Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica
A.A. 2023/2024

MACHINE-LEARNING CLASSIFICATION OF DAILY LIVING ACTIVITIES IN PARKINSON'S DISEASE

Relatori

Prof.sa Laura Gastaldi

Prof. Álvaro Gutiérrez

Candidato

Enrico Carbone

Summary:

Abstract......5

1 Evaluation of the current state of the art in activity recognition using inertial sensors......6

1.1 Disease overview......6

1.2 Application fields of activity recognition......8

1.2.1 Human activity recognition......9

1.2.2 Motor anomaly detection......11

1.3 Steps of human motion modelling system......13

1.4 Data acquisition: the role of sensors and wearable devices......13

1.5 Signal pre-processing......16

1.6 Feature extraction and selection......17

1.7 Training and classification: the role of Artificial Intelligence......19

1.7.1 Machine learning algorithms......22

1.7.2 Deep learning algorithms......23

2 Methodology......25

2.1 Dataset......25

2.2 Preprocessing and filtering......26

2.3 Feature selection and feature extraction......28

2.4 Signal segmentation......29

2.5 Classification approach......30

2.6 Results......32

3 Algorithm developed......37

3.1 Algorithm description......37

3.2 Results......39

4 Algorithm validation and results......44

4.1 Data collection......44

4.2 Results......46

5 Conclusion and future development......50

Bibliography......53

Acknowledgements......57

Abstract

The thesis work was carried out at the Polytechnic University of Madrid as part of a larger project called *NETremor*. This project involves the development of a methodology for the daily monitoring of tremor in Parkinson's (PD) and essential tremor (ET) patients, using a smartwatch. Classifying daily living activities (ADL) performed by PD patients and secondly to assess during which activities tremor occurs most are important milestone in evaluating the effectiveness of drug therapy.

In this thesis work, the main goal was to build up an algorithm based on artificial intelligence models, starting from inertial movement unit (IMU) signals that has been provided by a previous study to classify ADLs as efficiently as possible, and then to validate it by recording new data using a IMU sensor embedded in a smartwatch [1].

The thesis consists of seven chapters: the first one provides an overview of the current state of the art in activity recognition based on inertial sensors and the use of artificial intelligence-based methods. The second chapter presents the experimental protocol to collect data, the rationale behind feature selection, and is illustrated the first classification attempt and its results. In the third chapter the final classification model is explained. The fourth chapter consists of the validation of the model using data collected in laboratory and hospital and the analysis of the results. Finally, the fifth chapter presents the conclusion and the future development.

Chapter 1

Evaluation of the current state of the art in activity recognition using inertial sensors

The objective of this chapter is to comprehensively review and evaluate the existing literature on activity recognition using inertial sensors, with a specific focus on its application in monitoring tremor in Parkinson's and essential tremor patients. This evