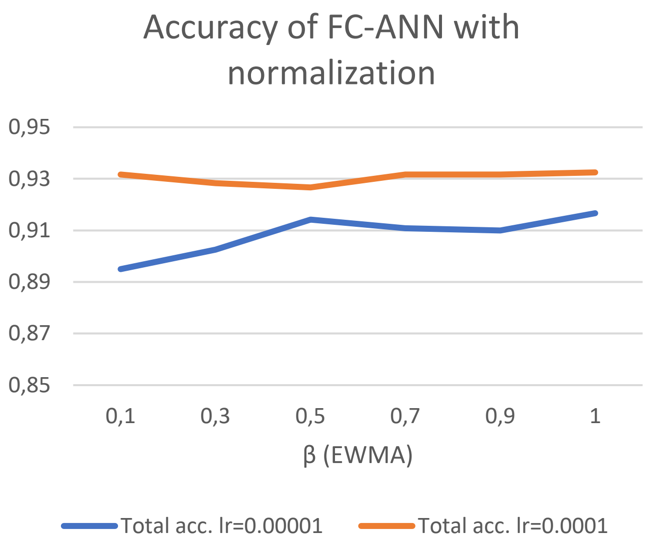

Figure 54. Accuracy of FC-ANN with normalization over the full datasets.

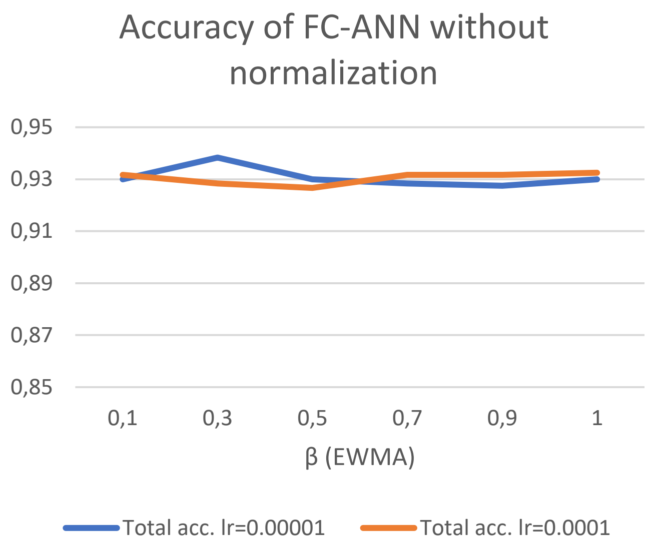

Figure 55. Accuracy of FC-ANN without normalization over the full dataset.

3.5.6.2 Results of 1D-CNN

The data contained in each cell of Tables 15 and 16 is the average of five training executions of the 1D-CNN with the given parameters. The training of the network has been executed 120 times, and it lasted 2 hours and 19 minutes, with an average of 35 seconds per execution.

In Figures 56, 57, 58 and 59 it can be seen how when using 1D-CNN, the accuracy is always better with a learning rate of 0.0001. Again, it can be seen in figures 56 and 58 how the accuracy tends to slightly decrease when the value of β is low and normalization is activated. Figures 57 and 59 shows how a value 0.3 of β seems to give better results again when normalization is deactivated.

Table 15. Average of five executions per each configuration of parameters for 1D-CNN. lr 0 0.00001

1D-CNN								
Mean	**lr = 0.00001**							
	Normalization				**No Normalization**			
EWMA b	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**
0.1	0.65	0.48	0.60	72	0.81	0.79	0.81	70
0.3	0.70	0.61	0.67	70	0.87	0.87	0.87	74
0.5	0.71	0.63	0.69	74	0.83	0.82	0.82	77
0.7	0.72	0.58	0.68	73	0.83	0.86	0.84	72
0.9	0.73	0.58	0.68	71	0.85	0.82	0.84	72
1	0.71	0.61	0.68	69	0.83	0.82	0.82	70

Table 16. Average of five executions per each configuration of parameters for 1D-CNN. lr = 0.0001

1D-CNN								
Mean	lr = 0.0001							
	Normalization				No Normalization			
EWMA b	Train acc.	Test acc.	Total acc.	Time (s)	Train acc.	Test acc.	Total acc.	Time (s)
0.1	0.96	0.81	0.91	72	0.98	0.91	0.96	74
0.3	0.97	0.88	0.95	66	0.98	0.89	0.96	67
0.5	0.98	0.86	0.94	67	0.98	0.89	0.96	67
0.7	0.98	0.85	0.94	66	0.98	0.91	0.96	66
0.9	0.98	0.82	0.94	67	0.98	0.90	0.96	68
1	0.99	0.86	0.95	67	0.98	0.90	0.96	67

Using this network structure seems to have a better effect when normalization is not activated. In Figures 56, 57, 58 and 59 it can be seen how for a learning rate of 0.0001, the accuracy is similar when using and not normalization, but when using a learning rate of 0.0001, the accuracy is enhanced.

The best model obtained using 1D-CNN has 0.98 of accuracy over the training set, 0.94 of accuracy over the test set and 0.97 of accuracy over the full dataset. This model was trained with a learning rate of 0.0001 and a value of β of 0.07 without normalization. There are two other configurations using a learning rate of 0.0001 that gave an accuracy over the full dataset of 0.97, but the accuracy over the training set was 0.99 and the accuracy over the test set was 0.92 in both. Since the interval between 0.99 and 0.92 is higher than the interval between 0.98 and 0.94, the second interval was selected because it looks that the high accuracy over the full dataset is due to a best generalization of the learned parameters since the accuracy over the test set is higher.

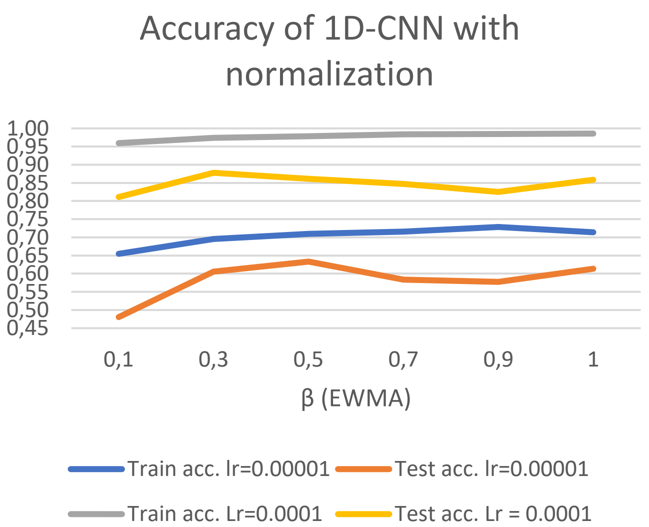

Figure 56. Accuracy of FC-ANN with normalization over training and test sets.

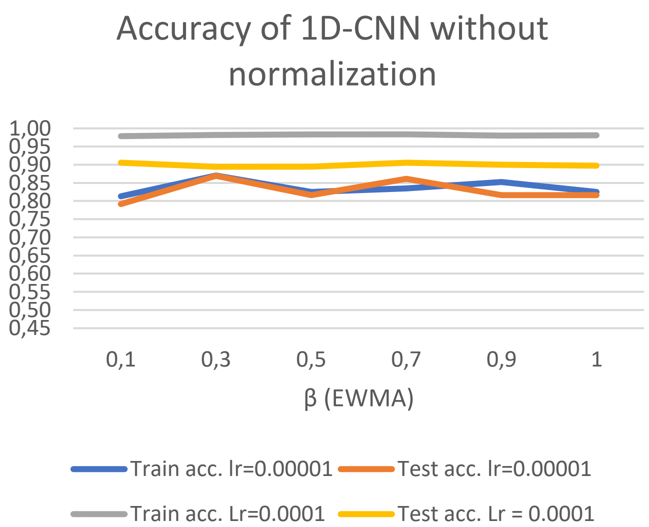

Figure 57. Accuracy of FC-ANN without normalization over training and test sets.

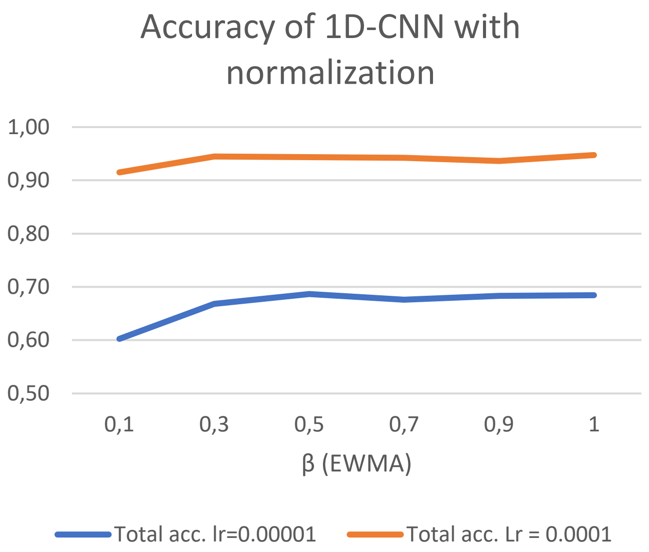

Figure 58. Accuracy of FC-ANN with normalization over the full datasets.

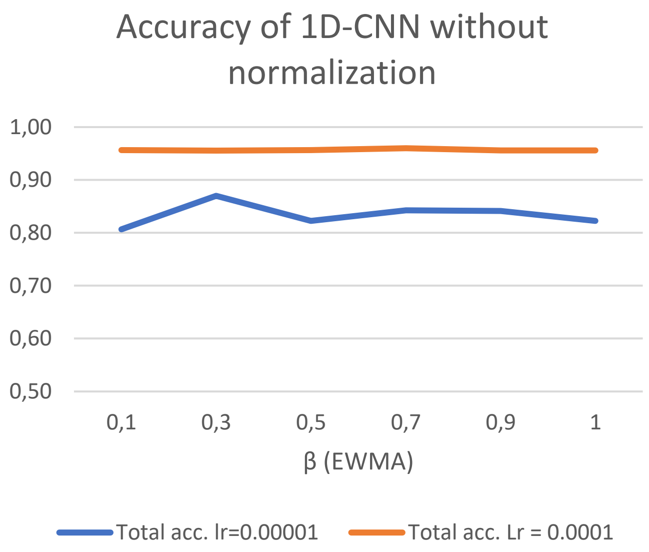

Figure 59. Accuracy of FC-ANN without normalization over the full dataset.

3.5.6.3 Results of LSTM

The data contained in each cell of Table 17 is the average of five training executions of the 1D-CNN with the given parameters. The training of the network has been executed 60 times, and it lasted 4 hours and 24 minutes, with an average of 4 minutes and 25 seconds per execution.

Since this is the network structure that employs more time in training, a learning rate of 0.0001 was applied before a learning rate of 0.00001 since it was the value that gave the best results in the other two network structures. Since the results of using a learning rate of 0.0001 are difficult to surpass, the network structure will not be trained using a learning rate of 0.00001.

Table 17. Average of five executions per each configuration of parameters for LSTM.

LSTM								
Mean	**lr = 0.0001**							
	Normalization				**No Normalization**			
EWMA b	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**	**Train acc.**	**Test acc.**	**Total acc.**	**Time (s)**
0.1	0.96	0.41	0.80	263.60	1.00	0.89	0.97	266.00
0.3	0.96	0.43	0.80	266.80	1.00	0.92	0.98	267.20
0.5	0.97	0.49	0.82	255.40	1.00	0.92	0.98	264.40
0.7	0.97	0.48	0.83	257.80	1.00	0.90	0.97	259.40
0.9	0.97	0.44	0.81	262.60	1.00	0.89	0.97	257.00
1	0.98	0.49	0.83	267.80	1.00	0.91	0.97	287.00

In figures 60, 61, 62 and 63, it can be seen how the accuracy of the executions that does not use normalization is always better. Again, when using normalization, it looks that the accuracy over the test set is better as the value of β increases. The accuracy over the test set when normalization is activated is the worst obtained from the three models, but the accuracy over the training set is high. This indicates that when normalization is activated the network is adjusting well to the training examples, but it does not generalize over examples that are not in the training set. When normalization is not activated, the accuracy over the training set is always 1 (it correctly classifies all examples in the training set), and the accuracy over the test set is high, always over 0.89. The accuracy over the full dataset, when normalization is activated, is always over 0.96.

The best model obtained using LSTM has 1.0 of accuracy over the training set, 0.94 of accuracy over the test set, and 0.98 of accuracy over the full dataset. These values have been achieved using a learning rate of 0.0001 and no normalization, with the value of β 0.5.

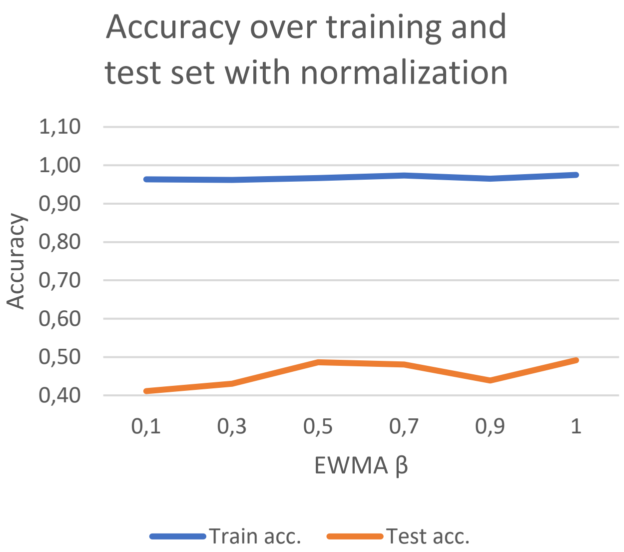

Figure 60. Accuracy of FC-ANN with normalization over training and test sets.

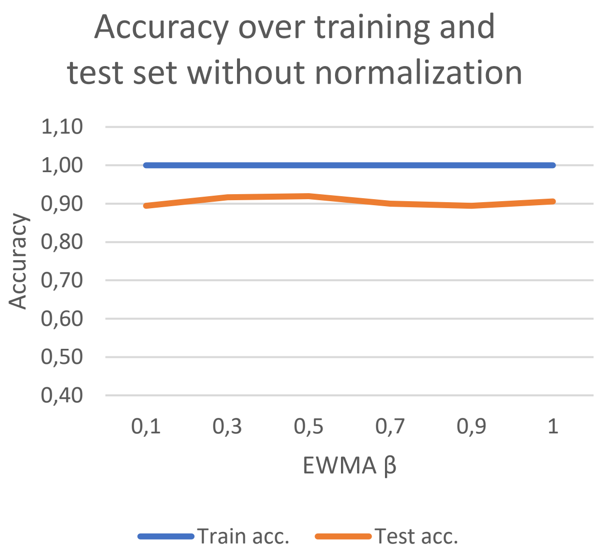

Figure 61. Accuracy of FC-ANN without normalization over training and test sets.

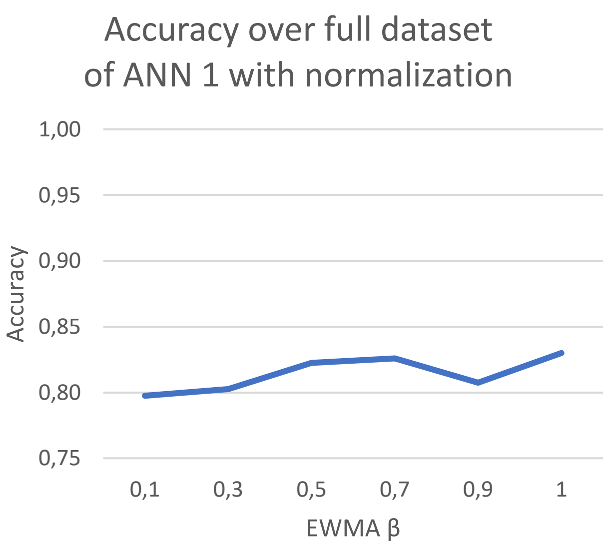

Figure 62. Accuracy of FC-ANN with normalization over training and test sets.

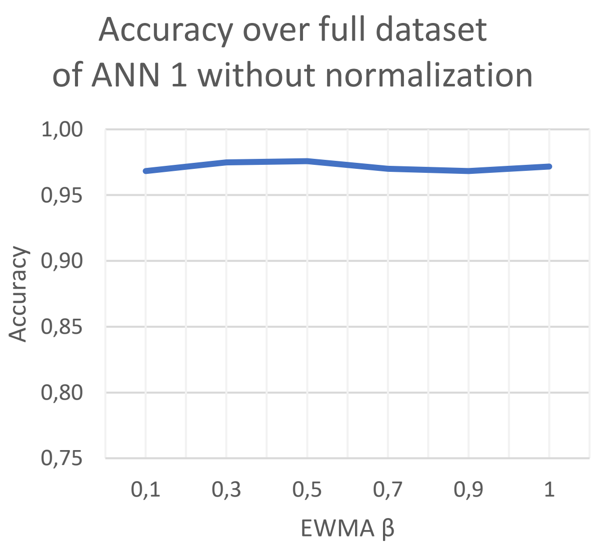

Figure 63. Accuracy of FC-ANN without normalization over training and test sets.

2.5.6.4 Comparison Between Classifiers

The three classifiers have obtained good results with normalization deactivated. In the Figures and Tables of the previous sections, it can be seen how FC-ANN obtained good results in all the executions, getting always accuracies of over 0.75 over the testing set, accuracies over 0.95 over the training set, and accuracies over 0.90 over the full dataset. This is different in 1D-CNN, which obtained poor results with a learning rate of 0.00001 and normalization, however, in the rest of the cases obtained always accuracies over 0.8 over the training set, accuracies over 0.79 over the test set, and accuracies over 0.81 over the full dataset.

Both, FC-ANN and 1D-CNN, have obtained better results when normalization is not activated and when the learning rate was 0.0001.

LSTM as well have obtained worse results when normalization was activated, but when it was deactivated, this network structure have obtained the best results, with an accuracy of 1.0 over the training set, an accuracy over 0.89 on the test set, and an accuracy over 0.97 over the full dataset.

1D-CNN was the classifier that took less time of training per each execution (35s), and LSTM the one which took more (4min 25s), closely followed by FC-ANN (3min 6s). The model which had to train more parameters was FC-CNN (519,686), which contrast with the relatively low number of parameters trained by LSTM (31,366) and 1D-CNN (11,686). Even with LSTM giving results slightly better than 1D-CNN, if a similar problem requires to train a new model, 1D-CNN would be the best option, since the lower training times compensate the slight difference in accuracy (Only 0.1 of difference in accuracy over the full dataset as shown in Table 18).

In Table 18, it can be seen the best results obtained by each network structure. The winner of the three classifiers is LSTM and this is the one that will be used in KSAS. This was the expected result since LSTM are the type of neural network most suitable for analyzing time series.

Table 18. Best accuracy obtained by each classifier.

Classifier	Train acc.	Test acc.	Total acc.
FC-ANN	0.97	0.90	0.95
1D-CNN	0.98	0.94	0.97
LSTM	1.00	0.94	0.98

3.6 Giving Feedback to the User

One of the disadvantages about online learning of a psychomotor activity most mentioned in the questionnaire of Appendix IV was the lack of feedback received by the students. This section describes the feedback strategy proposed in KSAS.

Since the device used is a smartphone, the three modalities of feedback visual, haptic and auditory can be given to the user. This is so because smartphones have speakers, screens and vibrators embedded.

The feedback strategy used in KSAS is very simple. The feedback is given by during the session. When the user executes a good movement, auditory feedback is given by congratulating verbally the user and playing a victory sound. When the user executes a wrong movement, haptic feedback is given by emitting a vibration of 500ms, and auditory feedback is given by correcting the user verbally and playing a buzz sound. The words for congratulating the user have been selected in a way that the app sounds motivating and polite, so the users do not feel frustrated when a movement is executed wrongly. For example, when a movement is correctly executed, the application congratulates him/her by saying phrases like "Great", "Perfect" or "Good Movement". However, when a movement is wrongly executed, the app just say that it has been wrongly executed and immediately proceed to as the user to repeat it, so the user does not feel that the app is complaining about him/her.

Visual feedback is not given during the session because the device is attached to the arm, and the screen is not always visible to the user. However, video examples of the movements are given before the execution and can be watched after the execution as well. This takes advantage of one of the advantages identified in the questionnaire of Appendix IV, which consist in that the use of videos, that the users can rewind, pause or change their speed, is useful for online learning.

3.7 AI Infrastructure of KSAS

Once the data has been captured, modeled, and a classifier has been trained, it is time to implement all those ideas in KSAS. The flow of the application is shown in Figure 64, which was already detailed in Section 2.1.

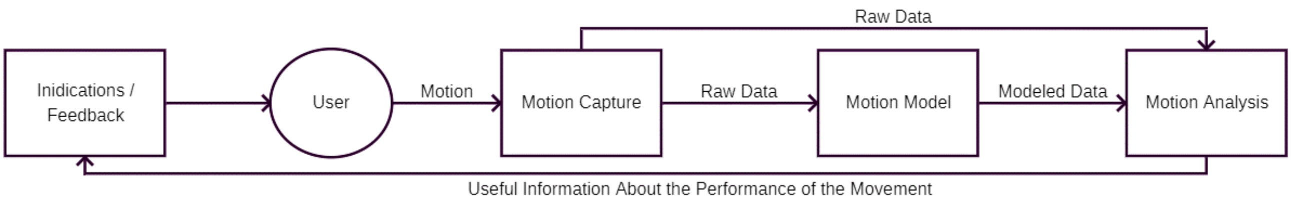

Figure 64. Flow of KSAS.

3.7.1 Capturing the Movements

When the application asks the user to execute a movement, it is needed to capture it. The code for capture the movements have already been developed in the app Motion Capture, described in Section 3.3.1, so it was ported to this application.

In KSAS, there is no need to store the captured data into a CSV file, so that part of the code was suppressed and now the movement is stored inside a matrix with the size (1, 18, n), where n is the length of the sequences. The matrix has those dimensions since the classifier requires the input to have the size (1, 18, 56), being 56 the longest length of sequence in the current dataset.

When the application asks the user to perform a movement, it starts to capture motion and waits three seconds before stopping. This gives a matrix of size (1, 18, n), with $n > 56$. An algorithm has been developed for looking for the consecutive sequence with length = 56 that maximizes the sum of absolute values of the values captured by the gyroscope. This is so because, as aforementioned, the values captured by the gyroscope are only different from of 0 in any axes when a movement is executed. After selecting the sequence, a matrix of size (1, 18, 56) is generated.

3.7.2 Modelling the Movements

The selected classifier requires the movements to be isolated, smoothed using EWMA, and expanded. Since now the data is being captured by the device instead of being part of a dataset, those tasks are easy to perform, and some of them are already done:

- **Isolate movements:** The movements are captured in isolation, so there is no need to isolate them.

- **Smooth the curves:** The EWMA algorithm was already developed using python, so the algorithm can be easy rewritten in Java.
- **Expand the movements:** The movements are already captured with the correct length of sequences.

Since the selected classifier was trained wit normalization deactivated, there is no need to normalize the movement.

3.7.3 Analyzing the Movements

The selected classifier was exported and converted into a Tensorflow Lite[35] model. This allows to import and use the classifier in an Android device. Then, the only needed step is to pass the input matrix of size (1, 18, 56) with the smoothen sequences to the classifier, which will return a one hot vector. The index of the highest value of the vector will correspond with the inferred class. The possible classes are: 1 – No movement detected, 2 – Upward Block, 3 – Inward Block, 4 – Outward Extended block, 5 – Downward Outward Block, and 6 – Rear Elbow Block.

3.7.4 Giving Feedback/Indications to the User

Once the movement has been classified, if the detected movement corresponds with the movement that the user had to execute, a victory sound will be played, and the app will verbally congratulate the user and ask him/her to execute the next movement. If the detected movement does not correspond with the movement that the user had to execute, a buzz sound will be played, the device will vibrate for 500ms, and the app will tell the user that the executed movement was the wrong one, and the app will ask the user to execute the movement again. If no movement is detected, the app will tell the user to execute the next movement.

As it can be seen, the feedback is mainly auditory, with a haptic component to indicate that the wrong movement has been executed. Visual feedback has not been included because the screen of the device is not accessible during most of the execution of the set.

Auditory feedback has been included as well when a sound is played when the application starts to capture the motion.

3.7.5 Activities of the Application

In Android, an activity represents a screen of the application. The application consists of four android activities that are explained below:

- **Activity 1 – Welcome to KSAS (Figure 65):** This activity welcomes the user and asks him/her to use the device as a wearable.
- **Activity 2 – Select a Set (Figure 66):** This activity asks the user to select a set to practice. A preview of the sets is shown in video, so the student can identify better which one he/she wants to select. Currently, there is only one available option, Blocking Set I.
- **Activity 3 – Blocking Set I (Figure 67):** The application shows a brief introduction to the set and asks the user to view a video showing the full set if he/she wants to review it or to start the practice of the set.
- **Activity 4 – Set Practice (Figure 68):** The application asks the user to get into the initial position and push start. When the user push start, the application asks him to execute a movement.
    - If the movement executed is the correct one, a victory sound is played, the app congratulates the user and asks him/her to execute the next movement.

[35] Tensorflow Lite webpage: https://www.tensorflow.org/lite

- If the movement has been wrongly executed or have executed the wrong movement, a buzz sound is played, the app tells the user that the movement is wrong and asks he/she to return to the previous position, then, the app asks the user to perform the movement again.
- If no movements have been detected, the app asks the user to execute the movement again.
- Once the user has completed the set, the app congratulates the user and tells him/her how many times he/she has committed an error. Then, the user can push start to practice the set again or go back and practice another set.

Figure 65. First activity of the app.

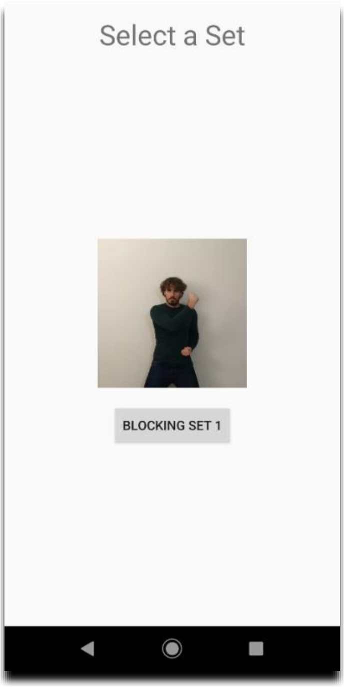

Figure 66. Second activity of the app.

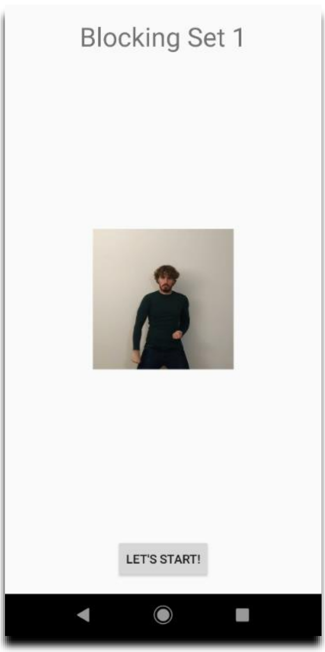

Figure 67. Third activity of the app.

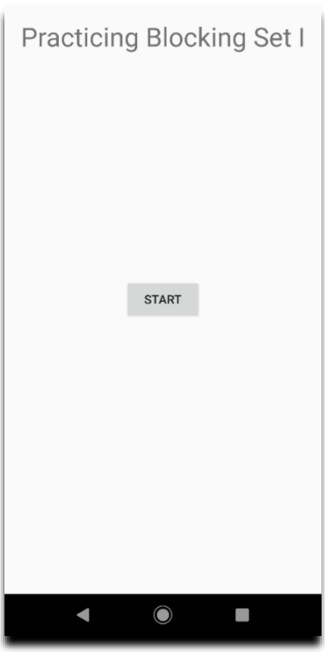

Figure 68. Fourth activity of the app.

4. DISCUSSION

This project has been organized in four phases that, as shown in Section 2.1, consist in: i) capture motion, ii) model motion, iii) analyze motion and iv) giving feedback to the user. With the purpose of researching if AI could assist in the learning of a psychomotor activity an AI-based application, that includes the selected technologies from the review of the state of the art, has been developed passing by the four mentioned phases.

In the following sections the results regarding the four phases and the development of the application are discussed, as well as the testing the application. The results of the phase of giving feedback are discussed after the results of testing the application since the users that participated in the testing provided some observations about it.

4.1 Capturing the Movements

In Section 3.3 the methods used for capturing the data were shown. The data of 20 voluntary participants was captured by means of the inertial sensors included in an Android device located in the forearm of the participants. A set of CSV files consisting in 18 characteristics extracted from the execution of five movements that are part of the Blocking Set I (Section 3.1.1), a set from Kenpo Karate formed by blocking movements, was created. Each characteristic is a time series of the captured signal. Each participant executed the set with both arms, so 40 executions of the set were captured. Static noise of the sensors when no movements were executed was captured as well from each arm, so information about what is not a movement was obtained. The capture of motion gave then 40 CSV files containing 18 characteristics extracted from the participants.

4.2 Modelling the Movements

The data captured has been modeled and prepared for its analysis. Each participant executed a total of five movements, and static noise was captured. Each movement was labelled and isolated from the others manually, obtaining a total of 240 movements (40 executions x 6 classes), classified in six classes (5 movements + static noise).

With the purpose of preparing the dataset for the analysis phases, which uses different types of neural networks, the data was normalized, smoothed and padded.

The normalization of the data was done locally, per each sequence. This way, the differences in the interval of values that different executions could produce are eliminated. Later in the analysis phase, the use of this kind of normalization produced generally worse results than when it was not applied, as it can be seen in Section 3.5.6.

The data has been smoothed by using the algorithm known as Exponential Weighted Moving Averages (EWMA). This algorithm allows to obtain smoother curves by adjusting a value called β. The analysis of the data while tuning this parameter shows that generally intermediate values o β (0.3-0.7) generates better results when normalization is not applied, but when normalization is applied, the results are better as the value of β increases as it can be seen in Section 3.5.6. Anyway, change the value of β provoke only smooth changes in the results.

After normalization and EWMA were applied, the data has been padded till every sequence in the dataset had a length of 56 (the maximum length of a sequence in the dataset). This is a prerequisite of the data when using it for feeding a neural network.

4.3 Analyzing the Movements

Once the dataset was created and the data has been modeled, three classifiers consisting in the use of three different kind of neural networks have been trained and compared. Those classifiers are: Fully Connected Artificial Neural Network (FC-ANN), 1-Dimmensional Convolutional Neural Network (1D-CNN), and a LSTM Neural Network (LSTM).

Those three classifiers have been trained using different configurations of hyperparameters (Section 3.5.2), of which the learning rate, the value β of EWMA and the use or not of normalization have been tuned. Per each configuration of parameters, the network have been trained five times and the averages of the five configurations can be seen in Section 3.5.6.

The results show that the three classifiers can obtain a good accuracy under the right configuration of parameters, however, the classifier that obtained the best results was LSTM, since it obtained accuracies of 1.0 over the training set and accuracies over 0.89 over the test set in every execution while normalization was deactivated and the learning rate was 0.0001. Similar results were obtained when using 1D-CNN, which obtained accuracies over 0.98 over the training set, and accuracies over 0.89 over the test set on the same conditions. Both classifiers behave worse when normalization was activated and when the learning rate was 0.00001. The FC-ANN obtained best accuracies with a learning rate of 0.00001 and normalization activated than LSTM and 1D-CNN but behaved slightly worse when the learning rate was 0.0001 and normalization was activated. A more detailed explanation of these results can be found in Section 3.5.6.

As expected, the best classifier was obtained by using LSTM with a learning rate of 0.0001, normalization deactivated and a value for β of 0.5. The accuracy over the training set was 1.0 and the accuracy over the test set was 0.94. This was then the classifier used in KSAS. It was expected that the LSTM was the best classifier since it has been designed for analyzing sequences.

4.4 KSAS (Kenpo Set Assisting System) and Feedback Strategy

An application for Android devices (i.e., a smartphone) has been created using the classifier selected after modeling and analyzing the data. When the application starts, asks the user to put the device as a wearable in the forearm, and then, asks him/her to select a set. A video of an execution of the set is shown, and then the user can proceed with practicing the set.

The application then asks the user to get into the starting position and asks him/her to execute the movements in sequence. If the user executes the right movement, a victory sound is played and the app congratulates the user, asking him/her to execute the next one. If the user executes a wrong movement, a buzz sound is played, the device vibrates, and the app informs the user that the executed movement is not the correct one, then asks the user to execute the movement again. If no movement is detected, the app asks the user to execute the movement again.

The development of this app is the result of applying the ideas and results from the different phases of motion capturing, analyzing, and modelling human motion, and a feedback strategy has been created. This application is able to capture motion data from the arm of a user by using the sensors of an Android device. Once the data is captured, the application models and prepares it for feeding the classifier. This is done by applying EWMA for smoothing the curves. Since the normalization was not applied in the selected classifier, it was not implemented in KSAS. The padding is not applied either since the captured sequence always have the correct length. Once the captured data has been modeled, a classifier is used to classify and recognize the movement, and feedback is given to the user depending of if the movement has been executed properly or if the wrong movement has been executed.

The feedback strategy used consist then in use auditory feedback for indicating when the user has executed the movement properly or not (The sound that is played), and for telling the user which movement should execute (When the app talks to the user). Haptic feedback is used as well for indicating when a movement has been wrongly executed (The device vibrates). Visual feedback has not been used since the screen of the device is not always used. However, the user can watch the execution of the set before proceeding with the practice.

4.5 Testing the Application

The application has been tested by one user, beginner in American Kenpo Karate. This user has been practicing American Kenpo Karate a month ago, and had learned the Blocking Set I. However, the user has not practiced the set since then, and the sequence of movement was not fresh.

The user was asked to perform some warming exercises and to start the application. No more indications from the researcher were given. The user put the device as wearable in her arm as the app requested and clicked next. Then the user selected Blocking Set I and watched a video of the full execution of the set.

The user then reached the four activity of the app, which asked her to get into the starting position and click start. Then the user started the execution of the movements, guided by the app. The user only failed once because she executed a wrong movement, but she could continue with the set after the app asked her to repeat the movement. After the execution finished, the user was asked to test the app again, with the other arm and in front of a mirror. This time, the user executed all the movements correctly.

After testing the app, the user was asked by the researcher to tell him what her thoughts about the app were. The user answered that using the device as wearable was not comfortable, and that it makes some movements more difficult to execute. The user emphasized that the execution of the movements in front of a mirror helped her by allowing her to observe better her own movements. Regarding the feedback strategy, the user complained about the app talking too much, and the annoying voice of the app. She said that it distracts her. The user also said that the use of a buzz sound with the vibration was a good idea for indicating when a movement was wrongly executed.

As an expert in the set, I also tested the app and found out that the second movement, the Inward Block, was recognized correctly when the beginner executed it, but not when I did it. After executing the movement, a great number of times, and observing the user execute it, I remembered that beginners tend to move the arm in a different manner while executing the movement, since they finish the first movement in a different angle. Since the application has been trained using mainly movements captured from persons who had never practice American Kenpo Karate, it is possible that the classifier has learned the movement wrongly. This highlights the importance of using a dataset conformed equally by data collected by experts and beginners.

An important fact that happened while testing the app was that the high number of times that I had to execute the movements caused tendinitis in my elbow. In future versions, the device could be chanted to a smaller one for avoiding annoyances, like the AX3 used in [8], and the app should be able to ask the user to stop practicing, or even to refuse to work, if the user has practiced the set many times. No more people has tested the app because of the quarantine state caused by the outbreak of COVID-19.

5 CONCLUSIONS

In this document, the possibilities of AI techniques to assist in the learning of martial arts have been researched. The hypothesis addressed is that it is possible to recognize, model, and analyze human motion, as well as give feedback to the user through the development of an AI-based application that executes in an Android device. The hypothesis has been validated by means of studying different methods for capturing, modeling and analyzing motion, as well as the creation of a feedback strategy, and by development of an Android able to assist in the learning of a set of movements from American Kenpo Karate known as Blocking Set I.

20 voluntaries participated in the motion capture phase. An Android device was placed in the forearm of each participant and the movements were captured from both arms. The characteristics recorded consist in sequences of signals obtained by the sensors of the Android device (Acceleration, rotation, magnetic field…).

Once the data has been captured, each movement has been locally normalized, smoothed using Exponentially Moving Averages (EWMA), and padded in a way so each sequence has the same length. This was done with the purpose of better structuring the dataset and preparing it for being analyzed. A dataset formed by 240 movements was obtained.

Then, the data has been analyzed using three different classifiers: A Fully Connected Artificial Neural Network (FC-ANN), A 1-Dimensional Convolutional Neural Network (1D-CNN), and a Long Short-Term Memory Neural Network (LSTM). The three classifiers have been trained over the dataset, giving generally good results. The analysis of the results shows that deactivating the normalization gave generally better results. The best classifier turned out to be LSTM, obtaining accuracies of 1.0 over the training set, and accuracies over 0.89 over the training set. In one of the cases the LSTM obtained 1.0 of accuracy over the training set, and 0.94 of accuracy over the testing set, and this one was the selected for developing the application.

This application, called KSAS (Kenpo Set Assisting System), runs on an Android device, and is able to capture the movements of a user, detecting if a movement has been correctly or wrongly executed by using a trained model, and give haptic and auditory feedback to the user. The application has been tested on one person, a beginner, that had a good impression of the app. Some feedback was obtained from the testing of the application, like the fact that the voice used in the application is annoying, that the execution in front of a mirror helps by providing a better self-observation, or that the use of haptic and auditory feedback was a good idea. The app has not been tested by more users because of the quarantine state caused by the outbreak of SARS-CoV-2.

6 FUTURE LINES OF RESEARCH

Many researches could derive from the one carried out in this Master's Thesis. Here, the developed application is able to assist the learning of a sequence of movements executed with one arm, but future lines of research could extend this application by using more devices and capturing data from more parts of the body. A set of futures lines of possible research are proposed here.

6.1 Development of a Net of Wearables

In the developed application, an Android device has been used for capturing the movements executed by the users. In a derived research, a net of wearables could be developed for wirelessly capturing of the data of different parts of the body and assist the learning of more complex and concurrent movements.

Some of this line of research have been advanced, and I have designed a low-cost Arduino-based wireless device using Fritzing[36] with this purpose. This kind of devices could be easily connected with each other through a Bluetooth module, as well as connected with Android devices or with a laptop for delegating heavy computer task.

The schematic design of the device is shown in Figure 69. The device is formed by an Arduino Nano, an inertial sensor MPU9250, an HC05 Bluetooth module, three LEDs of different colors for indicating states of the device, three 220 Ω resistors, one 1k Ω resistors and one 2k Ω resistor.

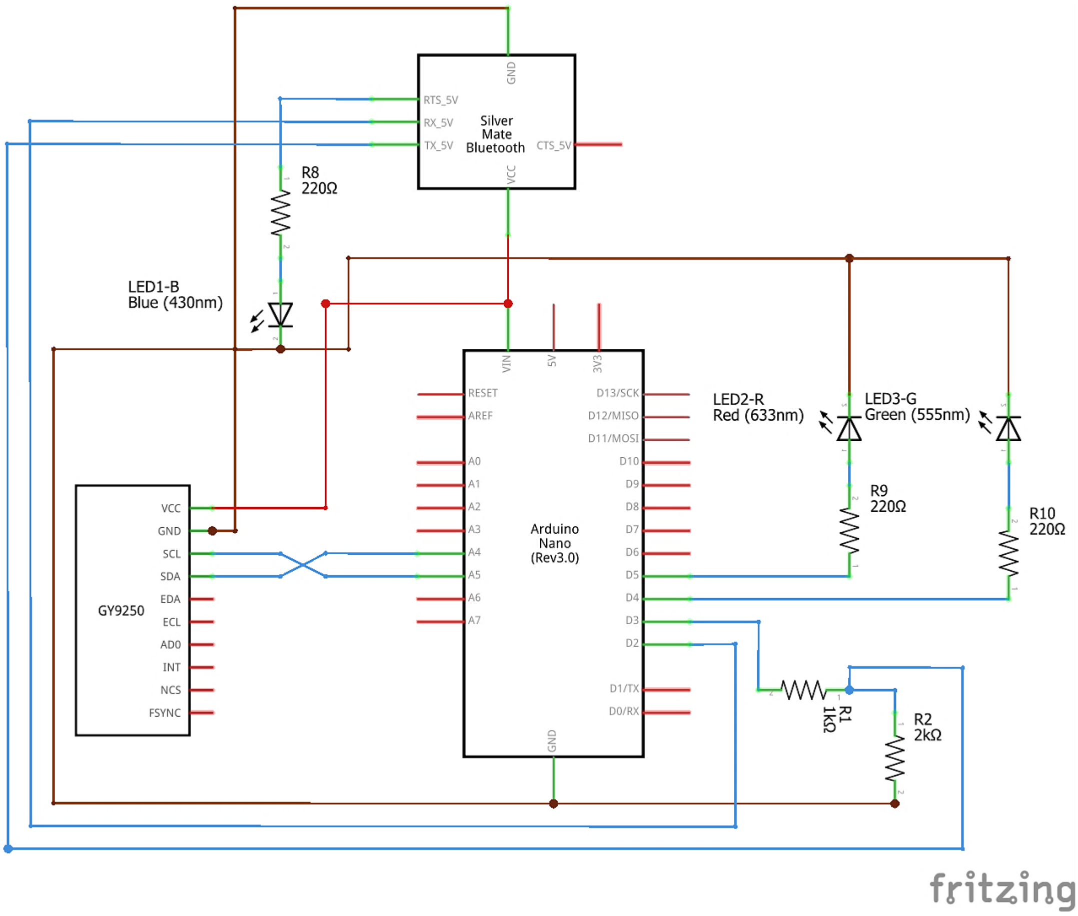

Figure 69. Schematic design of the wireless capture device.

[36] Fritzing webpage: https://fritzing.org/

Different modules and components could be added to this design for allowing more detailed feedback, such as vibrotactile or sound, directly from each wearable.

6.2 Teaching a Broader Set of Techniques and Movements

With the use of a net of wearables, movement capture from different parts of the body could be used for generating a 3D skeleton, or Labanotation scores. This could be useful for modelling more complex movements and assist the learning of complete sets, katas (forms) or techniques.

With a system of these characteristics, the full curriculum of a martial art could be modelled, and the system could follow up the student during the full learning process, learning the own characteristics of the user, as well as creating training routines based in the weaknesses or strengths of the student. It could also be used to assign the belt (level) to the users.

6.3 Considering Emotional Information in Martial Arts Learning

Martial arts practice involves the repetition over and over of movements, which can lead to different negative affective states such as boredom or frustration. In addition, in the questionnaire about the online learning of psychomotor activities the emotional situation of the user seemed to have an impact in the learning process. Thus, there is a need to research how the emotional state of the learner impacts in the learning of the martial art, especially considering an intra-subject approach that studies the user evolution along time.

Thus, the work carried out in this Master's Thesis has fed the INT^{2}AFF project (INTelligent INTra-subject development approach to improve actions in AFFect-aware adaptive educational systems) funded by the Spanish Ministry of Science, Innovation and Universities under Grant PGC2018-102279-B-I00. This project researchers how to deal with affective computing in education in different scenarios (including martial arts) along the Affective Life Cycle, which covers gathering affective information (using novel wearable devices and ambient features across sensor networks), detecting the affective state (features obtained with multimodal learning analytics from varied channels of interaction), modeling this state (labeling with self- or expert-based evaluations) and finally responding accordingly to the given situation (just-in-time multisensorial feedback).

6.4 Interaction Between People

The next level of assisting the learning of a student, is assisting the learning of more than one student that are interacting. An important aspect of martial art movements is that those movements have an application. Usually, the application of the movements involves the interaction between two or more subjects, and a system able to track their movements and the interaction between them, creating 3D visualizations of the movements and aiding them to apply the learnt techniques could be created. In relation to this, there is already a Final Carrier Project at the Computer Science School (developed by Jon Echeverría) which is developing the infrastructure to support the capture, modelling, analysis and feedback when training a kumite (combat) in Karate.

6.5 Training of Multiple Models

Right now, the dataset only contains the captured movements from 6 experts (i.e., persons who actively practice Kenpo Karate), but with an enough amount of captured movements from experts in Kenpo Karate, a multilevel application could be developed by training a classifier using movements captured only from experts, and another one using information captured from experts and beginners as well.

The classifier trained using movements from beginners and experts will be more permissive with some execution errors that beginners use to commit but will not allow errors in the sequence of movements. The classifier trained using movements only from experts will be stricter and will not allow execution errors, thus addressed for advanced users.

6.6 Testing with Users in the Wild

Due to the outbreak of COVID-19, the app could not be tested properly in more than two users, including the researcher writing this thesis. This leads to the necessity of testing the app with more people, beginners and experts, and get some feedback from them that could help to enhance the app and the feedback strategy. However, for the purposes of the research carried out, the testing done was enough to show the feasibility of the approach proposed and the validation of the AI techniques for modeling a specific movement in Kenpo Karate martial art.

6.7 Transfer Learning

The weights learned by the neural networks used in this research could be used for transferring the knowledge obtained from the movements' dataset to other martial arts, psychomotor activities, or even domains. This can be done by training the networks a few iterations over a new dataset.

In fact, the weights and networks obtained in this research are going to be applied, using transfer learning, to the dataset obtained in [164]. This dataset is formed by two movements collected from 20 volunteers and labelled by knee displacement, knee turn, and bokken hit.

6.8 Neuroevolution

In Section 1.4, the different subjects taken in the Master were mentioned, as well as how the obtained knowledge has been applied here. There is another subject that I took in the Master and is Evolutionary Computation.

There is a branch in Evolutionary Computation known as neuroevolution and consist in combining the plasticity of neural networks with the potential of evolutionary algorithms, so the weights of a neural network can be learned by means of an evolutionary algorithm while simultaneously the optimal network structure is explored. Neuroevolution can be applied in future lines of research for training better classifiers, including not only neural networks, but SVM and decision trees as well.

7 BIBLIOGRAPHY

[1] T. A. Green, *Martial Arts of the World: An Encyclopedia*. ABC-CLIO, 2001.

[2] B. Bu, H. Haijun, L. Yong, Z. Chaohui, Y. Xiaoyuan, and M. F. Singh, "Effects of martial arts on health status: A systematic review," *J. Evid. Based. Med.*, vol. 3, no. 4, pp. 205–219, 2010.

[3] S. Anne and D. Goldsmith, "Adolescent Participation in Traditional Martial Arts: Effects of Training on Risk Behaviors and Psychological Wellbeing," *Loma Linda Univ. Electron. Theses Diss. Pap. 122.*, 2013.

[4] Y. J. Ko, "Martial arts marketing: Putting the customer first," no. September, 2019.

[5] A. D. A. Mattiassi, "Fighting the game. Command systems and player-avatar interaction in fighting games in a social cognitive neuroscience framework," *Multimed. Tools Appl.*, vol. 78, no. 10, pp. 13565–13591, 2019.

[6] G. G. Muscolo and C. T. Recchiuto, "T.P.T. a novel Taekwondo personal trainer robot," *Rob. Auton. Syst.*, vol. 83, pp. 150–157, 2016.

[7] A. Westbrook and O. Ratti, *Aikido and the Dynamic Sphere: An Illustrated Introduction*. Tuttle Publishing, 2001.

[8] A. Corbi, O. C. Santos, and D. Burgos, "Intelligent framework for learning physics with aikido (martial art) and registered sensors," *Sensors (Switzerland)*, vol. 19, no. 17, pp. 1–18, 2019.

[9] A. Mroczkowski, "Using the Knowledge of Biomechanics in Teaching Aikido," 2012.

[10] E. K. Parker, *Infinite Insights Into Kenpo - Volumes 1-5*. BookSurge Publishing, 2009.

[11] A. M. Croom, "Embodying martial arts for mental health: Cultivating psychological wellbeing with martial arts practice," *Arch. Budo Sci. Martial Arts Extrem. Sport.*, vol. 10, pp. 59–70, 2014.

[12] J. M. Winkle and J. C. Ozmun, "Martial Arts: An Exciting Addition to the Physical Education Curriculum," *J. Phys. Educ. Recreat. Danc.*, vol. 74, no. 4, pp. 29–35, 2003.

[13] M. Theeboom and P. De Knop, "Asian Martial Arts and Approaches of Instruction in Physical Education," *Eur. J. Phys. Educ.*, vol. 4, no. 2, pp. 146–161, 1999.

[14] C. M. Gyves, "English Translation of General Qi Jiguang's 'Quanjing Jieyao Pian,'" vol. 298, 1993.

[15] M. Musashi, *The Book of Five Rings*. Japan, 1645.

[16] B. Cortés, *El Palo y el Sable: Teoría para el perfeccionamiento del manejo del sable por la esgrima del palo corto*. Biblioteca Nacional de España, 1851.

[17] B. Lee, *Tao of Jeet Kune Do*. Ohara Publications, 1975.

[18] M. Kitagawa and B. Windsor, *MoCap for Artists: Workflow and Techniques for Motion Capture*. 2008.

[19] M. Geroch, "Motion capture for the rest of us," *J. Comput. Sci. Coll.*, vol. 19, no. 3, pp. 157–164, 2004.

[20] O. C. Santos, "Psychomotor Learning in Martial Arts: an Opportunity for User Modeling, Adaptation and Personalization," *Adjun. Publ. 25th Conf. User Model. Adapt. Pers. -*

UMAP '17, pp. 89–92, 2017.

[21] O. C. Santos, "Artificial Intelligence in Psychomotor Learning: Modeling Human Motion from Inertial Sensor Data," *Int. J. Artif. Intell. Tools*, 2019.

[22] O. C. Santos, "Toward personalized vibrotactile support when learning motor skills," *Algorithms*, 2017.

[23] O. C. Santos, "Emotions and Personality in Adaptive e-Learning Systems: An Affective Computing Perspective," pp. 263–285, 2016.

[24] O. C. Santos, "Training the Body: The Potential of AIED to Support Personalized Motor Skills Learning," *Int. J. Artif. Intell. Educ.*, 2016.

[25] S. Bauer *et al.*, "Real-Time Range Imaging in Health Care : A Survey," pp. 228–254.

[26] B. Pueo, J. M. Jimenez-olmedo, U. D. A. España, and B. Pueo, "Application of motion capture technology for sport performance analysis," vol. 2041, pp. 241–247, 2017.

[27] J. G. Richards, "The measurement of human motion: A comparison of commercially available systems," *Hum. Mov. Sci.*, vol. 18, no. 5, pp. 589–602, 1999.

[28] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A study of vicon system positioning performance," *Sensors (Switzerland)*, vol. 17, no. 7, pp. 1–18, 2017.

[29] Z. Wang, C. Zhang, Z. Wang, Y. Shuang, and Y. Li, "Real-time three-dimensional measurement techniques - A review," *Optik (Stuttg).*, vol. 202, no. September 2019, p. 163627, 2020.

[30] S. K. Mada, M. L. Smith, L. N. Smith, and P. S. Midha, "An overview of Passive and Active Vision Techniques for Hand-Held 3D Data Acquisition," in *Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision*, 2003.

[31] Y. W. Kuan, N. O. Ee, and L. S. Wei, "Comparative study of intel R200, Kinect v2, and primesense RGB-D sensors performance outdoors," *IEEE Sens. J.*, vol. 19, no. 19, pp. 8741–8750, 2019.

[32] O. Wasenmüller and D. Stricker, "Comparison of kinect v1 and v2 depth images in terms of accuracy and precision," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10117 LNCS, pp. 34–45, 2017.

[33] D. Beltran and L. Basañez, "A comparison between active and passive 3D vision sensors: BumblebeeXB3 and microsoft kinect," in *Advances in Intelligent Systems and Computing*, 2014.

[34] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimed.*, vol. 19, no. 2, pp. 4–10, 2012.

[35] A. Shafaei and J. J. Little, "Real-time human motion capture with multiple depth cameras," *Proc. - 2016 13th Conf. Comput. Robot Vision, CRV 2016*, pp. 24–31, 2016.

[36] L. Zhang, J. Sturm, D. Cremers, and D. Lee, "Real-time human motion tracking using multiple depth cameras," *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 2389–2395, 2012.

[37] K. Berger, K. Ruhl, Y. Schroeder, C. Bruemmer, A. Scholz, and M. Magnor, "Markerless motion capture using multiple color-depth sensors," *VMV 2011 - Vision, Model. Vis.*, pp. 317–324, 2011.

[38] H. M. Zhu and C. M. Pun, "Human action recognition with skeletal information from

depth camera," *2013 IEEE Int. Conf. Inf. Autom. ICIA 2013*, no. August, pp. 1082–1085, 2013.

[39] X. Wei, P. Zhang, and J. Chai, "Accurate realtime full-body motion capture using a single depth camera," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–12, 2012.

[40] D. Regazzoni, G. De Vecchi, and C. Rizzi, "RGB cams vs RGB-D sensors: Low cost motion capture technologies performances and limitations," *J. Manuf. Syst.*, vol. 33, no. 4, pp. 719–728, 2014.

[41] X. Feng and P. Perona, "Human action recognition by sequence of movelet codewords," *Proc. - 1st Int. Symp. 3D Data Process. Vis. Transm. 3DPVT 2002*, pp. 717–721, 2002.

[42] A. Houssein Ahmed, K. Kpalma, and A. Osman Guedi, "Human detection using HOG-SVM, mixture of Gaussian and background contours subtraction," *Proc. - 13th Int. Conf. Signal-Image Technol. Internet-Based Syst. SITIS 2017*, vol. 2018-Janua, pp. 334–338, 2018.

[43] L. Zhang and Y. Liang, "Motion human detection based on background subtraction," *2nd Int. Work. Educ. Technol. Comput. Sci. ETCS 2010*, vol. 1, pp. 284–287, 2010.

[44] M. Dimitrijevic, V. Lepetit, and P. Fua, "Human Body Pose Recognition using Spatio-Temporal Templates," *ICCV Work. Model. People Hum. Interact.*, p. 8, 2005.

[45] N. Hasler, B. Rosenhahn, T. Thormahlen, M. Wand, J. Gall, and H.-P. Seidel, "Markerless Motion Capture with unsynchronized moving cameras," pp. 224–231, 2010.

[46] B. Rosenhahn, T. Brox, and J. Weickert, "Three-dimensional shape knowledge for joint image segmentation and pose tracking," *Int. J. Comput. Vis.*, vol. 73, no. 3, pp. 243–262, 2007.

[47] X. Ren, A. C. Berg, and J. Malik, "Recovering human body configurations using pairwise constraints between parts," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. I, pp. 824–831, 2005.

[48] H. ShangGuan and M. Ramakrishnan, "3D video-based motion capture using Convolutional Neural Networks," *2017 Int. Conf. Image Vis. Comput. New Zeal. (IVCNZ), Christchurch*, pp. 1–6, 2017.

[49] X. Zhao, H. Ning, Y. Liu, and T. Huang, "Discriminative estimation of 3D human pose using Gaussian processes," *Proc. - Int. Conf. Pattern Recognit.*, pp. 3–6, 2008.

[50] D. Mehta *et al.*, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, 2017.

[51] X. Zhou, S. Liu, G. Pavlakos, V. Kumar, and K. Daniilidis, "Human Motion Capture Using a Drone," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 2027–2033, 2018.

[52] K. Hoshino, M. Nakamura, Y. Nakai, Y. Ozone, S. Shimanoe, and Y. Noguchi, "A world camera for recording the game tactics in martial arts using bamboo swords," *ACM Int. Conf. Proceeding Ser.*, pp. 188–192, 2019.

[53] P. H. Han, Y. S. Chen, Y. Zhong, H. L. Wang, and Y. P. Hung, "My Tai-Chi coaches: An augmented-learning tool for practicing Tai-Chi Chuan," in *ACM International Conference Proceeding Series*, 2017.

[54] H. Zacharatos, C. Gatzoulis, Y. Chrysanthou, and A. Aristidou, "Emotion Recognition for Exergames using Laban Movement Analysis," *Proc. Motion Games - MIG '13*, pp. 61–66, 2013.

[55] K. Kolykhalova, A. Camurri, G. Volpe, M. Sanguineti, E. Puppo, and R. Niewiadomski, "A

multimodal dataset for the analysis of movement qualities in karate martial art," *Proc. 2015 7th Int. Conf. Intell. Technol. Interact. Entertain. INTETAIN 2015*, vol. 17, pp. 74–78, 2015.

[56] OmniVision Technologies, "OV7670/OV7171 CMOS VGA (OmniPixel ®) CAMERACHIP™ Sensor Omnivision Datasheet." 2006.

[57] J. Borenstein, H. R. Everett, L. Feng, and D. Wehe, "Mobile robot positioning: Sensors and techniques," *J. Robot. Syst.*, vol. 14, no. 4, pp. 231–249, 1997.

[58] R. Siegward and I. R. Nourbakhsh, *Introduction to Autonomous Mobile Robots*, 1st ed. The MIT press, 2004.

[59] M. Supej, "3D Measurements of alpine skiing with an inertial sensor motion capture suit and GNSS RTK system," *J. Sports Sci.*, vol. 28, no. 7, pp. 759–769, 2010.

[60] H. Yang and J. Ye, "A calibration process for tracking upper limb motion with inertial sensors," *2011 IEEE Int. Conf. Mechatronics Autom. ICMA 2011*, vol. 1, pp. 618–623, 2011.

[61] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors," *Xsens Motion Technol. BV, …*, no. January 2009, pp. 1–7, 2009.

[62] Z. Lin *et al.*, "Development of an ultra-miniaturized inertial measurement unit WB-3 for human body motion tracking," *2010 IEEE/SICE Int. Symp. Syst. Integr. SI Int. 2010 - 3rd Symp. Syst. Integr. SII 2010, Proc.*, pp. 414–419, 2010.

[63] J. M. Lambrecht and R. F. Kirsch, "Miniature low-power inertial sensors: Promising technology for implantable motion capture systems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 6, pp. 1138–1147, 2014.

[64] P. G. Jung, G. Lim, and K. Kong, "A mobile motion capture system based on inertial sensors and smart shoes," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 692–697, 2013.

[65] J. Wu, Z. Wang, S. Raghuraman, B. Prabhakaran, and R. Jafari, "Demonstration abstract: Upper body motion capture system using inertial sensors," *IPSN 2014 - Proc. 13th Int. Symp. Inf. Process. Sens. Networks (Part CPS Week)*, pp. 351–352, 2014.

[66] I. Prayudi, E. H. Seo, D. Kim, and B. J. You, "Implementation of an inertial measurement Unit based motion capture system," *URAI 2011 - 2011 8th Int. Conf. Ubiquitous Robot. Ambient Intell.*, pp. 425–429, 2011.

[67] G. To and M. R. Mahfouz, "Design of wireless inertial trackers for human joint motion analysis," *RWW 2012 - Proc. 2012 IEEE Top. Conf. Biomed. Wirel. Technol. Networks, Sens. Syst. BioWireleSS 2012*, pp. 49–52, 2012.

[68] A. D. Young, "From posture to motion: The challenge for real time wireless inertial motion capture," *Proc. 5th Int. ICST Conf. Body Area Networks, BodyNets 2010*, pp. 131–137, 2011.

[69] M. Brodie, A. Walmsley, and W. Page, "Fusion motion capture: a prototype system using inertial measurement units and GPS for the biomechanical analysis of ski racing," *Sport. Technol.*, vol. 1, no. 1, pp. 17–28, 2008.

[70] A. Szczęsna *et al.*, "Inertial motion capture costume design study," *Sensors (Switzerland)*, vol. 17, no. 3, pp. 6–9, 2017.

[71] M. Jin, J. Zhao, J. Jin, G. Yu, and W. Li, "The adaptive Kalman filter based on fuzzy logic for

inertial motion capture system," *Meas. J. Int. Meas. Confed.*, vol. 49, no. 1, pp. 196–204, 2014.

[72] C. Xiang, H. H. Hsu, W. Y. Hwang, and J. Ma, "Comparing real-time human motion capture system using inertial sensors with microsoft kinect," *Proc. - 2014 7th Int. Conf. Ubi-Media Comput. Work. U-MEDIA 2014*, pp. 53–58, 2014.

[73] W. Frey, "Application of inertial sensors and flux-gate magnetometer to real-time human body motion capture," 1996.

[74] H. Fourati, N. Manamanni, L. Afilal, and Y. Handrich, "Complementary observer for body segments motion capturing by inertial and magnetic sensors," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 1, pp. 149–157, 2014.

[75] F. Destelle *et al.*, "Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors," *Eur. Signal Process. Conf.*, no. 22, pp. 371–375, 2014.

[76] C. Wong, Z. Zhang, B. Lo, and G. Z. Yang, "Markerless motion capture using appearance and inertial data," *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014*, pp. 6907–6910, 2014.

[77] G. Pons-Moll, A. Baak, T. Helten, M. Müller, H. P. Seidel, and B. Rosenhahn, "Multisensor-fusion for 3D full-body human motion capture," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 663–670, 2010.

[78] M. Trumble, A. Gilbert, C. Malleson, A. Hilton, and J. Collomosse, "Total capture: 3D human pose estimation fusing video and inertial sensors," *Br. Mach. Vis. Conf. 2017, BMVC 2017*, pp. 1–13, 2017.

[79] D. Vlasic *et al.*, "Practical motion capture in everyday surroundings," *ACM Trans. Graph.*, vol. 26, no. 99, p. 35, 2007.

[80] I. Inc., "MPU-9259 Product Specification." 2016.

[81] N. Ward, M. Ortiz, F. Bernardo, and A. Tanaka, "Designing and measuring gesture using laban movement analysis and electromyogram," *Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Adjun. - UbiComp '16*, pp. 995–1000, 2016.

[82] P. V. S. Moreira, M. F. Goethel, and M. Gonçalves, "Neuromuscular performance of Bandal Chagui: Comparison of subelite and elite taekwondo athletes," *J. Electromyogr. Kinesiol.*, vol. 30, pp. 55–65, 2016.

[83] G. Shan, J. Evans, and S.-T. Chang, "Development of a real-time biofeedback tool for martial arts coaching practice," *2015 Int. Conf. Biomech. Sport.*, pp. 94–97, 2015.

[84] H. S. Aandahl, E. von Heimburg, and R. van Den Tillaar, "Effect of postactivation potentiation induced by elastic resistance on kinematics and performance in a roundhouse kick of trained martial arts practitioners," *J. Strength Cond. Res.*, vol. 32, no. 4, pp. 990–996, 2018.

[85] S. Fdili Alaoui, J. Françoise, T. Schiphorst, K. Studd, and F. Bevilacqua, "Seeing, Sensing and Recognizing Laban Movement Qualities," *Proc. 2017 CHI Conf. Hum. Factors Comput. Syst. - CHI '17*, pp. 4009–4020, 2017.

[86] J. Françoise, S. Fdili Alaoui, T. Schiphorst, and F. Bevilacqua, "Vocalizing dance movement for interactive sonification of laban effort factors," in *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*, 2014.

[87] S. Yabukami *et al.*, “Motion capture system of magnetic markers using three-axial magnetic field sensor,” *IEEE Trans. Magn.*, vol. 36, no. 5 I, pp. 3646–3648, 2000.

[88] S. Hashi, M. Toyoda, S. Yabukami, K. Ishiyama, Y. Okazaki, and K. Arai, “Development of wireless magnetic motion capture system for multi-marker detection,” in *INTERMAG 2006 - IEEE International Magnetics Conference*, 2006, vol. 42, no. 10, p. 524.

[89] E. Fléty, “3D Gesture Acquisition Using Ultrasonic Sensors,” *Trends Gestural Control Music*, pp. 193–208, 2000.

[90] K. Ashhar, C. B. Soh, and K. H. Kong, “A Wearable Ultrasonic Sensor Network for Analysis of Bilateral Gait Symmetry,” *2017 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4455–4458, 2017.

[91] Y. Qi, C. B. Soh, E. Gunawan, and K. Low, “A Wearable Wireless Ultrasonic Sensor Network for Human Arm Motion Tracking,” *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 5960–5963, 2014.

[92] Y. Qi, C. B. Soh, E. Gunawan, and K. Low, “Ambulatory Measurement of Three-Dimensional Foot Displacement During Treadmill Walking Using,” vol. 19, no. 2, pp. 446–452, 2015.

[93] D. Laurijssen, S. Truijen, W. Saeys, W. Daems, and J. Steckel, “An Ultrasonic Six Degrees-of-Freedom Pose Estimation Sensor,” vol. 17, no. 1, pp. 151–159, 2017.

[94] T. Sato, “Design and Implementation of a Robust and Real-time Ultrasonic Motion-capture System,” *2011 Int. Conf. Indoor Position. Indoor Navig.*, pp. 1–6.

[95] D. Laurijssen, S. Truijen, W. Saeys, and J. Steckel, “Three Sources , Three Receivers , Six Degrees of Freedom : An Ultrasonic Sensor for Pose Estimation & Motion Capture,” *2015 IEEE SENSORS*, pp. 1–4, 2015.

[96] R. J. Przybyla, H. Tang, S. E. Shelton, D. A. Horsley, and B. E. Boser, “3D Ultrasonic Gesture Recognition,” *2014 IEEE Int. Solid-State Circuits Conf. Dig. Tech. Pap.*, pp. 210–211, 2014.

[97] Y. Qifan, T. Hao, Z. Xuebing, and L. Yin, “Dolphin : Ultrasonic-based Gesture Recognition on Smartphone Platform,” *2014 IEEE 17th Int. Conf. Comput. Sci. Eng.*, pp. 1461–1468, 2014.

[98] B. Van Dam, Y. Murillo, M. Li, and S. Pollin, “In-air ultrasonic 3D-touchscreen with gesture recognition using existing hardware for smart devices,” *2016 IEEE Int. Work. Signal Process. Syst.*, pp. 74–79, 2016.

[99] X. Li, H. Dai, L. Cui, and Y. Wang, “SonicOperator : Ultrasonic Gesture Recognition with Deep Neural Network on Mobiles,” *2017 IEEE SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov.*, pp. 1–7, 2017.

[100] M. West, S. I. Limited, and S. Centre, “Developing High Quality Data Models,” *Dev. High Qual. Data Model.*, 2011.

[101] L. Kovar, M. Gleicher, and F. Pighin, “Motion graphs,” *ACM SIGGRAPH 2008 Classes*, p. 51, 2008.

[102] Z. Li, Y. Zhou, S. Xiao, C. He, Z. Huang, and H. Li, “Auto-Conditioned Recurrent Networks for Extended Complex Human Motion Synthesis,” pp. 1–13, 2017.

[103] R. . Agrawal and A. K. Ratnadip, *An Introductory Study on Time Series Modeling and*

Forecasting, vol. 1302.6613. 2013.

[104] A. Karapurkar, "Modeling Human Activities," *Sch. Pap. Arch. Dep. Comput. Sci. Univ. Maryland, Coll. Park. MD*, vol. 20742, pp. 1–12, 2008.

[105] V. Ramakrishna, T. Kanade, and Y. Sheikh, "Reconstructing 3D human pose from 2D image landmarks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7575 LNCS, no. PART 4, pp. 573–586, 2012.

[106] W. Zhang, Z. Liu, L. Zhou, H. Leung, and A. B. Chan, "Martial Arts, Dancing and Sports dataset: A challenging stereo and multi-view dataset for 3D human pose estimation," *Image Vis. Comput.*, vol. 61, pp. 22–39, 2017.

[107] Y. Shinagawa, J. I. Nakajima, and T. L. Kunii, "Capturing and analyzing stability of human body motions using video cameras," *Comput. Animat. Conf. Proc.*, pp. 48–57, 1997.

[108] M. Müller, *Information retrieval for music and motion*. Springer, 2007.

[109] O. College, *Anatomy & Physiology*. OpenStax College, 2013.

[110] B. O. Community, "Blender - a 3D modelling and rendering package." Blender Foundation, Stichting Blender Foundation, Amsterdam.

[111] J. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied regression analysis: a research tool*, 2nd ed. Springer-Verlag New York, 1998.

[112] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. 2010.

[113] A. Ng, Y. Bensouda Mourri, and K. Katanforoosh, "Deep Learning Specialization (Online Courses)." Coursera.

[114] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," vol. 323, no. 6088, pp. 533–536, 1986.

[115] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[116] R. Vezzani, D. Baltieri, and R. Cucchiara, "HMM Based Action Recognition with Projection Histogram Features," *Recognizing Patterns Signals, Speech, Images Videos - ICPR 2010 Contests*, pp. 23–26, 2010.

[117] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," *Proc. 1992 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 92, pp. 379–385, 1992.

[118] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," *Int. Work. Hum. Behav. Underst.*, pp. 29–39, 2011.

[119] E. P. Ijjina and C. K. Mohan, "Human action recognition based on motion capture information using fuzzy convolution neural networks," *2015 Eighth Int. Conf. Adv. Pattern Recognit.*, pp. 1–6, 2015.

[120] Y. Hou, Z. Li, P. Wang, and W. Li, "Skeleton Optical Spectra-Based Action Recognition Using Convolutional Neural Networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 807–811, 2018.

[121] T. Hachaj, M. Piekarczyk, and M. R. Ogiela, "Human actions analysis: Templates generation, matching and visualization applied to motion capture of highly-skilled karate athletes," *Sensors (Switzerland)*, vol. 17, no. 11, 2017.

[122] F. Bevilacqua, F. Guédy, N. Schnell, E. Fléty, and N. Leroy, “Wireless sensor interface and gesture-follower for music pedagogy,” in *Proceedings of the 7th International Conference on New Interfaces for Musical Expression, NIME ’07*, 2007.

[123] A. H. Guest, *Labanotation: the system of analyzing and recording movement*. Routledge, 2014.

[124] O. C. Santos and M. H. Eddy, “Modeling Psychomotor Activity: Current Approaches and Open Issues,” *UMAP’17 Adjun.*, pp. 305–310, 2017.

[125] J. Hat, “Movementxml : a Representation of Semantics of Human Movement Based on Labanotation,” *Channels*, 2006.

[126] A. Sankhla *et al.*, “Automated translation of human postures from kinect data to labanotation,” *Natl. Conf. Comput. Vision, Pattern Recognition, Image Process. Graph.*, pp. 494–505, 2018.

[127] W. Choensawat, M. Nakamura, and K. Hachimura, “GenLaban: A tool for generating Labanotation from motion capture data,” *Multimed. Tools Appl.*, vol. 74, no. 23, pp. 10823–10846, 2015.

[128] K. Ikeuchi, Z. Ma, Z. Yan, S. Kudoh, and M. Nakamura, “Describing Upper-Body Motions Based on Labanotation for Learning-from-Observation Robots,” *Int. J. Comput. Vis.*, vol. 126, no. 12, pp. 1415–1429, 2018.

[129] C. J. Gavagan and M. G. Sayers, “A biomechanical analysis of the roundhouse kicking technique of expert practitioners: A comparison between the martial arts disciplines of Muay Thai, Karate, and Taekwondo,” *Public Libr. Sci. ONE*, vol. 12, no. 8, pp. 351–356, 2017.

[130] M. Błaszczyszyn, A. Szczęsna, M. Pawlyta, M. Marszałek, and D. Karczmit, “Kinematic analysis of Mae-Geri kicks in beginner and advanced Kyokushin karate athletes,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 17, 2019.

[131] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso, “Preprocessing techniques for context recognition from accelerometer data,” *Pers. Ubiquitous Comput.*, vol. 14, no. 7, pp. 645–662, 2010.

[132] J. Liu, A. Shahroudy, D. Xu, and G. Wang, “Spatio-temporal lstm with trust gates for 3d human action recognition,” *Eur. Conf. Comput. Vis.*, pp. 816--833, 2016.

[133] W. Zucchini, I. L. MacDonald, and R. Langrock, *Hidden Markov models for time series: an introduction using R*. CRC press, 2017.

[134] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, “Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review,” *Psychon. Bull. Rev.*, vol. 20, no. 1, pp. 21–53, 2013.

[135] E. Lantz, S. Bryson, D. Zeltzer, M. T. Bolas, B. De La Chapelle, and D. Bennett, “The future of virtual reality: Head mounted displays versus spatially immersive displays,” in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996*, 1996.

[136] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, “Surround-screen projection-based virtual reality: The design and implementation of the CAVE,” in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1993*, 1993.

[137] J. J. LaViola, “A discussion of cybersickness in virtual environments,” *ACM SIGCHI Bull.*,

2000.

[138] T. Komura, B. Lam, R. W. H. Lau, and H. Leung, “e-Learning martial arts,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4181 LNCS, pp. 239–248, 2006.

[139] X. Chen *et al.*, “ImmerTai: Immersive Motion Learning in VR Environments,” *J. Vis. Commun. Image Represent.*, vol. 58, pp. 416–427, 2019.

[140] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd ed. Prentice Hall, 2002.

[141] F. Hülsmann, J. P. Göpfert, B. Hammer, S. Kopp, and M. Botsch, “Classification of motor errors to provide real-time feedback for sports coaching in virtual reality — A case study in squats and Tai Chi pushes,” *Comput. Graph.*, vol. 76, pp. 47–59, 2018.

[142] N. Powell and J. Lumsden, “Exploring novel auditory displays for supporting accelerated skills acquisition and enhanced performance in motorsport,” *Proc. 21st Int. Conf. Audit. Disp.*, pp. 173–180, 2015.

[143] J. Dyer, P. Stapleton, and M. Rodger, “Sonification of Movement for Motor Skill Learning in a Novel Bimanual Task: Aesthetics and Retention Strategies,” pp. 99–102, 2017.

[144] Y. Zheng and J. B. Morrell, “A vibrotactile feedback approach to posture guidance,” *2010 IEEE Haptics Symp. HAPTICS 2010*, pp. 351–358, 2010.

[145] K. Bark, J. W. Wheeler, S. Premakumar, and M. R. Cutkosky, “Comparison of skin stretch and vibrotactile stimulation for feedback of proprioceptive information,” *Symp. Haptics Interfaces Virtual Environ. Teleoperator Syst. 2008 - Proceedings, Haptics*, pp. 71–78, 2008.

[146] B. Weber, S. Schätzle, T. Hulin, C. Preusche, and B. Deml, “Evaluation of a vibrotactile feedback device for spatial guidance,” *2011 IEEE World Haptics Conf. WHC 2011*, pp. 349–354, 2011.

[147] S. . H. T. . P. C. and H. G. Schaetzle, “Evaluation of Vibrotactile Feedback to the Human Arm,” *EuroHaptics 2006*, pp. 557–560, 2006.

[148] J. Van Der Linden, E. Schoonderwaldt, J. Bird, and R. Johnson, “MusicJacket - Combining motion capture and vibrotactile feedback to teach violin bowing,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 1, pp. 104–113, 2011.

[149] J. Lieberman and C. Breazeal, “Development of a wearable vibrotactile feedback suit for accelerated human motor learning,” *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 23, no. 5, pp. 4001–4006, 2007.

[150] R. W. Lindeman, R. Page, Y. Yanagida, and J. L. Sibert, “Towards full-body haptic feedback: The design and deployment of a spatialized vibrotactile feedback system,” *Proc. ACM Symp. Virtual Real. Softw. Technol. VRST*, pp. 146–149, 2004.

[151] R. W. Byrne and A. E. Russon, “Learning by imitation : A hierarchical approach,” pp. 667–721, 1998.

[152] M. Čoh, D. Jovanović-Golubović, and M. Bratić, “Motor learning in sport,” *Facta Univ. - Ser. Phys. Educ. Sport*, vol. 2, no. 1, pp. 45–59, 2004.

[153] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.2017*. 2018.

[154] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by Watching: Extracting Reusable Task Knowledge from Visual Observation of Human Performance," *IEEE Trans. Robot. Autom.*, vol. 10, no. 6, pp. 799–822, 1994.

[155] M. Okumura, A. Kijima, K. Kadota, K. Yokoyama, H. Suzuki, and Y. Yamamoto, "A Critical Interpersonal Distance Switches between Two Coordination Modes in Kendo Matches," *PLoS One*, vol. 7, no. 12, 2012.

[156] S. Phunsa, N. Potisarn, and S. Tirakoat, "Edutainment - Thai art of self-defense and boxing by motion capture technique," *Proc. - 2009 Int. Conf. Comput. Model. Simulation, ICCMS 2009*, pp. 152–155, 2009.

[157] G. Mustapha, M. Zakaria, and W. R. Wan, "The Mechanical Aspects of Martial Arts: Total Time of Execution and Kinematics of Kaedah A," *Proc. Int. Colloq. Sport. Sci. Exerc. Eng. Technol. 2014 (ICoSSEET 2014)*, vol. 2014, no. ICoSSEET 2014, pp. 3–12, 2014.

[158] G. Mustapha, M. F. A. Razak, M. S. M. Hamzah, N. H. M. Yahya, and J. Mahmud, "The development of a low cost motion analysis system: Cekak Visual 3D v1.0," *Int. J. GEOMATE*, vol. 11, no. 2, pp. 2248–2252, 2016.

[159] P. Hämäläinen, T. Ilmonen, J. Höysniemi, M. Lindholm, and A. Nykänen, "Martial arts in artificial reality," *CHI 2005 Technol. Safety, Community Conf. Proc. - Conf. Hum. Factors Comput. Syst.*, no. 1, pp. 781–790, 2005.

[160] L. Zhang *et al.*, "KaraKter: An autonomously interacting Karate Kumite character for VR-based training and research," *Comput. Graph.*, vol. 72, pp. 59–69, 2018.

[161] B. Emad *et al.*, "iKarate: Improving Karate Kata," *Procedia Comput. Sci.*, vol. 170, no. 2019, pp. 466–473, 2020.

[162] Y. Kamaya, T. Naka, M. Yamada, and S. Miyazaki, "Information visualization for virtual martial arts training," *Proc. - 2018 NICOGRAPH Int. NICOINT 2018*, pp. 66–69, 2018.

[163] E. Wu and H. Koike, "FuturePose - Mixed reality martial arts training using real-time 3D human pose forecasting with a RGB camera," *Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vision, WACV 2019*, pp. 1384–1392, 2019.

[164] A. Corbí and O. C. Santos, "MyShikko: Modelling knee walking in Aikido practice," in *UMAP 2018 - Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, 2018.

[165] M. Zago *et al.*, "Dynamic balance in elite karateka," *J. Electromyogr. Kinesiol.*, vol. 25, no. 6, pp. 894–900, 2015.

[166] Y. Shen, H. Wang, E. S. L. Ho, L. Yang, and H. P. H. Shum, "Posture-based and action-based graphs for boxing skill visualization," *Comput. Graph.*, vol. 69, pp. 104–115, 2017.

[167] A. Soga, "Support System Using Motion Data for Creating Solo Performances of Shorinji Kempo," *Proc. - 2017 Int. Conf. Cult. Comput. Cult. Comput. 2017*, vol. 2017-Decem, pp. 161–162, 2017.

[168] V. E. Di Bacco, M. Taherzadeh, O. Birot, and W. H. Gage, "The effects of single versus multiple training sessions on the motor learning of two Krav Maga strike techniques, in women," *PeerJ*, vol. 2020, no. 2, 2020.

[169] D. Mehta *et al.*, "Monocular 3D human pose estimation in the wild using improved CNN supervision," in *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, 2018.

[170] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014.

[171] T. Hachaj, M. R. Ogiela, and M. Piekarczyk, “The open online repository of karate motion capture data: A tool for scientists and sport educators,” in *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings*, 2018.

APPENDIX I – METHODS USED IN THE REVIEWED PUBLICATIONS ABOUT MARTIAL ARTS

Below, a table with information about the reviewed publications regarding applications or the study of martial arts is shown. In the first column, the reference of the publication is presented based in the numeration of Section 7. The columns are described as follow:

- The first column corresponds with the martial arts used in the publication.
- The second column corresponds with the motion capture methods applied. In the cases of depth sensors, inertial suits and of sets of cameras with markers, the system used is indicated between parentheses. When RGB cameras are used as auxiliary method, the purpose of its use is indicated between parentheses.
- The third column correspond with the modelling method applied. If the publication also model actions, the method is indicated as well.
- The fourth column correspond with the analysis method applied. Methods using mathematical equations or statistical processing are grouped under the name of "Mathematical and/or Statistical".
- The fifth column correspond with the feedback method applied. The kind of system used is indicated between parentheses.

Table 19. Martial arts and methods employed in each document

Paper	Martial Arts	Motion Capture Method	Motion Modelling Method	Motion Analysis Method	Feedback Method
[129]	· Muay Thai · Karate · Tae Kwon Do	· Set of Cameras with Reflective Markers (Qualisys)	· Motion: Succession of partial 3D Skeletons	· Mathematical and/or Statistical	· None
[155]	· Kendo	· Set of Cameras with Reflective Markers (Qualisys) · RGB Camera (Record scene)	· Motion: Succession of 3D Skeletons	Mathematical and/or Statistical	· None
[55]	· Karate	· Set of Cameras with Reflective Markers (Qualisys) · RGB Camera (Synchronization of data)	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical	· None
[52]	· Tankendo	· RGB Camera	· None	· None	· None
[107]	· Shorinji Kempo	· Set of RGB Cameras	· Motion: Succession of 3D Silhouettes	· Mathematical and/or Statistical	· None
[141]	· Tai Chi	· Set of Cameras with Reflective Markers (OptiTrack)	· Motion: Succession of 3D Skeleton · Action: Random Forest	· Random Forest (Feature Extraction) · SVM (Classification)	· Auditory (Verbal) · Visual (VR - CAVE)
[83]	· Non specified	· Set of Cameras with Reflective Markers (Non-Specified) · EMG	· Motion: Succession of 3D Skeletons and series of characteristics of motion	· Mathematical and/or Statistical	· None
[165]	· Karate	· Set of Cameras with Reflective Markers (BTS)	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical	· None
[156]	· Muay Thai	· Set of Cameras with Reflective Markers (MotionAnalysis)	· Motion: Succession of 3D Skeletons	· None	· None
[84]	· Kick Boxing · Tae Kwon Do	· Set of Cameras with Reflective Markers (Qualisys) · EMG	· Motion: Succession of 3D Skeletons and series of characteristics of motion	· Mathematical and/or Statistical	· None
[138]	· Boxing	· Set of Cameras with Reflective Markers (Non-Specified)	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical	· Visual (VR - HMD)
[163]	· Non-Specified	· RGB Camera (Capture)	· Motion: Succession of 2D silhouettes and 3D Skeletons (Inferred from 2D silhouettes). · Action: RNN	· RNN (Forecast pose) · MA (Filtering noise) · Kalman Filter (Filtering noise)	· Visual (VR – HMD)

Paper	Martial Arts	Motion Capture Method	Motion Modelling Method	Motion Analysis Method	Feedback Method
				· Threshold value (Filtering noise)	
[121]	· Karate	· Inertial Sensors (Shadow 2.0)	· Motion: Succession of 3D Skeletons · Action: Templates	· DTW (Compare execution with templates)	· None
[161]	· Karate	· Depth Sensor (Kinect V2)	· Motion: Succession of 3D Skeletons	· F-DTW	· Visual (Screen)
[139]	· Tai Chi	· Depth Sensor (Kinect)	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical · DTW	· Visual (VR - CAVE) · Visual (VR – HMD) · Visual (Screen)
[162]	· Non specified	· Depth Sensor (Kinect V2)	· Motion: Succession of 3D Skeletons	· None	· Visual (VR – HMD)
[160]	· Karate	· Set of Cameras with Reflective Markers (ART and Vicon)	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical	· Visual (VR – CAVE) · Visual (VR – HMD)
[130]	· Karate	· Set of Cameras with Reflective Markers (Vicon) · RGB Cameras	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical	· None
[159]	· Kung fu	· RGB Camera	· Motion: Succession of 2D silhouettes	· Mathematical and/or Statistical	· Visual (AR – Screen) · Auditory (Sounds)
[106]	· Karate · Tai Chi	· Multiple RGB Cameras · Depth Sensor (Bumblebee II) · Set of Cameras with Reflective Markers (MotionAnalysis)	· Motion: Succession of 3D Skeletons, 2D silhouettes and 3D silhouettes	· Mathematical and/or Statistical	· None
[53]	· Tai Chi	· RGB Camera (Fisheye camera) · Drone	· Motion: Succession of 2D silhouettes	· None	· Visual (AR – HMD)
[82]	· Tae Kwon Do	· Set of cameras with Reflective Markers (Vicon) · EMG	· Motion: Succession of 3D Skeletons and series of characteristics of motion	· Mathematical and/or Statistical	· None
[166]	· Boxing	· Set of cameras with Reflective Markers (Non-specified)	· Motion: Succession of 3D Skeletons · Actions: Motion graphs	· None	· None

Paper	Martial Arts	Motion Capture Method	Motion Modelling Method	Motion Analysis Method	Feedback Method
[167]	· Shorinji Kempo	· Set of cameras with Reflective Markers (Non-Specified)	· Motion: Succession of 3D Skeletons	· None	· Visual (Screen)
[158]	· Seni Silat	· Depth Camera (Kinect)	· Motion: Succession of 3D Skeletons	· None	· Visual (Screen)
[168]	· Krav Maga	· Set of cameras with reflective markers (Vicon)	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical	· None
[157]	· Seni Silat	· Depth Camera (Kinect)	· Motion: Succession of 3D Skeletons	· Mathematical and/or Statistical	· None

APPENDIX II – CHARACTERISTICS OF REVIEWED PUBLICATIONS ABOUT MARTIAL ARTS

Below, a table with information about the reviewed publications regarding applications or the study of martial arts is shown. In the first column, the reference of the publication is presented based in the numeration of Section 7. The columns are described as follow:

- The first column indicates if there is interaction between users, or if users are studied in isolation.
- The second indicates if an external dataset is being used, and which dataset is.
- The third column indicates the number of users using when capturing the motion.
- The fourth column indicates the number of users used to test the system and obtain results. If it is the same group as the group in the previous column, it has been indicated.
- The fifth column indicates if the publication involves the use of AR or VR techniques.
- The sixth column indicates if the application is a real-time application or not.
- The seventh column indicates if weapons are being used in the study.

Table 20. A set of characteristics of the study carried out in each document.

Paper	Interaction between subjects	Using External Dataset	Number of users for motion capture	Number of users for results	AR/VR	Real Time	Use of weapons
[129]	· No	· No	· 24	· 24 (Same group)	· No	· No	· No
[155]	· Yes (2 Subjects)	· No	· 6	· 24 (Same group)	· No	· No	· Yes
[55]	· No	· No	· 7	· 7 (Same group)	· No	· No	· No
[52]	· Yes (2 Subjects)	· No	· 2	· 2 (Same group)	· No	· No	· Yes
[107]	· No	· No	· 1	· 1 (Same group)	· No	· No	· No
[141]	· No	· No	· 74	· 1	· VR (CAVE)	· Yes	· No
[83]	· No	· No	· 1	· 1 (Same group)	· No	· Yes	· No
[165]	· No	· No	· 10	· 10 (Same group)	· No	· No	· No
[156]	· Yes (2 subjects)	· No	· 2	· 2 (Same group)	· No	· No	· No
[84]	· No	· No	· 16	· 16 (Same group)	· No	· No	· No
[138]	· No	· No	· 2	· 2 (Same group)	· VR (HMD)	· Yes	· No
[163]	· No	· [169], [170]	· None	· 10	· VR (HMD)	· Yes	· No
[121]	· No	· No	· 2	· 2 (Same group)	· No	· No	· No
[161]	· No	· [171]	· None	· 5	· No	· Yes	· No
[139]	· No	· No	· 30	· 18	· VR (CAVE, HMD)	· Yes	· No
[162]	· No	· No	· Non specified	· Non specified	· VR (HMD)	· Yes	· No
[160]	· No	· No	· 3	4	· VR (CAVE, HMD)	· Yes	· No
[130]	· No	· No	· 26	· 26 (Same group)	· No	· No	· No
[159]	· No	· No	· Non specified	· 46	· AR (Screen)	· Yes	· No
[106]	· No	· No	· 5 (2 martial artists)	· 5 (Same group)	· No	· No	· No
[53]	· No	· Non specified	· Non specified	· 60	· AR (HMD)	· Yes	· No
[82]	· No	· No	· 14	· 14 (Same group)	· No	· No	· No
[166]	· No	· No	· 4	· 4 (Same group)	· No	· No	· No
[167]	· No	· No	· 1	· 6	· No	· No	· No
[158]	· No	· No	· None	· None	· No	· Yes	· No
[168]	· No	· No	· 16	· 16 (Same group)	· No	· No	· No
[157]	· No	· No	· 4	· 4 (Same group)	· No	· Yes	· No

APPENDIX III – MOTION CAPTURE PARTICIPANTS' INFORMATION

This research and the motion capture process has followed the guidelines of the Ethics Committee to carry out the study, as well as the indications of the Master. When the motion data was captured, the participants were requested to fulfill an anonym questionnaire. The information gathered from this questionnaire has been useful for determining some characteristics of the dataset such as number of participants that practices martial arts, age of participants or number of participants that practices Kenpo Karate. The questionnaire fulfilled can be seen below.

Cuestionario sobre actividad física y artes marciales

A continuación, realizaremos una serie de preguntas sobre la actividad física que realiza. Estas preguntas se realizan con el fin de recabar información sobre los participantes y finalmente poder comparar la ejecución de los movimientos en función de diferentes factores como pueden ser las actividades físicas practicadas, el tiempo durante el cual no se han practicado o la edad.

INFORMACIÓN PERSONAL

1.1. En primer lugar, vamos a recopilar algunos datos personales. El tratamiento de estos viene descrito en el documento adjunto a esta encuesta. Los datos identificativos como la ID no serán compartidos en ningún momento, siendo eliminados cuando terminen los experimentos, anonimizando así los datos.

Profesión:			
Edad:		ID (No rellenar):	
Género:		Fecha:	

ACTIVIDAD FÍSICA

2.1. ¿Practica usted algún arte marcial? De ser así indique a continuación el nombre del arte marcial (o las artes marciales) que practica, así como las horas de entrenamiento semanales, tiempo que lleva practicándolo, el cinturón alcanzado, y, si procede, el número de años que lleva sin practicar el arte marcial. Puede poner hasta un máximo de 3 artes marciales.

Nombre del arte marcial	Días de entrenamiento a la semana	Horas de entrenamiento diario	Tiempo practicando actividad física (años)	Nivel alcanzado (cinturón)	Años sin practicar

2.2. ¿Practica usted alguna otra actividad física? De ser así, indique a continuación el nombre de la actividad, junto a las horas de entrenamiento semanales, el tiempo que lleva practicándolo y, si procede, el número de años que lleva sin practicar esa actividad. Puede poner hasta un máximo de 3 actividades físicas.

Nombre de la Actividad Física	Días de entrenamiento a la semana	Horas de entrenamiento diario	Tiempo practicando actividad física (años)	Años sin practicar

TRABAJO

3.1. Algunos trabajos implican la realización de una actividad física intensa. En el caso de que su trabajo implique una actividad física intensa y no se corresponda con las actividades indicadas en los apartados anteriores, por favor, indique el número de horas semanales que dedica en su trabajo a realizar una actividad física intensa, así como el tiempo que lleva practicándolo. En el caso de no realizarlo actualmente, indique también cuanto tiempo lleva sin practicarlo.

Nombre de la Actividad Física*	Días de trabajo a la semana	Horas que dedica a la actividad en un día típico	Tiempo en el trabajo (años)	Años sin practicar

* En nombre de la actividad física indicar en que consiste esta actividad (Levantar peso, montar en bicicleta, correr, trabajos de construcción, cavar…).

SEDENTARISMO

4.1. A continuación, indique el número de horas diarias que dedica en un día típico a estar sentado, tumbado o recostado; ya sea en el trabajo, tiempo de ocio o de desplazamiento. (No incluya el tiempo que pasa durmiendo).

Horas de sedentarismo en un día típico

In this questionnaire, some questions regarding their age, gender, martial arts practiced, sports practiced and jobs that implies physical activity were performed. Finally, just information regarding gender, martial arts and age was used.

For storing the information gathered from the participants, a MySQL relational database was created. The main table of the database is called Participant and contains information about the user such as age, gender, hand or hours of sedentary lifestyle in a day. This table also has two id values: id_participant, which identifies a user uniquely in the table, and id_current_job, which identifies the current job of the user, to see if it involves physical activity.

The participants were asked if they were practicing any martial art in the moment of the questionnaire or in the past, so they could fill the answer with multiple options. A table called Martial_Art has been created for storing the possible martial arts that a participant can practice and the countries of origin of the martial arts, and a table called Practice_Martial_Art has been created for associating a participant with the martial arts that he/she has practiced. This table contains information about the number of years that the participant has been practicing a martial art, the number of days per week, the number of hours per day, the number of years inactive, and the highest level reached.

Furthermore, the participants were asked if they were practicing any sport in the moment of the questionnaire or in the past, again, they could answer the question with multiple options. The structure of the tables regarding sports is similar to the structure of the tables used for martial arts. A table named Sport was created for storing the possible sports that a participant can practice, and a table called Practice_Sport was created for associating a participant with a sport, including information such as how many years have been them practicing the sport, how many days per week, how many hours a day and how many years have been them inactive.

The participants were also asked if they had a job that involved any physical activity. The structure of the tables is again similar to the structure of the sport and martial arts tables. A table called Job was created for storing the possible jobs that a participant can have, and a table called Practice_Job was created for associating participants with jobs, including information about how many years have been them in that job, how many days per week, how many hours a day and how many years have been them inactive.

Finally, a table named Experiment was created for storing the paths of the files containing the information about the motion captured for the participants from both hands.

The structure of the database can be seen in Figure 70.

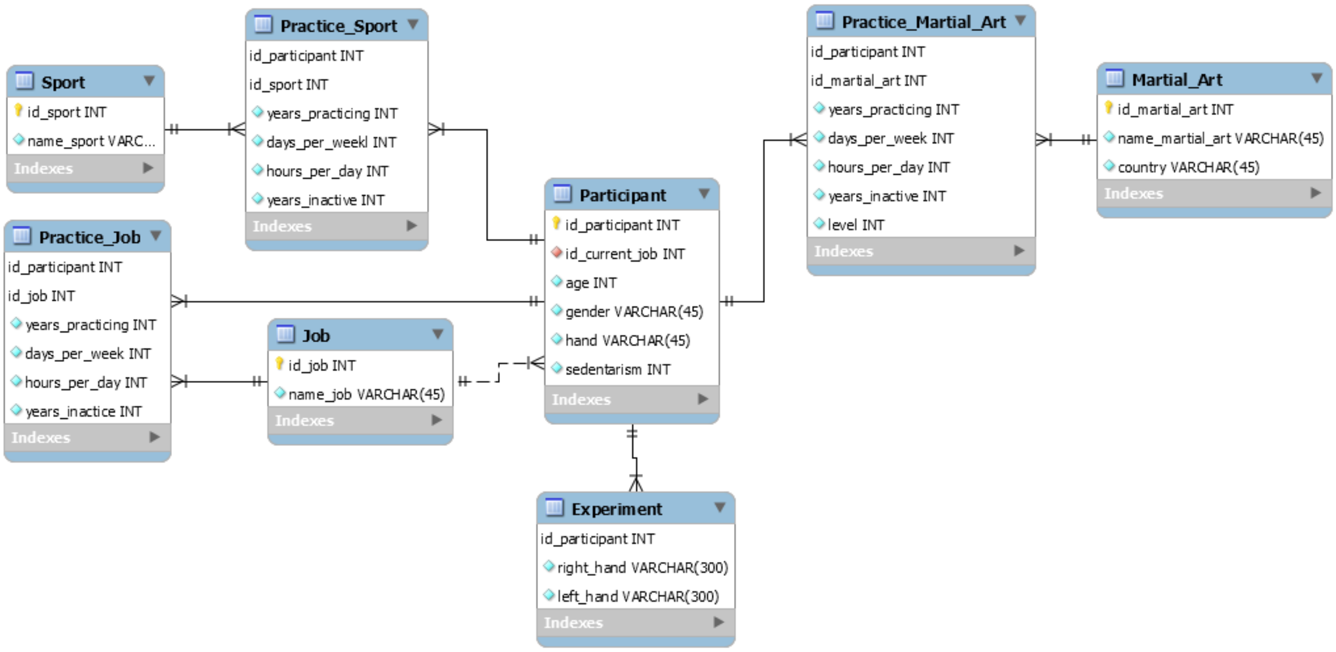

Figure 70. Structure of the database for storing users' information.

The participants were asked as well to sign a consent. This consent indicates that the user has been informed about the research purposes and allows the researcher to use the captured data in the project. The documents with the information given to the participants and the document that they had to sign are shown in the next pages.

HOJA DE INFORMACIÓN SOBRE EL PROYECTO DE INVESTIGACIÓN Y/O EXPERIMENTACIÓN

Título del Proyecto: Trabajo Fin de Máster 'Análisis y Modelado y de la actividad psicomotora en Artes Marciales' (título provisional)
Investigador/a Principal: Olga C. Santos (Directora del TFM)
Grupo de Investigación: Alberto Casas Ortiz (Estudiante del TFM)
Máster Universitario en Inteligencia Artificial Avanzada
Facultad/Escuela: Escuela Técnica Superior de Ingenieros Informáticos (ETSI Informática)

La legislación vigente establece que la participación de toda persona en un proyecto de investigación y/o experimentación requerirá una previa y suficiente información sobre el mismo y la prestación del consentimiento por parte de los sujetos que participen en dicha investigación/experimentación. A tal efecto, a continuación, se detallan los objetivos y características del proyecto de investigación arriba referenciado, como requisito previo a la prestación del consentimiento y a su colaboración voluntaria en el mismo:

1) **OBJETIVOS:** Estudiar la forma más adecuada de describir los movimientos que se realizan en las artes marciales, para posteriormente desarrollar sistemas inteligentes que ayuden tanto a mejorar la técnica como a prevenir lesiones durante su práctica.

2) **DESCRIPCIÓN DEL ESTUDIO:** Primeramente, se pedirá al participante que cumplimente un cuestionario con información personal (e.g., demográfica, parámetros físicos, experiencia con artes marciales, actividad física…). Después, uno de los investigadores explicará los ejercicios que se van a realizar. En concreto, se propondrá la realización de movimientos sencillos propios de Kenpo Karate, un arte marcial que incluye bloqueos, golpes o desplazamientos. Una vez los participantes hayan aclarado sus dudas, se realizará una serie de ejercicios de calentamiento para prevenir lesiones. A continuación, se dejará a los participantes practicar los movimientos indicados y se procederá a su captura. Estos movimientos serán capturados a través de una serie de dispositivos como pueden ser cámaras o sensores colocados sobre determinadas partes del cuerpo. La recogida de datos no durará más de 15 minutos por participante en cada sesión.

3) **POSIBLES BENEFICIOS**: Se espera que el participante obtenga una mejor comprensión sobre la forma en que debe realizar determinados movimientos característicos de Kenpo Karate. Además, el objetivo final de la investigación es desarrollar un sistema que facilite el aprendizaje de los movimientos y evite lesiones al realizarlos.

4) **POSIBLES INCOMODIDADES Y/O RIESGOS DERIVADOS DEL ESTUDIO:** Es posible que el uso de los sensores pueda suponer una ligera incomodidad en la realización de los movimientos. No se considera que pueda haber riesgo para la salud ya que los movimientos propuestos son básicos y comunes en las artes marciales, y los participantes del estudio deberán tener cierta experiencia en su práctica. En todo caso, se procederá a realizar un calentamiento previo para evitar lesiones. Además, los participantes deberán tener un seguro médico en vigor que cubra la práctica de artes marciales.

5) **PREGUNTAS E INFORMACIÓN**: Puede contactar con la investigadora y directora del TFM, Olga Santos Martín (ocsantos@dia.uned.es) o con Alberto Casas Ortiz (albertocasasortiz@gmail.com), estudiante que está realizando este estudio en el contexto de su TFM. Se le entrega esta hoja informativa y un consentimiento que indica que está de acuerdo con su participación en el mismo, el cual deberá firmar y entregar antes de empezar el estudio.

6) **PROTECCIÓN DE DATOS**: Este proyecto requiere la utilización y manejo de datos de carácter personal que, en todo caso, serán tratados conforme a las normas aplicables garantizando la confidencialidad de los mismos. Cada participante tendrá asociado un código utilizado para identificarle en el estudio, de manera que los datos queden anonimizados. Existirá un listado interno con la relación entre cada participante y su código, accesible solo por los investigadores del proyecto, con uso exclusivo para este estudio y que también será destruido cuando finalice la recopilación y análisis de datos. Los datos serán tratados de forma anónima y podrán ser utilizados para la realización del TFM y otros estudios relacionados, así como en la redacción de artículos científicos derivados de la investigación realizada. Las grabaciones en video o fotos (si las hubiera) se utilizarán para el análisis por parte de los investigadores y serán destruidas cuando finalice la recopilación y análisis de los datos, salvo que expresamente el participante autorice a utilizarlas para divulgación científica en presentaciones y artículos mostrando su conformidad con el uso de su imagen en la casilla correspondiente del Consentimiento Informado que firme el participante.

La participación de este proyecto de investigación es voluntaria y puede retirarse del mismo en cualquier momento.

Y para que conste por escrito a efectos de información de los participantes a los que se solicita su participación voluntaria en el proyecto antes mencionado, se ha formulado y se entrega la presente hoja informativa.

En .... a ...... de.................................de...............

Nombre y firma del Investigador principal

Fdo: Olga Santos Martín

CONSENTIMIENTO INFORMADO

D./Dª...

He leído la hoja de información que se me ha entregado anexa a este documento, y la he comprendido en todos sus términos.

He sido suficientemente informado y he podido hacer preguntas sobre los objetivos y metodología aplicada en el proyecto 'Modelado y análisis dela actividad psicomotora en Artes Marciales' que se desarrolla como Trabajo Fin de Máster del Máster Universitario en Inteligencia Artificial Avanzada de la ETS de Ingeniería Informática de la UNED, y para el que se ha pedido mi colaboración.

A continuación indico si autorizo o no el uso con fines de divulgación científica (por ejemplo, para ilustrar presentaciones o artículos en revistas que reporten la investigación realizada) las imágenes tomadas durante las sesiones de recogida de datos para la realización del Trabajo Fin de Máster anteriormente mencionado, así como en investigaciones asociadas en las que participen dichos investigadores:

AUTORIZO: Sí ☐ No ☐

Comprendo que mi participación es voluntaria y que puedo retirarme del estudio,

- cuando quiera;
- sin tener que dar explicaciones y exponer mis motivos; y
- sin ningún tipo de repercusión negativa para mí.

Por todo lo cual, PRESTO MI CONSENTIMIENTO para participar en el proyecto de investigación asociado a la realización de un Trabajo Fin de Máster antes citado y para que mis datos de carácter personal sean tratados, según la normativa vigente y la política de protección de datos de la UNED, para el uso exclusivo en este proyecto.

En .... a .... de .... de 20..........

Fdo. ....

APPENDIX IV – QUESTIONNAIRE ABOUT ADVANTAGES AND DISADVANTAGES OF ONLINE TRAINING

In this appendix, the results of the questionnaire carried out about psychomotor online learning are shown. It has two parts, one focused on teachers and one focused on learners. First, the results from the point of view of the teachers are analyzed, and then, the results from the point of view of the students.

The questionnaire was designed to be totally anonymous and voluntary. Students and teachers were asked some questions regarding their age, the psychomotor activity they learn/teach online, and about their experience teaching/learning online those psychomotor activities. Then, in order to land the problem to the participants with some specific examples, they are put in a set of advantageous and disadvantageous situations that could potentially arise from our experience to see if they have encountered with them. Before and after the questions regarding the different situations, they have a free question with the purpose to get their comments, specifically after reading the example situations, if they have encountered with those situations or with new situations that we have not considered at the beginning of the questionnaire but that they have recalled after reading the sample situations proposed. Finally, a set of questions regarding their psychological status during online psychomotor learning are asked, to analyze if the emotional state can have some influence in psychomotor learning.

Before submitting the questionnaire and to comply with data protection, participants had to confirm that they allow the researchers to use their answers (which are in any case anonymously collected) in this Master's Thesis and for further investigation.

IV.I Point of View of the Teacher

The number of teachers that participated in the questionnaire were 2 (n=2), and their answers are shown below. Since we only have answers from two teachers, the results are not representative and generalizable for more teachers, but we are going to contrast the teachers' results with the results obtained in the students' part of the questionnaire, so we can forecast if the results obtained in the teachers' part of the questionnaire about their students are accurate.

IV.I.I Information About the Teachers

In this subsection it is shown the information about the teachers' answers, including age, genre and psychomotor activity that they teach.

In the graphics it can be seen how both teachers are woman. One of them teaches dance, and the other teaches to play a musical instrument. One of the teachers is between 25 and 34 years old, and the other between 35 and 44. The mean of activities taught by a teacher is 1, which coincides with the mode. The standard deviation is 0 since both teachers teach 1 activity. One of the teachers only teach through interactive class, but the other uses recorded videos as well.

The questions and the answers of the participants are shown in the next pages.

1.1. What is your age?

Table 21. Age range of the teachers.

Age range	Absolute Frequency
<18	0
18-24	0
25-34	1
35-44	1
45-54	0
>65	0
Total	**2**

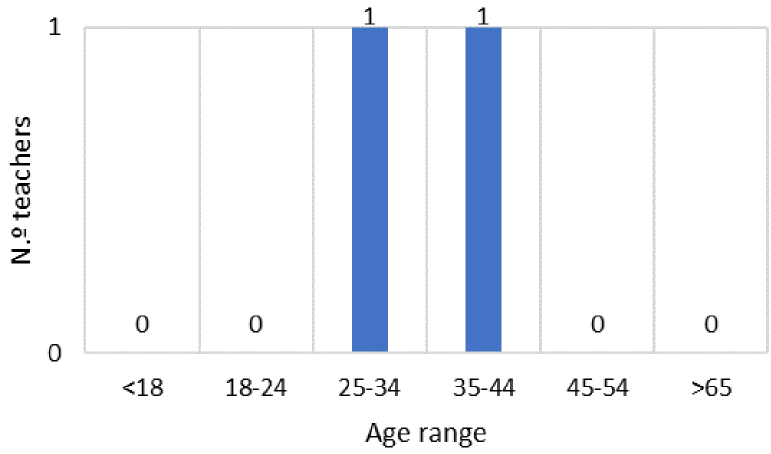

Figure 71. Age range of the teachers.

1.2. What is your gender?

Table 22. Gender of the teachers.

Gender	Absolute Frequency
Woman	2
Man	0
Total	**2**

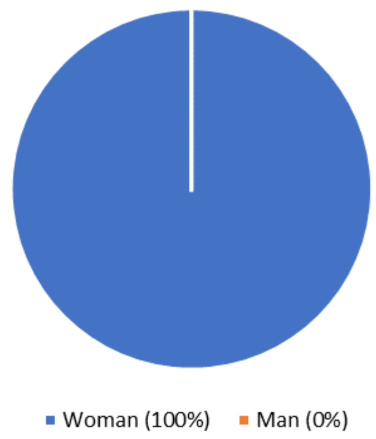

Figure 72. Gender of the teachers.

1.3. In the current quarantine situation, are you performing any online psychomotor activity?

Table 23. Activities taught by the teachers.

Activity	Absolute Frequency
Dance	1
Play musical instruments	1
Physical training	0
Martial Arts	0
Yoga	0
Performance	0
Drawing	0
Total	**2**

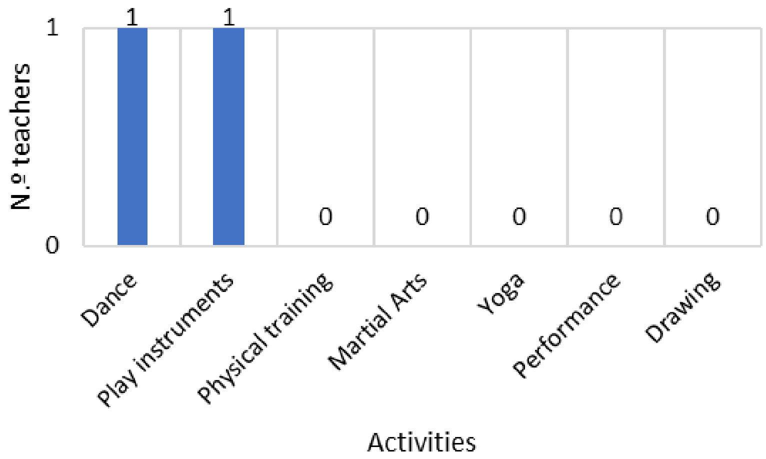

Figure 73. Activities taught by the teachers.

Table 24. Statistics about number of activities taught by teachers.

	Nº Activities taught per teacher
Mean	1
Mode	1
Standard Deviation	0

1.4. Which kind of media do you use to perform the psychomotor activity?

Table 25. Kind of platforms used by teachers.

Kind of Media used to teach	Absolute Frequency
Interactive classes	1
Recorded Videos	0
Both	1
Total	**2**

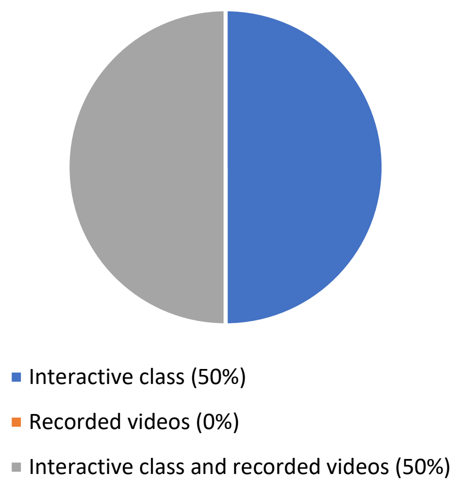

Figure 74. Kind of platforms used by teachers.

IV.I.II How the Teachers Teach Psychomotor Activities Online

In this subsection it is shown the information about which tools the teachers use for teaching online the psychomotor activities, their experience with those tools and the characteristics of the classes.

Both teachers use Zoom. One of them uses YouTube, and the other uses Jitsi and Skype as well. The mean of platforms used per teacher is 2, with a standard deviation of 1,414213562. The mode cannot be calculated because there is not a more frequent value.

Both teachers use a laptop, and one of them use a smartphone as well. The mean of devices used per teacher is 1.5, and the standard deviation is 0,7071067812. Again, the mode cannot be calculated because there is not a more frequent value.

The students of one of the teachers attend to classes alone, while the students of the other teacher attend to classes with somebody. Both teachers teach alone and have been teaching online between two and three months (since the quarantine started in Spain). One of them have taught less than 10 sessions online, but the other has taught between 30 and 39.

The number of students that use to attend the classes is less than 10 for one of the teachers and between 10 and 19 for the other. One of the teachers claims to have an advanced level when using tools for online teaching, the other claims to have a medium level. In any case, both have installed their applications to teach online.

This section included two free question where the teachers could answer anything. The first question was: “Which advantages have you found when training psychomotor activities online?”. One of the teachers answered, “My students can assist classes when it suits them”, and the other responded “Punctuality and comfort”. The second question was: “Which disadvantages have you found when training psychomotor activities online?”. One of the teachers responded: “I can’t see my students execute the movements or asking me for help.”, and the other responded “Some things are difficult to teach without a direct interaction with the student.”

The questions and the answers of the participants are shown in the next pages.

1.1. Which platforms do you use/have used to teach online?

Table 26. Platforms used by teachers.

Platform	Absolute Frequency
YouTube	1
Zoom	2
Skype	1
Instagram Direct	0
WhatsApp	0
Facebook	0
Jitsi	1
Hangouts	0
RTVE A la carta	0
Total	**5**

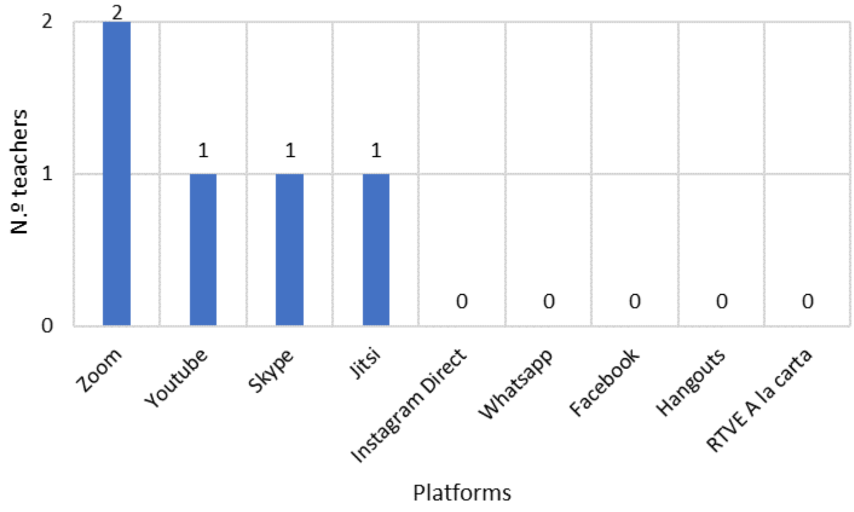

Figure 75. Platforms used by teachers.

Table 27. Statistics about the number of platforms used by teachers.

	Nº platforms used by teacher
Mean	2
Mode	Null
Standard Deviation	1,414213562

2.2. Which devices do you use in your online classes?

Table 28. Kind of devices used by teachers.

Device	Absolute Frequency
Laptop	2
Smartphone	1
Smart TV	0
Tablet	0
Desktop computer	0
Total	**3**

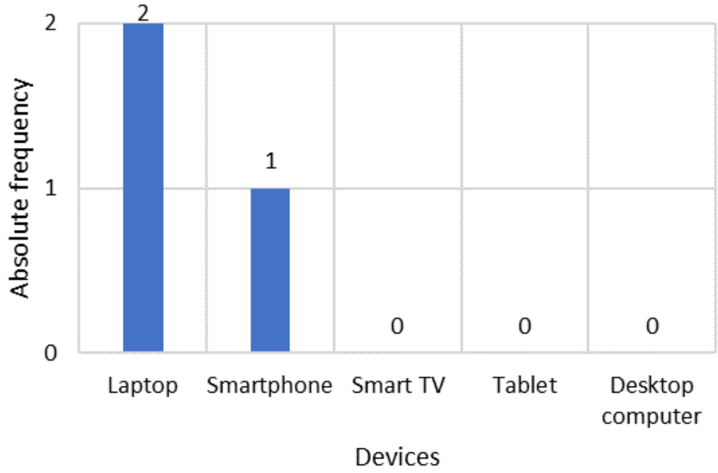

Figure 76. Kind of devices used by teachers.

Table 29. Statistics about the number of devices used by teachers.

	Nº devices used by teacher
Mean	1.5
Mode	Null
Standard Deviation	0,7071067812

2.3. Do you teach with somebody's help? (E.g. your couple, roommate, relative…)

Table 30. How the teachers teach

Do you teach with somebody's help?	Absolute frequency
I teach alone	2
I teach with somebody	0
Total	**2**

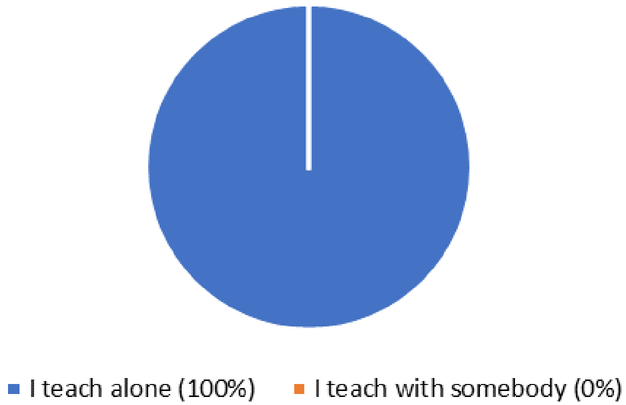

Figure 77. How the teachers teach.

2.4. Do your students attend your classes with somebody? (E.g. their couple, roommate, relative…)

Table 31. How students attend the online classes.

Do your students attend alone?	Absolute frequency
They attend with somebody	1
They attend alone	1
Total	**2**

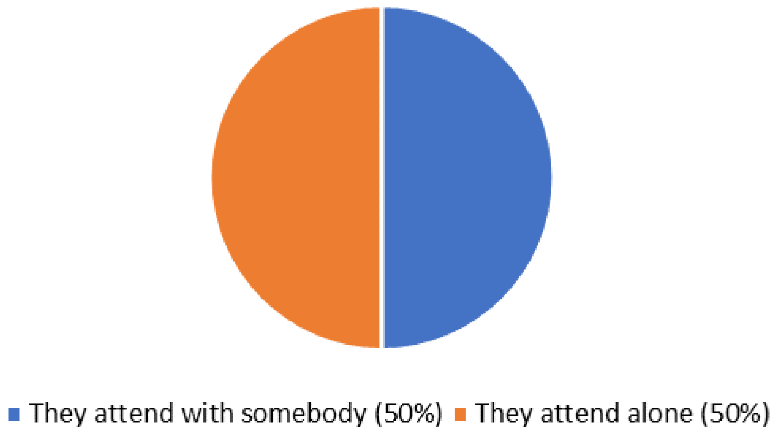

Figure 78. How students attend the online classes.

2.5. How many time have you been teaching online?

Table 32. Time that the teachers have been teaching online.

How many time have you been teaching online?	Absolute frequency
< 2 weeks	0
>= 2 weeks < 1 month	0
>= 1 month < 2 month	0
>= 2 month < 3 month	2
>= 3 month < 4 month	0
>= 4 months	0
Total	**2**

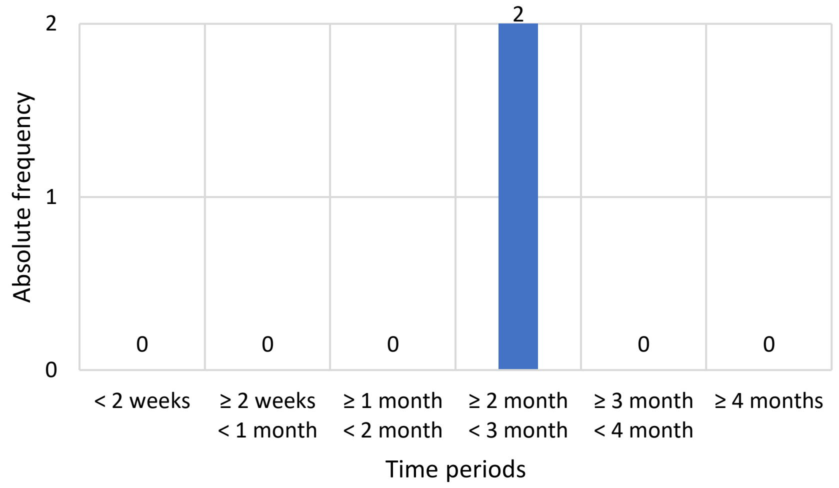

Figure 79. Time that the teachers have been teaching online.

2.6. How many sessions do you estimate that you have imparted in that time?

Table 33. Number of sessions taught by teachers.

Nº sessions taught	Absolute frequency
< 10	1
10 - 19	0
20- 29	0
30-39	1
40-49	0
50-59	0
60-69	0
70-79	0
80-89	0
90-99	0
>100	0
Total	**2**

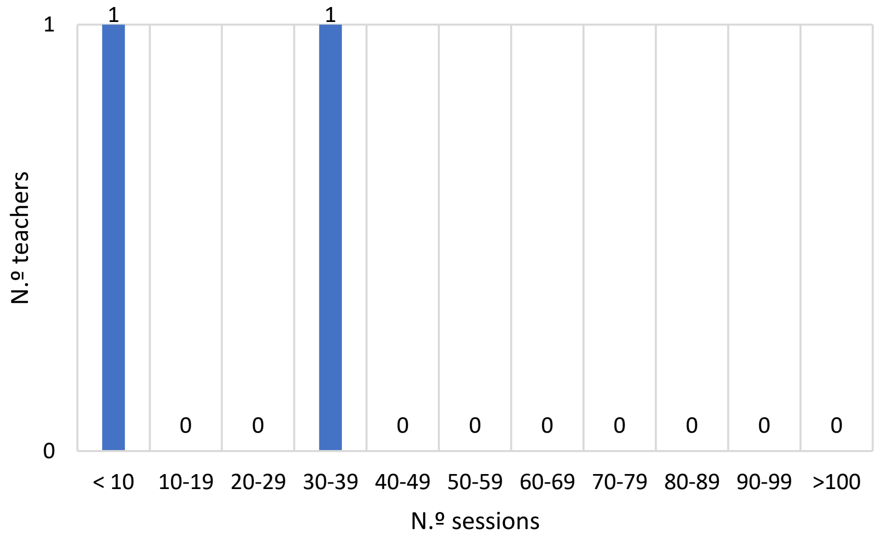

Figure 80. Number of sessions taught by teachers.

2.7. How many students use to attend your online classes?

Table 34. Number of students that attend the online classes.

Nº students per class	Absolute frequency
< 10	1
10-19	1
20-29	0
30-39	0
40-49	0
>=50	0
Total	**2**

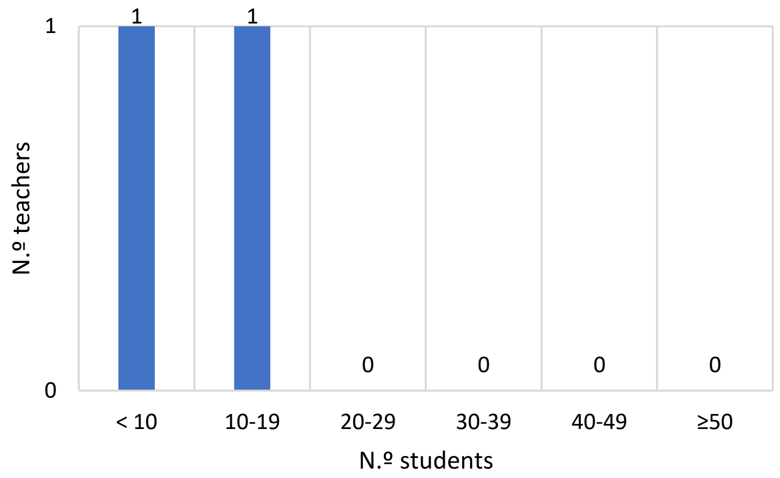

Figure 81. Number of students that attend the online classes.

2.8. Which is your level using applications and technologies of online learning?

Table 35. Level of the teachers when using online teaching tools.

Level using online teaching tools	Absolute frequency
Basic	0
Medium	1
Advanced	1
Total	**2**

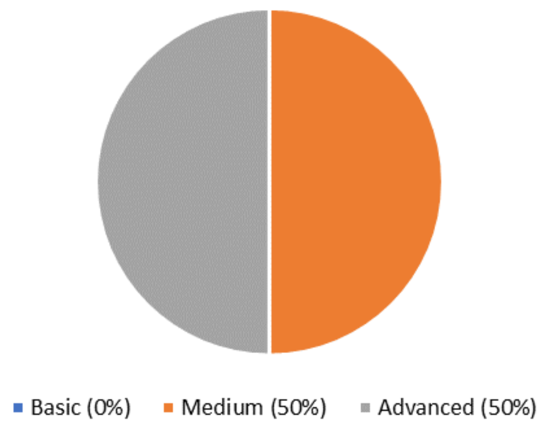

Figure 82. Level of the teachers when using online teaching tools.

2.9. Have you installed the necessary tools/apps by yourself?

Table 36. Information about how the teachers have installed the online learning tools.

Have you installed the necessary applications for teaching online?	Absolute frequency
I installed it	2
I needed help	0
Total	**2**

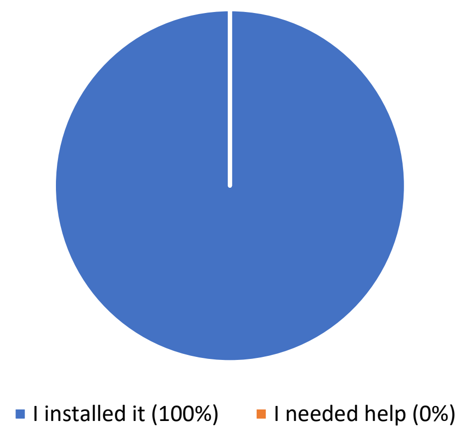

Figure 83. Information about how the teachers have installed the online tools.

2.10. Which advantages have you found when teaching psychomotor activities online? (Answers translated from Spanish)

“Mis alumnos pueden tomar la clase cuando les venga bien.” – “My students can assist classes when it suits them.”

“Puntualidad y comodidad.” – “Punctuality and comfort.”

2.11. Which disadvantages have you found when teaching psychomotor activities online? (Answers translated from Spanish)

“No puedo ver a mis alumnos ejecutar los movimientos o pedirme ayuda.” – I can’t see my students execute the movements or asking me for help.”

“Algunas cosas son difíciles de enseñar sin una interacción directa con el estudiante.” – “Some things are difficult to teach without a direct interaction with the student.”

IV.I.III Advantageous Situations

In this subsection, the teachers were exposed to a set of advantageous situations that they might have encounter when teaching psychomotor activities online.

Both teachers consider advantages of online teaching the fact that i) the classes can be recorded so the student can retake the classes, ii) the students can send them a video executing the movements, so the teachers can follow the progress of the student, and iii) the classes can be taken at different hours than in-person classes and can even be extended as there is no another activity that requires access to the physical space in the classroom.

In turn, they disagree in three of the questions. One of the teachers considers that their students are more uninhibited and that they assist more assiduously in online classes, but the other teacher disagrees. Only one of them would continue with online classes in the future, when the quarantine situation ends.

At the end of this section of the questionnaire, there is a free question, so the teachers can indicate some advantages that come to their mind or they want to say something regarding their answers. Only one of the teachers answered this question and said that another advantage of online learning is that there is no need to go anywhere, so you save money on that.

3.1. Do you feel that your students are more uninhibited or outgoing since they are more comfortable in their environment, and they can perform without drawing the attention of their partners?

Table 37. Answers to question 3.1 in teachers' questionnaire.

Answer	Absolute frequency
Yes	1
No	1
Total	**2**

Figure 84. Answers to question 3.1 in teachers' questionnaire.

3.2. Do you think that the fact that your students can record the classes (or you can distribute them), so they can repeat and review them is an advantage of online learning?

Table 38. Answers to question 3.2 in teachers' questionnaire.

Answer	Absolute frequency
Yes	2
No	0
Total	**2**

Figure 85. Answers to question 3.2 in teachers' questionnaire.

3.3. Do you consider that the fact that your students can send you their progress in video, so you can review them more carefully and give them a better feedback, is and advantage of online learning?

Table 39. Answers to question 3.3 in teachers' questionnaire.

Answer	Absolute frequency
Yes	2
No	0
Total	**2**

Figure 86. Answers to question 3.3 in teachers' questionnaire.

3.4. Do you think that your students assist more assiduously to online classes, since they don't have to leave their environment?

Table 40. Answers to question 3.4 in teachers' questionnaire.

Answer	Absolute frequency
Yes	1
No	1
Total	**2**

Figure 87. Answers to question 3.4 in teachers' questionnaire.

3.5. Do you consider that the fact that you can teach online at different hours than on in-person teaching, extending the classes since you don't depend on physical installations, is an advantage?

Table 41. Answers to question 3.5 in teachers' questionnaire.

Answer	Absolute frequency
Yes	2
No	0
Total	**2**

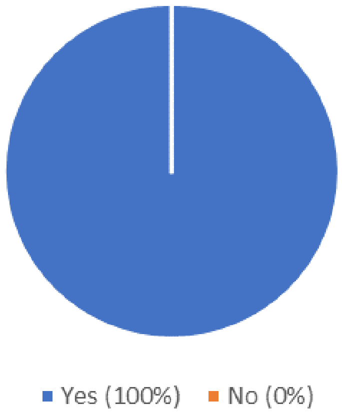

Figure 88. Answers to question 3.5 in teachers' questionnaire.

3.6. Will you continue teaching online in the future, even if the quarantine situation ends?

Table 42. Answers to question 3.6 in teachers' questionnaire.

Answer	Absolute frequency
Yes	1
No	1
Total	**2**

Figure 89. Answers to question 3.6 in teachers' questionnaire.

3.7. Is there any other advantage that comes to your mind or do you want to add something about your answers? (Translated from Spanish)

"Ventaja sobre todo en que no hay desplazamientos y el gasto que conlleva." – "A great advantage is that you don't have to move, which could suppose a great spending."

IV.I.IV Disadvantageous Situations

In this subsection, the teachers were exposed to a set of disadvantageous situations that they might have encountered when teaching psychomotor activities online.

Both teachers found a problem in the delay of the connection and think that the lack of communication is a disadvantage, but neither of them think that their students are training/practicing less in online classes than on in-person classes. One of them thinks that the fact that they cannot observe properly the movements of her/his students in online learning is a problem, that it is more difficult to perceive the progress of the student and that they are less participative in online classes than on in-person classes. The other teacher disagrees.

At the end of this section of the questionnaire, there is a free question, so the teachers can indicate some disadvantages that come to their mind or they want to say something regarding their answers. The teacher that teaches music said that the delay in the sound is a problem because they cannot practice with multiple instruments, and the teacher that uses YouTube said that the fact that the students need an account to interact with her is a problem.

4.1. Is the delay in your connection a problem? E.g. There is delay between the moment you explain something, and your explanation reach your students.

Table 43. Answers to question 4.1 in teachers' questionnaire.

Answer	Absolute frequency
Yes, the delay is a problem	2
No, the delay is not a problem	0
There is no delay	0
Total	**2**

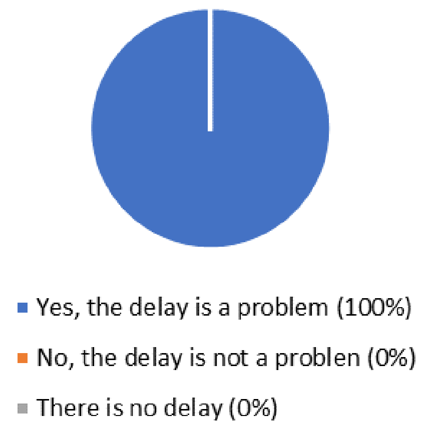

Figure 90. Answers to question 4.1 in teachers' questionnaire.

4.2. The lack of communication supposes any problem when teaching online? E.g. Your students cannot talk with you, only can talk through text or is it difficult to correct the movements of the student.

Table 44. Answers to question 4.2 in teachers' questionnaire.

Answer	Absolute frequency
Yes	2
No	0
Total	**2**

Figure 91. Answers to question 4.2 in teachers' questionnaire.

4.3. The fact that you cannot observe properly your students' movements suppose any problem when teaching online? E.g. You have to ask your students to move or turn to observe better their movements, or you cannot even observe the movements.

Table 45. Answers to question 4.3 in teachers' questionnaire.

Answer	Absolute frequency
Yes	1
No	1
Total	2

Figure 92. Answers to question 4.3 in teachers' questionnaire.

4.4. Do you feel that it is more difficult to notice the progress of your students in the online classes?

Table 46. Answers to question 4.4 in teachers' questionnaire.

Answer	Absolute frequency
Yes	1
No	1
Total	2

Figure 93. Answers to question 4.4 in teachers' questionnaire.

4.5. Do you feel that your students are les participative in online classes?

Table 47. Answers to question 4.5 in teachers' questionnaire.

Answer	Absolute frequency
Yes	1
No	1
Total	**2**

Figure 94. Answers to question 4.5 in teachers' questionnaire.

4.6. Do you feel that your students are not training/practicing enough in online classes as they do on in-person classes?

Table 48. Answers to question 4.6 in teachers' questionnaire.

Answer	Absolute frequency
Yes	0
No	2
Total	**2**

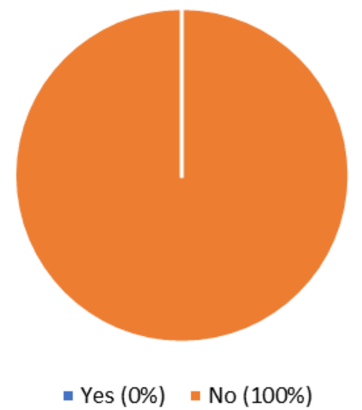

Figure 95. Answers to question 4.6 in teachers' questionnaire.

4.7. Is there any other disadvantage that comes to your mind, or do you want to add something about your answers? (Translated from Spanish)

"Al dar las clases por YouTube, necesito que mis alumnos tengan que crear una cuenta para poder interactuar conmigo." – "Since I teach using YouTube, I need my students to have an account to interact with me."

"El retardo en el sonido es un problema ya que si se quiere tocar con distintos instrumentos impide que se pueda hacer bien." – "The delay in the sound is a problem because if I want to play different instruments it cannot be done properly."

IV.I.V Emotional State During Classes

Finally, the teachers had to answer to a set of questions about their emotional state or how they feel during online classes, to probe if it affects them emotionally. They were also asked about the perception of the emotional state of their students.

Both teachers feel the same nervous in online and on in-person classes, and both said that they will not abandon online classes. However, one of them said that will not impart more online classes in the future. Both teachers feel their students motivated in online classes as they were on in-person classes, but they think that their students feel more satisfied after an in-person class, and that the fact that their students are not in person with the rest of students and with the teacher affect them negatively.

They disagree in some questions. One of the teachers considers that her students feel less nervous in online classes than on in-person classes, that they feel more motivated during in-person classes, and that they feel more bored in online classes. However, the other teacher considers that her students feel the same in both classes. One of the teachers feel more motivated during in-person classes, more bored during online classes, more satisfied after an in-person class, and the fact that she is not with her students affects her negatively, but the other teacher feels the same in both, online and in-person classes, in those cases.

Only one of the teachers answered the open question about her mood in the classes and she said: "The fact that I cannot see my students, their evolution, and encourage them is demotivating."

5.1. Do you feel less nervous during online classes than during in-person classes?

Table 49. Answers to question 5.1 in teachers' questionnaire.

Answer	Absolute frequency
I feel less nervous in online classes	0
I feel less nervous on in-person classes	0
I feel the same in both cases	2
Total	**2**

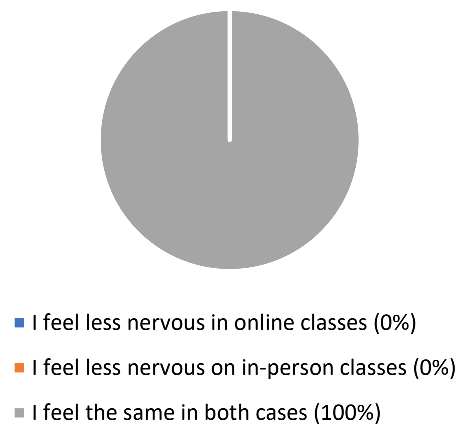

Figure 96. Answers to question 5.1 in teachers' questionnaire.

5.2. Do you feel your students less nervous during online classes than during in-person classes?

Table 50. Answers to question 5.2 in teachers' questionnaire.

Answer	Absolute frequency
I feel my students less nervous in online classes	1
I feel my students less nervous on in-person classes	0
I feel my students the same in both classes	1
Total	2

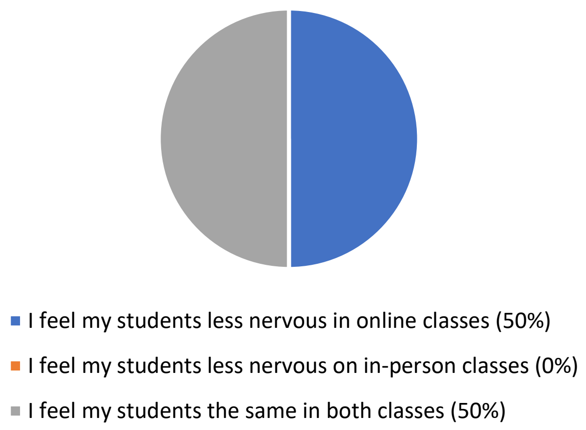

Figure 97. Answers to question 5.2 in teachers' questionnaire.

5.3. Do you feel more motivated during in-person classes than during online classes?

Table 51. Answers to question 5.3 in teachers' questionnaire.

Answer	Absolute frequency
I feel more motivated on in-person classes	1
I feel more motivated in online classes	0
I feel the same in both classes	1
Total	2

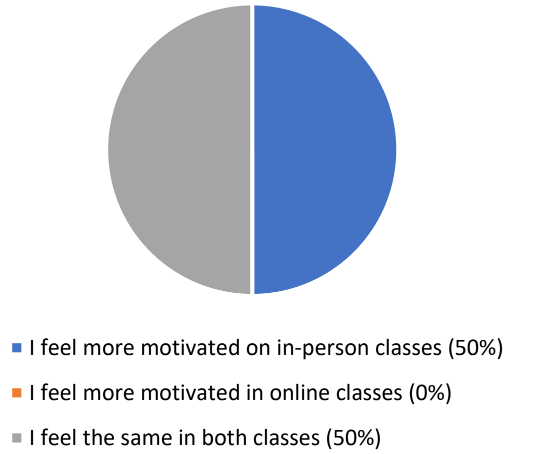

Figure 98. Answers to question 5.3 in teachers' questionnaire.

5.4. Do you feel your students more motivated during in-person classes than during online classes?

Table 52. Answers to question 5.4 in teachers' questionnaire.

Answer	Absolute frequency
I feel my students more motivated on in-person classes	0
I feel my students more motivated in online classes	0
I feel my students the same in both classes	2
Total	2

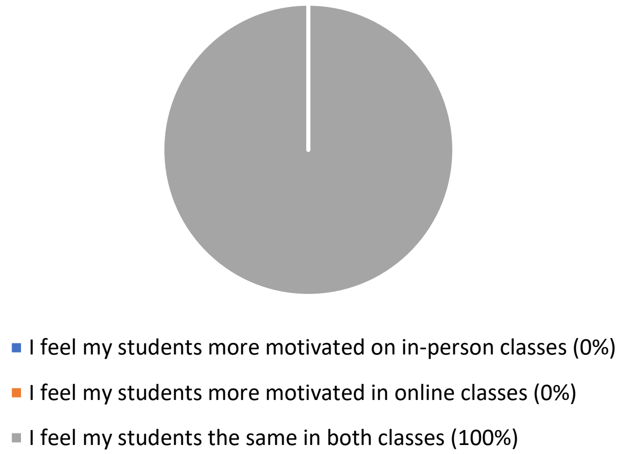

Figure 99. Answers to question 5.4 in teachers' questionnaire.

5.5. Do you feel more bored during online classes than during in-person classes?

Table 53. Answers to question 5.5 in teachers' questionnaire.

Answer	Absolute frequency
I feel more bored in online classes	1
I feel more bored on in-person classes	0
I feel the same in both classes	1
Total	2

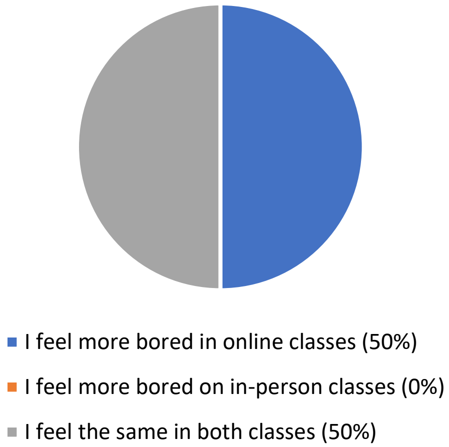

Figure 100. Answers to question 5.5 in teachers' questionnaire.

5.6. Do you feel your students more bored during online classes than during in-person classes?

Table 54. Answers to question 5.6 in teachers' questionnaire.

Answer	Absolute frequency
I feel my students more bored in online classes	1
I feel my students more bored on in-person classes	0
I feel my students the same in both classes	1
Total	2

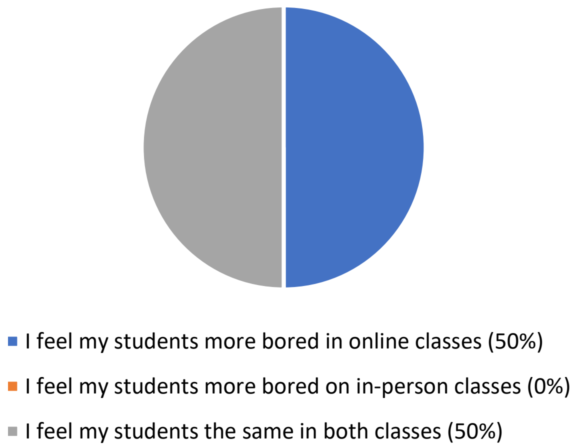

Figure 101. Answers to question 5.6 in teachers' questionnaire.

5.7. How affect you the fact that you are not with your students in person?

Table 55. Answers to question 5.7 in teachers' questionnaire.

Answer	Absolute frequency
It affects me positively	0
It affects me negatively	1
It does not affect me	1
Total	2

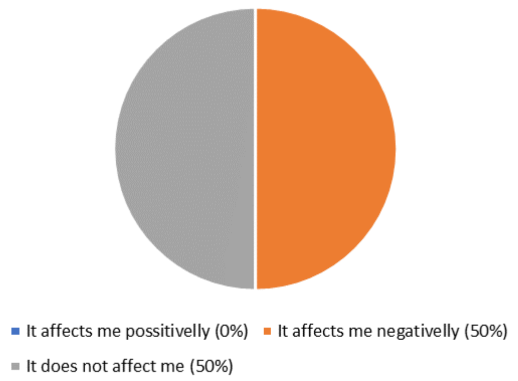

Figure 102. Answers to question 5.7 In teachers' questionnaire.

5.8. How do you think that affect your students the fact that they are not with you and their partners in person?

Table 56. Answers to question 5.8 in teachers' questionnaire.

Answer	Absolute frequency
It affects them positively	0
It affects them negatively	2
It does not affect them	0
Total	2

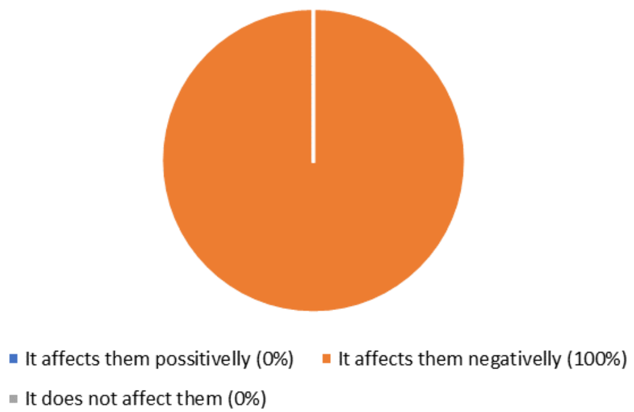

Figure 103. Answers to question 5.8 in teachers' questionnaire.

5.9. Do you feel more satisfied after an online class than after an in-person class?

Table 57. Answers to question 5.9 in teachers' questionnaire.

Answer	Absolute frequency
I feel more satisfied after in-person classes	1
I feel more satisfied after online classes	0
I feel the same after both classes	1
Total	2

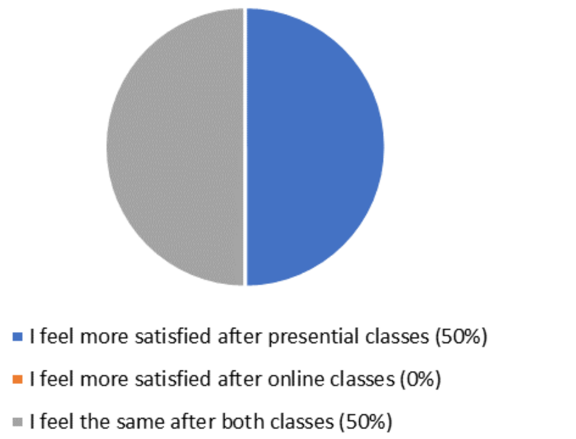

Figure 104. Answers to question 5.9 in teachers' questionnaire.

5.10. Do you feel your students more satisfied after an online class than after an in-person class?

Table 58. Answers to question 5.10 in teachers' questionnaire.

Answer	Absolute frequency
I feel my students more satisfied after in-person classes	2
I feel my students more satisfied after online classes	0
I feel my students the same after both classes	0
Total	2

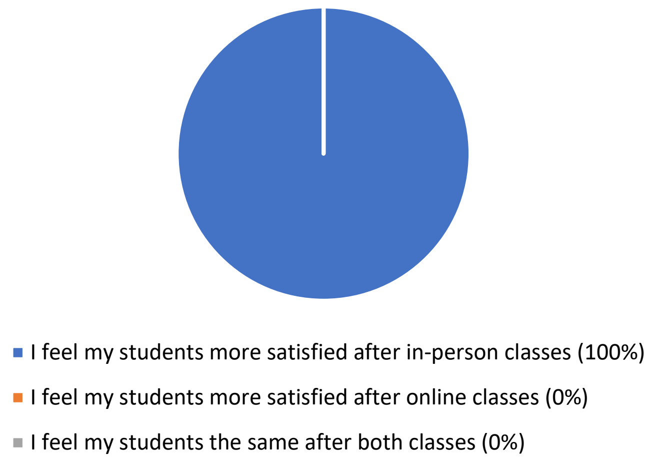

Figure 105. Answers to question 5.10 in teachers' questionnaire.

5.11. Have you though in imparting more online classes in the future?

Table 59. Answers to question 5.11 in teachers' questionnaire.

Answer	Absolute frequency
Yes	1
No	1
Total	2

Figure 106. Answers to question 5.11 in teachers' questionnaire.

5.12. Have you though in abandon online classes?

Table 60. Answers to question 5.12 in teachers' questionnaire.

Answer	Absolute frequency
Yes	0
No	2
Total	2

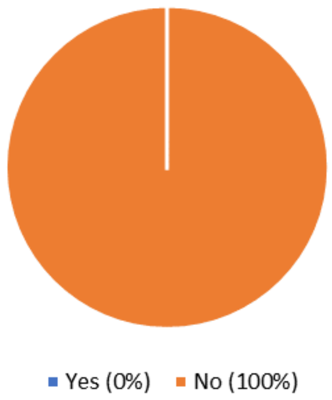

Figure 107. Answers to question 5.12 in teachers' questionnaire.

5.13. Is there anything about your mood that comes to your mind, or do you want to add something about your answers? (Translated from Spanish)

"Me desmotiva no poder ver a mis alumnos y no poder ver cómo evolucionan y darles ánimos y hacerles reír" – "The fact that I cannot see my students, their evolution, and encourage them is demotivating."

IV.II Point of View of the Student

The number of students that participated in the questionnaire were 27 (n=27), and their answers are shown below. Some of the results obtained here are complained in the last section, to forecast if the perception of the teachers about how their students feel are accurate.

IV.II.I Information About the Students

In this subsection it is shown the information about the students' responses, including age, genre and psychomotor activity that they are learning.

It can be seen how more than half of the students are women (59.3%), and the ages of the students are in the range 25-34 years old (16 of 27 students).

The most popular activities practiced by the students are Physical Training (13 of 46 apparitions), Dance (9 of 46 apparitions) and Martial Arts (8 of 46 apparitions). Most of the students learn two activities. The mean of activities learned per student is 1.7, with a standard deviation of 0.77. 48.1% of the students learn by using recorded videos,18.6% learn by attending interactive classes, and the remaining 33% learn using both.

The questions and the responses of the participants are shown in the next pages.

1.1. What is your age?

Table 61. Age range of the students.

Age range	Absolute Frequency
<18	0
18-24	8
25-34	16
35-44	0
45-54	2
>65	1
Total	**2**

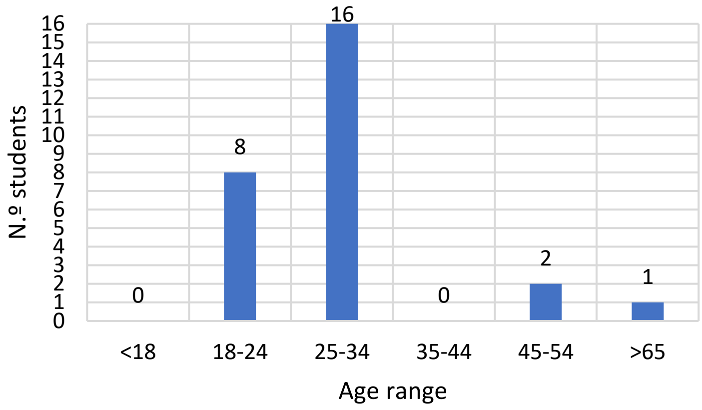

Figure 108. Age range of the students.

1.2. What is your gender?

Table 62. Gender of the students.

Gender	Absolute Frequency
Woman	16
Man	11
Total	**2**

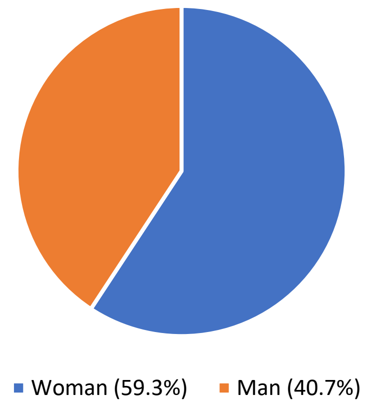

Figure 109. Gender of the students.

1.3. In the current quarantine situation, are you performing any online psychomotor activity?

Table 63. Activities taught by the students.

Activity	Absolute Frequency
Dance	9
Play musical instruments	7
Physical training	13
Martial Arts	8
Yoga	5
Performance	3
Drawing	1
Total	**2**

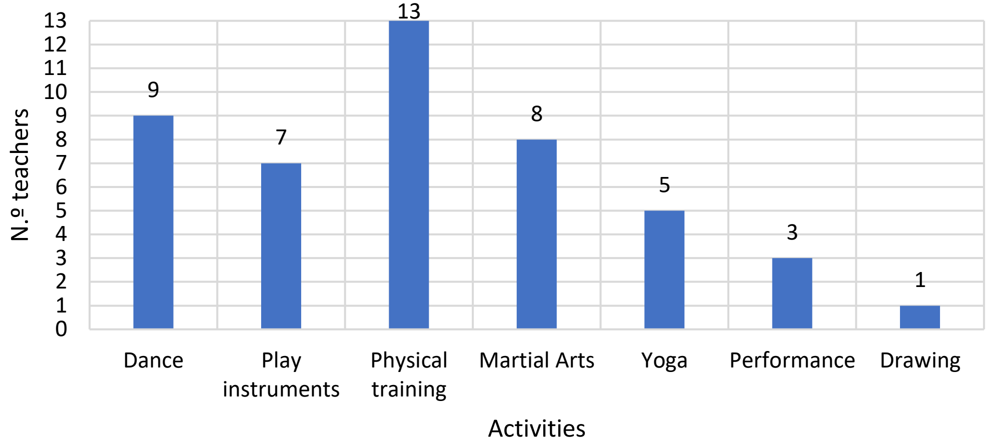

Figure 110. Activities taught by the students.

Table 64. Statistics about number of activities taught by students.

	Nº Activities taught per student
Mean	1.7037037
Mode	2
Standard Deviation	0.77533193

1.4. Which kind of media do you use to perform the psychomotor activity?

Table 65. Kind of platforms used by students.

Kind of Media used to learn	Absolute Frequency
Interactive classes	5
Recorded Videos	13
Both	9
Total	**2**

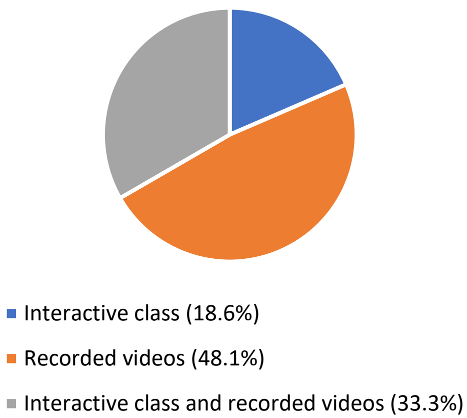

Figure 111. Kind of platforms used by students.

IV.II.II How the Students Learn Psychomotor Activities Online

In this subsection it is shown the information about which tools the students use for learning online the psychomotor activities, their experience with those tools and the characteristics of the classes.

Most of students use YouTube for learning online (19 apparitions of 48), and the rest of them are mainly distributed between Zoom (8 apparitions of 48), Instagram Direct (6 apparitions of 48), Skype (5 apparitions of 48) and WhatsApp (5 apparitions of 48). Facebook and Hangouts are the less used. Two students added Jitsi and RTVE a la Carta, which have one apparition each one. Those two services were no contemplated when creating the questionnaire, and the use of an open question have allowed to identify them. The more frequent number of platforms used for learning online is 1, with a mean of 1.77 and a standard deviation of 0.8.

The most common devices used for learning online are laptops and smartphones (16 apparitions of 49 each one). This is followed by Smart TVs (8 apparitions of 49) and Tablets (7 apparitions of 49). The device less used was the desktop computer. The reason of this could be that laptops, smartphones and tablets are highly portable, and can be moved to different parts of the environments. Further, smart TVs and Tablets usually provide better displays and resolution.

Only 22.2% of the students attend classes with somebody, the rest of them do it alone. Most of the students have started to attend online classes since the quarantine started (3 months before the questionnaire was taken), and only four of them have been taken online classes before. This explains why most of them have only taken between 1 and 49 online classes. The number of students that assist to the classes simultaneously use to be less than 10 (10 apparitions of 27). Five of the students attend the classes alone.

Most students identify themselves as skilled (29,6%) or partially skilled (55.6%) when using the tools and technologies needed for online learning. This could explain why 96.3% of students have installed their own tools and applications themselves, without needing help.

At the end of the section, there are two open questions regarding the advantages and disadvantages of online learning identified by the students. The answers of the students indicate that the main advantages that they can find in online learning are that they don’t have to move from home to practice their psychomotor activities, and the fact that they can take the classes at any time. Three of the students said that it is good that they can take the classes at their own pace. The fact that the videos of the classes can be paused, repeated, or the speed can be changed has been also identified as positive. Some other advantages mentioned by the students are that they can keep in shape in their own homes, comfort, safety, they save money in transportation, privacy, possibility of take the classes during a quarantine, and that there is a broad catalog of videos available on YouTube. The students also identified some disadvantages. The most common disadvantage identified was that technical problems with the hardware used or with the internet connection may occur. The lack of interaction with other persons and the lack of space at home for executing the movements were identified as disadvantages as well. Other disadvantages identified were the lack of constancy, the lack of material for execute some activities, the lack of communication with the teacher, and the fact that the movements cannot be properly observed.

The questions and the responses of the participants are shown in the next pages.

2.1. Which platforms do you use/have used to learn online?

Table 66. Platforms used by students.

Platform	Absolute Frequency
YouTube	19
Zoom	8
Instagram Direct	6
Skype	5
WhatsApp	5
Facebook	2
Hangouts	1
Jitsi	1
RTVE A la carta	1
Total	48

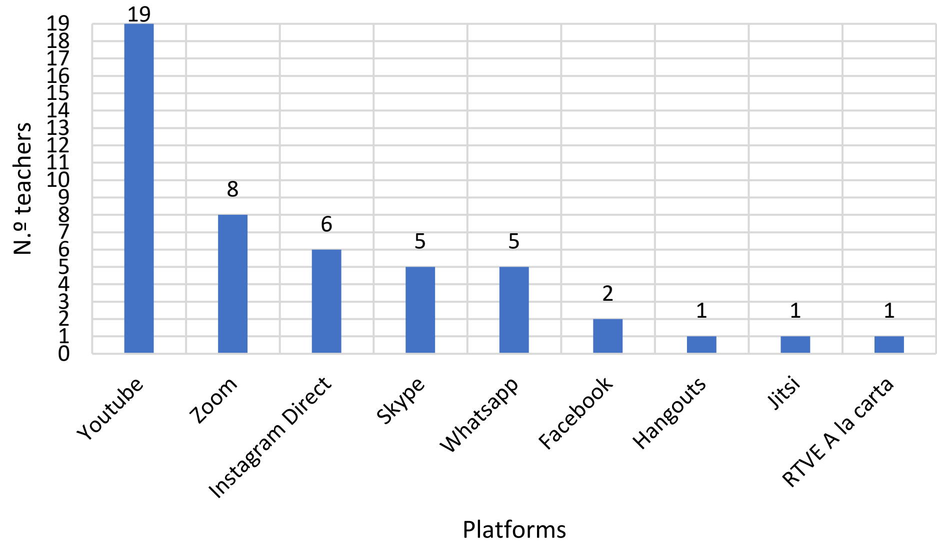

Figure 112. Platforms used by students.

Table 67. Statistics about the number of platforms used by students.

	Nº platforms used by student
Mean	1.77777778
Mode	1
Standard Deviation	0.80064077

2.2. Which devices do you use for attending online classes?

Table 68. Kind of devices used by students.

Device	Absolute Frequency
Laptop	16
Smartphone	16
Smart TV	8
Tablet	7
Desktop computer	2
Total	49

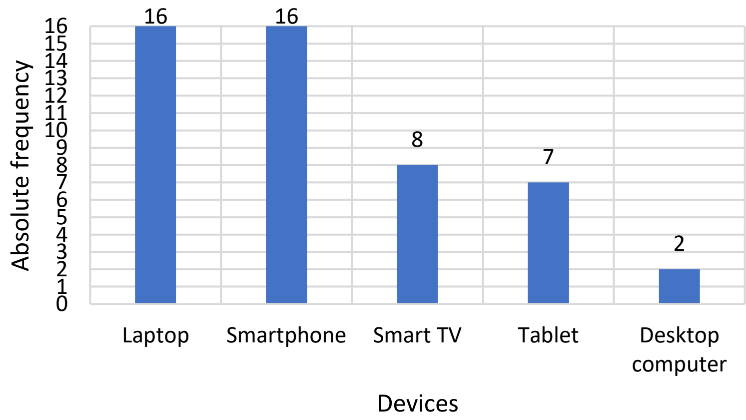

Figure 113. Kind of devices used by students.

Table 69. Statistics about the number of devices used by students.

	Nº devices used by students
Mean	1.81481481
Mode	1
Standard Deviation	0.96225045

2.3. Do you attend classes with somebody? (E.g. your couple, roommate, relative…)

Table 70. How the students learn.

Do you teach with somebody's help?	Absolute frequency
I teach alone	6
I teach with somebody	21
Total	27

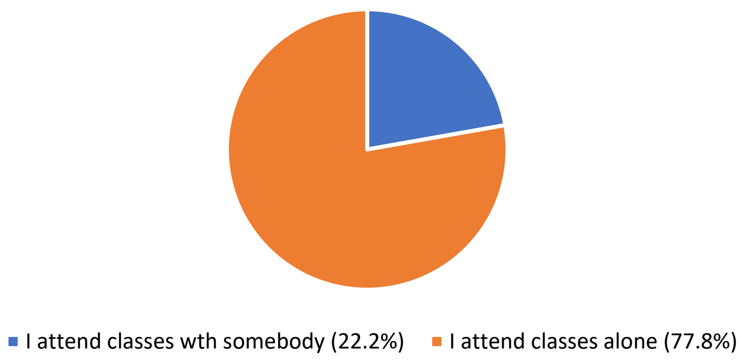

Figure 114. How the students learn.

2.4. How many time have you been attending online classes?

Table 71. Time that the students have been learning online.

How many time have you been learning online?	Absolute frequency
< 2 weeks	4
>= 2 weeks < 1 month	2
>= 1 month < 2 month	5
>= 2 month < 3 month	7
>= 3 month < 4 month	5
>= 4 months	4
Total	**27**

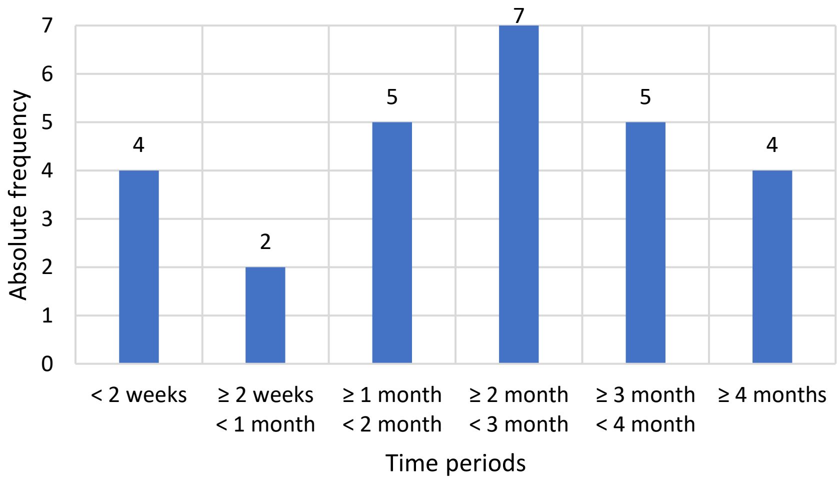

Figure 115. Time that the students have been teaching online.

2.5. How many sessions do you estimate that you have attended in that time?

Table 72. Number of sessions attended by students.

Nº sessions taught	Absolute frequency
< 10	6
10 - 19	9
20- 29	4
30-39	2
40-49	3
50-59	0
60-69	1
70-79	0
80-89	1
90-99	0
>100	1
Total	**27**

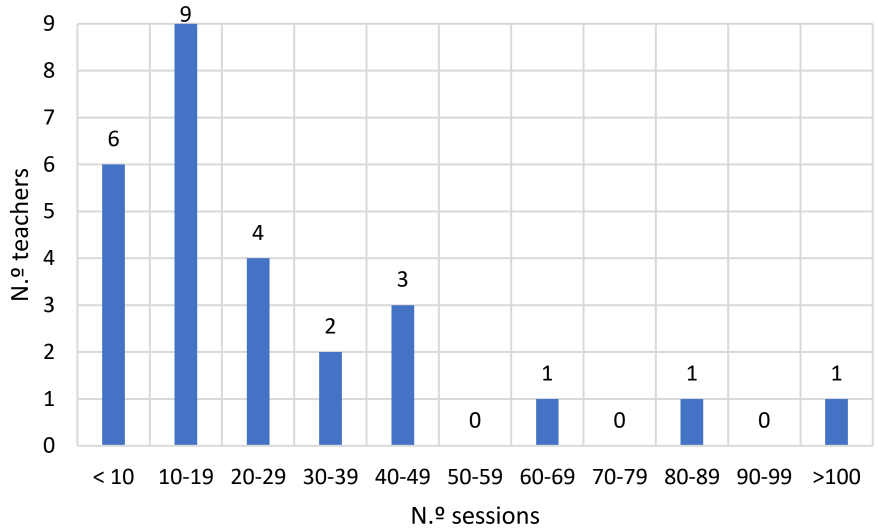

Figure 116. Number of sessions attended by students.

2.6. How many students use to attend online classes at the same time as you?

Table 73. Number of students that attend the online classes.

Nº students per class	Absolute frequency
0	5
< 10	10
10-19	4
20-29	3
30-39	1
40-49	0
>=50	4
Total	27

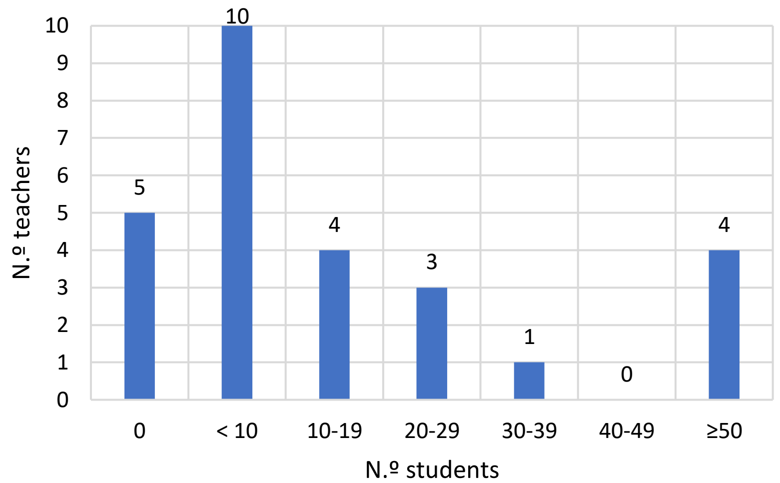

Figure 117. Number of students that attend the online classes.

2.7. Which is your level using applications and technologies of online learning?

Table 74. Level of the students when using online learning tools.

Level using online learning tools	Absolute frequency
Basic	4
Medium	15
Advanced	8
Total	**27**

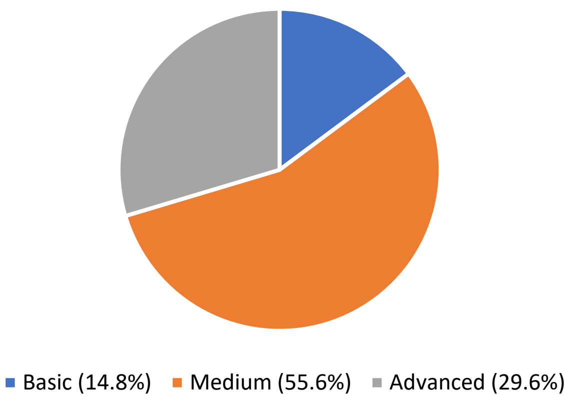

Figure 118. Level of the students when using online learning tools.

2.8. Have you installed the necessary tools/apps by yourself?

Table 75. Information about how the students have installed the online learning tools.

Have you installed the necessary applications for learning online?	Absolute frequency
I installed it	26
I needed help	1
Total	**27**

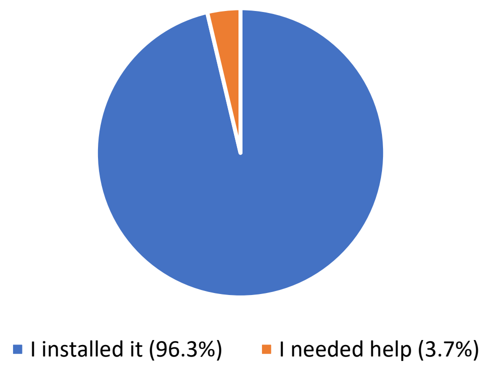

Figure 119. Information about how the students have installed the online tools.

2.9. Which advantages have you found when learning psychomotor activities online? (Answers translated from Spanish)

- “La única ventaja es no tener que desplazarte al lugar.” – “The only advantage is that I don’t have to move to the place.”
- “No necesitas desplazarte.” – “You don’t need to move from home.”
- “seguridad y comodidad.” – “Safety and comfort.”
- “Comodidad desde casa, el ritmo adaptable al nivel físico, poder verlas en diferido. Respecto a los vídeos grabados, lo que he mencionado más el amplio catálogo de internet.” – “Comfort from home, rhythm adaptable at physical level, classes can be seen referred. Regarding the recorded video, which I mentioned, plus the great catalog from internet.”
- “Que puedo hacer todo a mi ritmo.” – “I can do everything at my own pace.”
- “Ponerse en forma en casa sin necesidad de salir a la calle con control y motivación.” – Get fit at home, without needing to go out with control and motivation.”
- “Poder relajarte sin salir de casa. Yoga lo hago pagando a la profesora y ella nos pasa los videos recién grabados. El resto de ejercicio lo hago en YouTube y no tengo que pagar gimnasio, lo cual es muy práctico.” – “I can relax at my own home. I practice Yoga by paying my teacher, so she can send us the videos newly recorded. I do the rest of exercises with YouTube videos and I do not have to pay the gym, which is so practical.”
- “No dejar de realizarlas, comparando con la forma presencial, son mucho peores.” – “Not stopping attending classes. If compared with in-person classes, online are worse.”
- “Ninguna.” – “None.”
- “Ninguna.” – “None.”
- “Estar sola te obliga a centrarte más en lo que estás haciendo, fijarte bien en los movimientos y ser más crítica contigo misma” – “Being alone forces yourself to focus more in what you are doing, observe the movements and being more critic with yourself.”
- “puedes parar el video y repetir el ejercicio las veces que necesites.” – “Stopping the video and repeating the exercises as needed.”
- “Autonomía.” – “Autonomy.”
- “Honestamente prefiero presencial, pero esto se puede hacer en casa y ayuda a mantener.” – “Honestly, I prefer face to face classes, but this can be done at home and help maintain (¿).”
- “La comodidad de hacerlo a distancia desde tu propia casa.” – “The comfort of doing it online from home.”
- “Pocas la verdad.” – “Honestly, not too much.”
- “Ninguna.” – “None.”
- “La accesibilidad.” – “Accessibility.”
- “Flexibilidad de horario y repetición de las clases que más me han gustado.” – “Flexible Schedule and repetition of the classes that I liked the most.”
- “Poder realizarlas cuando yo quiera.” – “I can take the classes whenever I want.”
- “Mantenimiento físico.” – “Physical maintenance.”
- “Mantener actividad y tono muscular.” – “Keeping activity and muscular tone.”
- “Disponibilidad horaria e intimidad.” – “Schedule availability and privacy.”
- “Dibujar siempre lo he hecho a modo de técnicas de dibujo no he ido a clases, las clases de gimnasio han sido para activar el cuerpo y a diferencia del gimnasio en el tema de preparación se ahorra tiempo” – “I have always drawn using drawing techniques, I have not attended classes. The gym classes were to activate the body, and differently from the gym, you save time in getting prepared.”
- “Puedes cambiar la velocidad del video para ver las partes complicadas a cámara lenta y las partes habladas o menos importantes al doble de velocidad. También pues retroceder y ver el video tantas veces como sea necesario. Y lo más importante para mí es hacerlo a mi ritmo sin deadlines.” – “You can change the video speed for observe complicated parts at slow-motion, and putt he video at double speed for the spoken or less important parts. You can also rewind the video and watch it as many times as needed. The most important thing to me is doing it at my own pace and without deadlines.”
- “No hay que desplazarse.” – “I don’t have to move to the place.”
- “Puedes estar en forma desde casa.” – “You can be in shape from home.”

2.10. Which disadvantages have you found when learning psychomotor activities online? (Answers translated from Spanish)

- "Falta de contacto personal, peor visibilidad e imposibilidad de realizar determinados ejercicios correctamente, sobre todo en artes marciales donde el contacto es esencial." – "Lack of personal interaction, worse visibility, impossible to execute specific exercises correctly, especially in martial arts where personal interaction is essential."
- "La corrección de posturas." – "Posture correction."
- "ninguna." – "None."
- "El profesor no te puede corregir la postura." – "The teacher cannot correct the postures."
- "Quizá no hay nadie que corrija mis posturas si lo hago mal." – "May be there is nobody to correct my postures if I do it wrongly."
- "No poder hacer dudas. No saber si algún ejercicio se está haciendo mal y no tener la oportunidad de recibir un feedback para corregirlo." – "Impossibility of solving doubts. Impossibility of knowing if an exercise is done wrongly and getting feedback."
- "No sé si algunos ejercicios los hago de forma correcta." – "I don't know if I am executing some exercises correctly."
- "Conexión, latencia, puerta de ruido." – "Connection, latency, noise gate."
- "Es mejor presencial." – "Face to face is better."
- "Poco espacio, no es un lugar asociado al deporte ya que es tu casa y ocasionalmente problemas de red" – "Lack of space. Your house is not a place associated with sport, and occasionally there are network issues."
- "Falta de espacio y en algunos casos de visibilidad (al no tener mucho hueco o enfocas a la parte superior del cuerpo o a la inferior, las correcciones se alargan y se pierde dinamismo)." – "Lack of space, and in some cases, lack of visibility (since there is no much space, you can either or focus the camera in the upper or lower body. The corrections are extended, and dynamism is lost."
- "Internet no siempre funciona correctamente." – "Internet is not always working properly."
- "Problemas técnicos de micrófono." – "Technical problems with my microphone."
- "Falta de contacto, problemas de audio o vídeo, imposibilidad de hacer cosas en común..." – "Lack of contact, problems with audio and video, impossibility of doing things with people…"
- "Creo que existen ciertas barreras digitales que es muy complicado sortear." – "I think that there are some digital barriers that we have to overcome."
- "Sobre todo he notado que hay cierta pérdida en el lenguaje verbal y no verbal, cierto desfase que suele depender de la velocidad de conexión, plataforma etc. que sobre todo afecta mucho a cualquier práctica que requiera sincronía." – "Some verbal and non-verbal language is lost, there is some delay that often depends on the network speed, platform… it affects mainly practices that requires synchronization."
- "No es lo mismo, el teatro es cara a cara. Se está haciendo lo que se puede, pero como que se pierde la esencia. Además, yo tengo a veces problemas de Wifi y me saca de la llamada y es un poco coña o." – "It's not the same, theater is face to face. We are doing our best, but it loses it essence. Further, I have sometimes problems with my WIFI, and it takes me out of the videocall, this is insane."
- "La atención que te dedica el profesor para corregirte es bastante más limitada." – "The attention that the teacher put in you for correcting you is much more limited."
- "Pierdes muchísimo: ves menos los detalles, bailar solo o en grupo anímicamente es muy distinto, te despejas menos por seguir encerrado, seguir al profesor es más complicado." – "A lot is lost. You see less details, dance alone or in group is too different psychically, you cannot evade since you still locked, follow the teacher is more complicated."

- “En mi caso la clase es en diferido y esto implica que el feedback hacia el entrenador es diferente. Tienes que presentar más atención de hacer bien los movimientos porque no le tienes delante para que te corrija.” – “In my case, the class is deferred, which implies that the feedback towards the trainer changes. You must focus more in doing the movements correctly since there is no one to correct you.”
- “Ninguno.” – “None.”
- “No poder hacer todas las actividades por falta de espacio.” – “I cannot do all the activities because of the lack of space.”
- “Falta de espacio.” – “Lack of space.”
- “Ninguno, solamente algún fallo eléctrico o mal funcionamiento de wifi.” – “None, just some electrical failures, or WIFI issues.”
- “La constancia.” – “Constancy.”
- “Las clases en persona tienen la ventaja de que el profesor/a puede corregir los malos hábitos, algo que me parece esencial para un mejor aprendizaje. La ausencia de consejos personalizados en cursos online es la desventaja más grande.” – “Face to face classes have the advantage that the teacher can correct the bad habits, something that I think is essential for a good learning. The lack of personalized feedback in online curses is the biggest disadvantage.”
- “No es lo mismo ya que el material no es el óptimo.” – “It is not the same because the material is not the best.”
- “No tengo material necesario para hacerlo bien.” – “I don’t have the required materials for doing it properly.”

Table 76. Advantages of online learning of psychomotor activities identified by the students.

Advantages of online learning identified by the students	Number of students that considered it an advantage in the free question
Doing it from home, without going out	9
Flexibility in Schedule	6
None	4
Doing it at my own pace	3
You can stop/repeat/change the speed of the video	3
Keep in shape from home	3
Comfort	2
Safety	1
Broad catalogue of videos online	1
Save money	1
Possibility of continuing with the classes	1
Forces yourself to focus and self-observe	1

Table 77. Disadvantages of online learning identified by the students.

Disadvantages of online learning identified by the students	Number of students that considered it an advantage in the free question
Lack of feedback	8
Lack of interaction with other people	5
Lack of visibility of the movements	3
Technical problems	8
None	3
Lack of communication	1
Lack of space	4
Lack of constancy	1
Lack of materials	2

IV.II.III Advantageous Situations

In this subsection, the students were exposed to a set of advantageous situations that they might have encounter when learning psychomotor activities online.

It seems that online classes does not really help the students to be pore uninhibited during online classes, since only half of them (48.1%) have answered that they feel more outgoing during online classes, also, less than half of the students (44.4%) feel that they assist less assiduously to online classes than to in-person classes. Most (81.5%) of the students consider advantages the use of recorded classes that can be later visualized and practiced, as well as the fact that they can send recorded videos to their teachers for receiving better feedback (77.8%), and the fact that they can attend classes in different hours (81.5%). Even when the students have admitted that online learning entails a set of advantages, only half of students (55.6%) answered that they will continue attending online classes even if the quarantine ends.

At the end of the section, a free question was asked again regarding advantages of online learning. Eight students answered this question and some of the advantages identified by the users in the previous section were emphasized again, including the fact that students save money on transportation. Two new advantages have been identified here, and they are that some activities can be better at home, and that a good trainer can help students to improve and to keep training. Some of the answers emphasized some of the disadvantages previously mentioned, like the lack of constancy and the lack of personal interaction.

3.1. Do you feel more uninhibited or outgoing since you are more comfortable in your environment, and you can perform without drawing the attention of your partners?

Table 78. Answers to question 3.1 in students' questionnaire.

Answer	Absolute frequency
Yes	13
No	14
Total	**27**

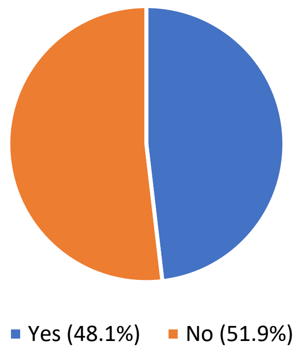

Figure 120. Answers to question 3.1 in students' questionnaire.

3.2. Do you think that the fact that you can record the classes (or your teacher can distribute them), so you can repeat and review them is an advantage of online learning?

Table 79. Answers to question 3.2 in students' questionnaire.

Answer	Absolute frequency
Yes	22
No	5
Total	**27**

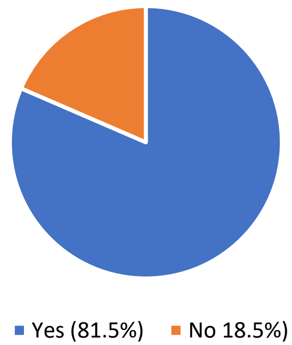

Figure 121. Answers to question 3.2 in students' questionnaire.

3.3. Do you consider that the fact that you can send your progress in video to your teacher, so he/she can review it more carefully and give you a better feedback, is and advantage of online learning?

Table 80. Answers to question 3.3 in students' questionnaire.

Answer	Absolute frequency
Yes	21
No	6
Total	**27**

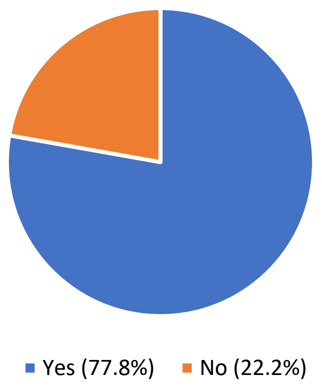

Figure 122. Answers to question 3.3 in students' questionnaire.

3.4. Do you think that you assist more assiduously to online classes, since you don't have to leave your environment?

Table 81. Answers to question 3.4 in students' questionnaire.

Answer	Absolute frequency
Yes	12
No	15
Total	**27**

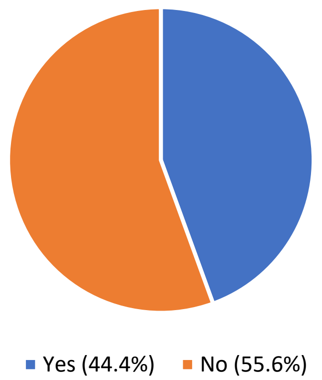

Figure 123. Answers to question 3.4 in students' questionnaire.

3.5. Do you consider that the fact that you can attend online classes at different hours than on in-person classes, extending the classes since you don't depend on physical installations, is an advantage?

Table 82. Answers to question 3.5 in students' questionnaire.

Answer	Absolute frequency
Yes	22
No	5
Total	**27**

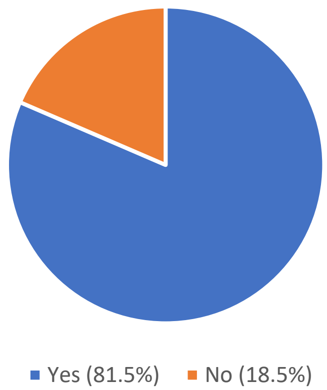

Figure 124. Answers to question 3.5 in students' questionnaire.

3.6. Will you continue teaching online in the future, even if the quarantine situation ends?

Table 83. Answers to question 3.6 in students' questionnaire.

Answer	Absolute frequency
Yes	15
No	12
Total	**27**

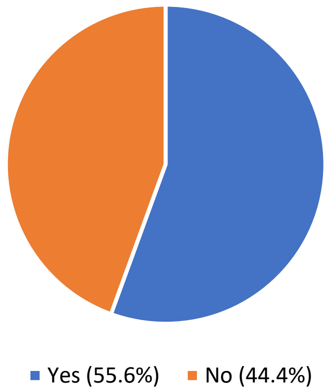

Figure 125. Answers to question 3.6 in students' questionnaire.

3.7. Is there any other advantage that comes to your mind, or do you want to add something about your answers? (Translated from Spanish)

- “Quiero concluir diciendo que determinadas actividades sí pueden ser mejores en casa, como baile individual. Pero en lo que se refiere a las artes marciales considero esencial las clases presenciales.” – “I want to conclude saying that specific activities can be better at home, like dancing alone. But regarding martial arts, I consider that face to face classes are essential.”
- “Ahorro de tiempo en el trasporte al lugar.” – “You save money in transportation.”
- “Asistir a clases presenciales tiene sus ventajas de que suelen ser de pago y ya simplemente por haber pagado por ello ya te obligas a asistir, por el contrario, con las clases online que hago, al ser gratuitas, puedo dejar de hacerlas si no me apetece o si no me viene bien sin remordimientos de haberme gastado dinero y no hacerlo.” – “Assisting to in-person classes has it advantages. In-person classes are often paid, and that simple fact forces you to assist. On the other hand, since the online classes that I attend are free, I can stop doing them if I want without any regret for spending money on it and not doing it.”
- “Al igual que con todo, necesitas ser constante con los ejercicios. Al no tener a nadie pendiente de ti o la obligación de asistir a una clase a una hora determinada en un lugar X, es más fácil dejarlo.” – “As with everything, you need to be constant with the exercises. Since nobody is aware of you, and you don’t have the obligation of attend classes at a certain time, it is easier to give up.”
- “En mi opinión, grabarse y tener videos ayuda, pero sin clases presenciales sirve de poco. Necesitas una base presencial y luego cuando ya sabes los movimientos sigues un vídeo.” – “In my opinion, recording yourself and having videos can help, but it is useless without in-person classes. You need an on-person basis, and once you know the movements, you can continue with a video.”
- “Mi motivación por el ejercicio físico se ha visto incrementada notablemente desde que empecé a entrenar todos los días. A pesar de que sean videos, tengo un buen entrenador que hace que quiera continuar todos los días. Me siento mucho mejor desde que empecé a hacer ejercicio y algunos días lo combino con baile. Está siendo mi vía de escape para no echar tanto de menos la natación que suele ser el deporte habitual que realizo.” – “My motivation by physical exercise has been incremented since I started to train everyday. Regardless it is a video, I have a good trainer that makes me want to continue every day. I feel much better since I started to exercise, and some days I combine it with dance. This is my way of evading from missing swimming, which is the sport that I habitually practice.”
- “No” – “No.”
- “Necesito el contacto con otro compañero para progresar.” – “I need interaction with a partner to progress.”

IV.II.IV Disadvantageous Situations

In this subsection, the students were exposed to a set of disadvantageous situations that they might have encountered when learning psychomotor activities online.

Seems like there is more consensus in the disadvantageous situations. For most of the students, the lack of communication with the teacher (70.4%) and the fact that the teacher cannot properly observe their movements (74.1%) suppose a problem. Students feel that they are participating less in online classes (66.7%), that they are not training enough in online classes (74.1%) and that it is more difficult to notice that they are progressing in online classes (63%). Even when technical problems were one of the most identified problems by the students in the open questions, only 25.9% of the students consider the delay in the connection a problem, so the technical problem must occur elsewhere.

Again, a free question for expanding the possible disadvantages previously identified was asked. A student considered that online learning only entails disadvantages in online learning, and other student emphasizes again the lack of interaction with other persons. The lack of space was mentioned again, and two new disadvantages have been identified. The first one is that practicing alone does not motivate students, and the second one is that you cannot wind down the same in online classes as on in-person classes.

4.1. Is the delay in your connection a problem? E.g. There is delay between the moment your teacher explains something, and it reaches you.

Table 84. Answers to question 4.1 in students' questionnaire.

Answer	Absolute frequency
Yes, the delay is a problem	7
No, the delay is not a problem	6
There is no delay	14
Total	**27**

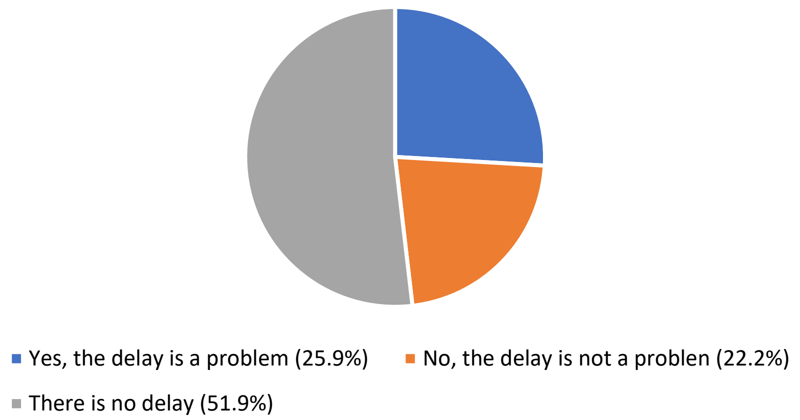

Figure 126. Answers to question 4.1 in students' questionnaire.

4.2. The lack of communication supposes any problem when learning online? E.g. You cannot talk with your teacher, you can only talk through text, or is it difficult for your teacher to correct your movements.

Table 85. Answers to question 4.2 in students' questionnaire.

Answer	Absolute frequency
Yes	19
No	8
Total	**27**

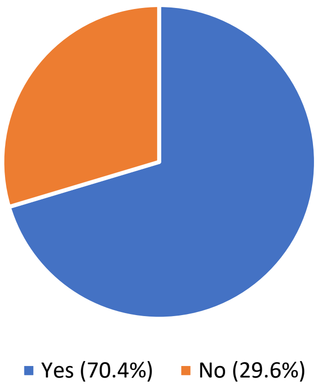

Figure 127. Answers to question 4.2 in students' questionnaire.

4.3. The fact that your teacher cannot observe properly your movements suppose any problem when learning online? E.g. Your teacher has to ask you to move or turn to observe better your movements, or your teacher cannot even observe your movements.

Table 86. Answers to question 4.3 in students' questionnaire.

Answer	Absolute frequency
Yes	20
No	7
Total	**27**

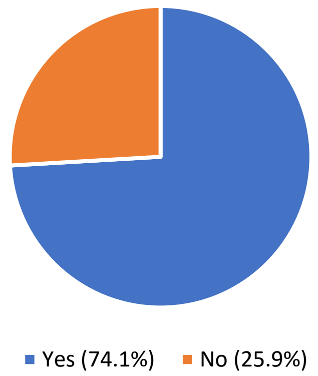

Figure 128. Answers to question 4.3 in students' questionnaire.

4.4. Do you feel that it is more difficult to notice your progress in the online classes?

Table 87. Answers to question 4.4 in students' questionnaire.

Answer	Absolute frequency
Yes	17
No	10
Total	**27**

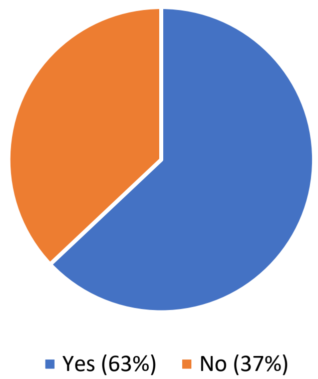

Figure 129. Answers to question 4.4 in students' questionnaire.

4.5. Do you feel that you are les participative in online classes?

Table 88. Answers to question 4.5 in students' questionnaire.

Answer	Absolute frequency
Yes	18
No	9
Total	**27**

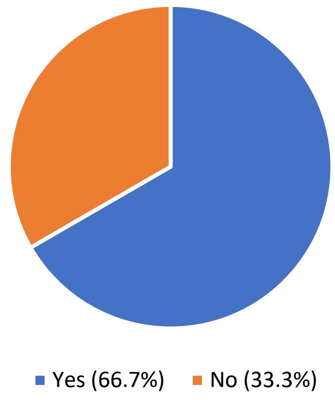

Figure 130. Answers to question 4.5 in students' questionnaire.

4.6. Do you feel that you are not training/practicing enough in online classes as you do on in-person classes?

Table 89. Answers to question 4.6 in students' questionnaire.

Answer	Absolute frequency
Yes	20
No	7
Total	**27**

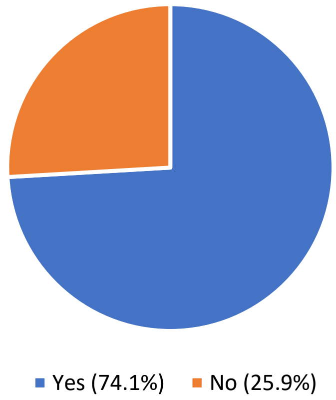

Figure 131. Answers to question 4.6 in students' questionnaire.

4.7. Is there any other disadvantage that comes to your mind, or do you want to add something about your answers? (Translated from Spanish)

- “Considero que en las artes marciales las clases online son todo desventajas.” – “I consider that everything is a disadvantage when learning martial arts online.”
- “Pero una gran desventaja es que no vale para todos los deportes, evidentemente, y el espacio en casa no es el adecuado para todos los movimientos.” – “A great disadvantage is that it is not suitable for every sport, and the space available at home is not enough for all movements.”
- “Nada.” – “Nothing.”
- “No desconectas igual. Al salir de casa y bailar en grupo te despejas mucho más.” – “You cannot wind down. Going out of home and dance in group allow you to wind down more.”
- “En aikido se necesita la energía del atacante para poder armonizar con ella, solo prácticas la posición y la técnica, pero sin atacante falta lo principal de este arte.” – “In aikido, you need the energy of the aggressor so you can harmonize whit it, you only practice position and technique, but without an aggressor the main thing of this art is missing.
- “No tiene aliciente practicar solo.” – “Practice alone has no incentive.”

IV.II.V Emotional State During Classes

Finally, the students had to answer to a set of questions about their emotional state or how they feel during online classes, to probe if it affects them emotionally.

The results of this questionnaire indicate that online classes involucres negative emotional states in the students. It can be seen how 74.1% of students feel more motivation on in-person classes than on in-person classes, and how 63% of students feel more bored in online classes. The facts that the students cannot be with their teachers and partners in-person affects negatively 59.3% of students, and 77.8% of students feel more satisfied after an in-person class. 63% of students would not join more online classes in the future, but also a 63% of them have not though in abandon online classes. This could be because during the quarantine status, they have no more options for practicing their psychomotor activities.

In this section an open question about the emotional status of the students was asked, and eight students answered. One of the students answered that she feels that she works more on in-person classes, since the executed exercises are more active, However, for other kind of exercises that have a fixed position like Yoga or abs, she feels very satisfied after an online class. Other student answered that he is bored in online classes, but maybe in classes regarding other psychomotor activities it would be different. A third student indicated that takes a great effort to take an online class. Again, a student answered that person interaction is important. Another studied answered that even when online classes provide privacy, the fact that other students are looking at you can motivate you. Finally, a student said that it is complicated not to be distracted when you are in an online class with other students and your camera is not on.

Those answers again give a hint about the one of the problems of online learning. There is no people or teacher with you, giving feedback and taking the exercises together, so the motivation created by the group is lost. However, in some exercises that can be done alone, an online class can be beneficial since a student can focus in self-observation and in the executed exercises.

5.1. Do you feel less nervous during online classes than during in-person classes?

Table 90. Answers to question 5.1 in students' questionnaire.

Answer	Absolute frequency
I feel less nervous in online classes	9
I feel less nervous on in-person classes	4
I feel the same in both cases	14
Total	**27**

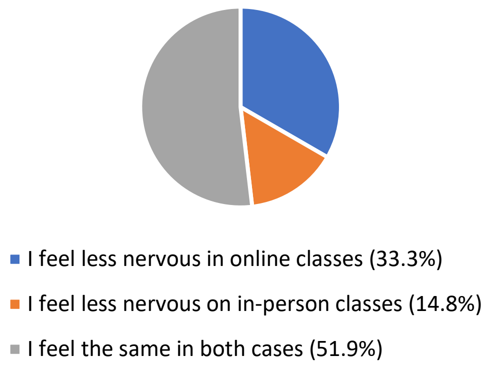

Figure 132. Answers to question 5.1 in students' questionnaire.

5.2. Do you feel more motivated during in-person classes than during online classes?

Table 91. Answers to question 5.2 in students' questionnaire.

Answer	Absolute frequency
I feel more motivated on in-person classes	20
I feel more motivated in online classes	1
I feel the same in both classes	6
Total	**27**

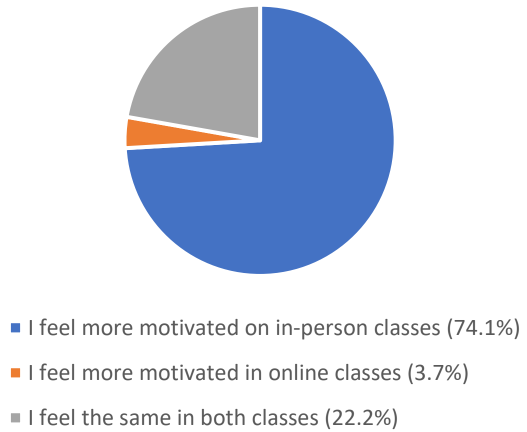

Figure 133. Answers to question 5.2 in students' questionnaire.

5.3. Do you feel more bored during online classes than during in-person classes?

Table 92. Answers to question 5.3 in students' questionnaire.

Answer	Absolute frequency
I feel more bored in online classes	17
I feel more bored on in-person classes	2
I feel the same in both classes	8
Total	**27**

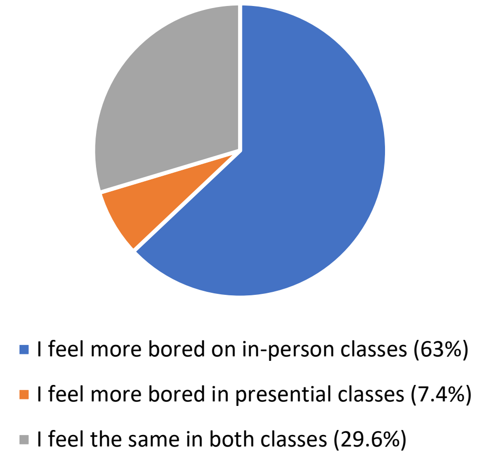

Figure 134. Answers to question 5.3 in students' questionnaire.

5.4. How affect you the fact that you are not with your teacher and partners in person?

Table 93. Answers to question 5.4 in students' questionnaire.

Answer	Absolute frequency
It affects me positively	2
It affects me negatively	16
It does not affect me	9
Total	**27**

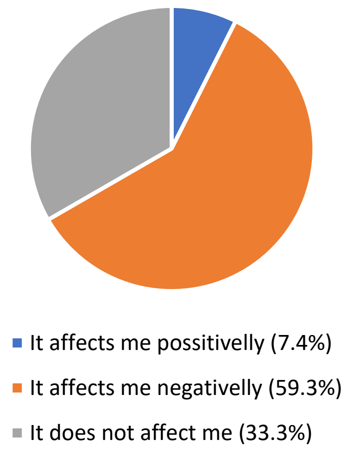

Figure 135. Answers to question 5.4 in students' questionnaire.

5.5. Do you feel more satisfied after an online class than after an in-person class?

Table 94. Answers to question 5.5 in students' questionnaire.

Answer	Absolute frequency
I feel more satisfied after in-person classes	21
I feel more satisfied after online classes	1
I feel the same after both classes	5
Total	**27**

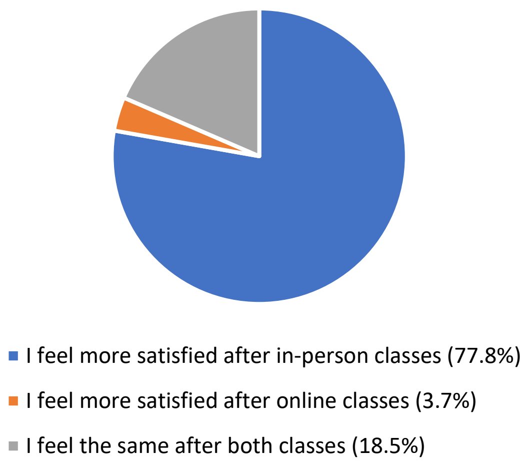

Figure 136. Answers to question 5.5 in students' questionnaire.

5.6. Have you though in abandon online classes?

Table 95. Answers to question 5.6 in students' questionnaire.

Answer	Absolute frequency
Yes	10
No	17
Total	**27**

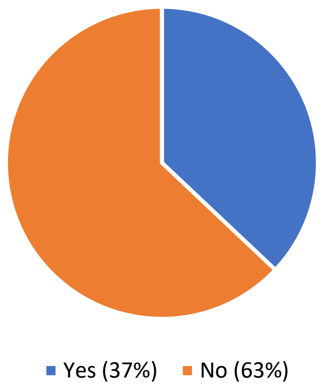

Figure 137. Answers to question 5.6 in students' questionnaire.

5.7. How affect you the fact that you are not with your students in person?

Table 96. Answers to question 5.7 in students' questionnaire.

Answer	Absolute frequency
Yes	10
No	17
Total	**27**

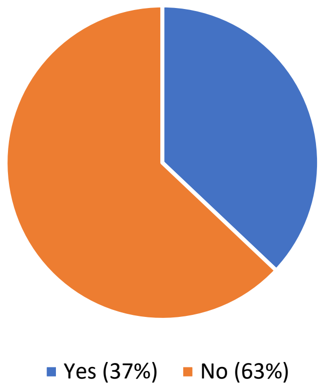

Figure 138. Answers to question 5.7 In students' questionnaire.

5.8. Is there anything about your mood that comes to your mind, or do you want to add something about your answers? (Translated from Spanish)

- "Siento que trabajo más en las presenciales, se ejecutan todos los ejercicios de los deportes más activos, y baile. Sin embargo, en actividades como Yoga, Pilates, estiramientos, abdominales, actividades de suelo, y de posición fija como sentadillas, me siento muy satisfecha después de la clase online." – "I feel that I work more at in-person classes, all exercises and the most active sports are executed, and dance. However, in activities like Yoga, Pilates, stretching, abs, floor activities, and fixed position like squats, I feel so satisfied after an online class."
- "No" – "No."
- "En mi caso concreto, creo que me aburro más en las clases online debido a que el dinamismo de la clase presencial se pierde. Pero es probable que en otros tipos de clase online no notase diferencia." – "In my case, I think that I am more bored in online classes due to the lost of dynamism of in person classes. But it is possible that in other types of online classes I won't notice it."
- "Tengo que hacer un esfuerzo de voluntad muy grande en las clases online." – "I have to make a great effort in online classes.
- "Nada sustituye al contacto presencial." – "Nothing can replace personal interaction."
- "Es cierto que las clases online te dan esa discreción o intimidad. Pero el hecho de saber que todos a tu alrededor te pueden estar observando es un punto de motivación a favor." – "It is true that online classes give you privacy and discretion. But knowing that everybody around you could be observing you, is a motivational factor."
- "Cuando la clase es online vía Zoom con varios alumnos y no es necesario encenderla cámara es más complicado no distraerse." – "When the class is imparted using Zoom with several students and it is not necessary to turn on the camera, it is more difficult not to get distracted."
- "Echo de menos ir al gimnasio." – "I miss the gym."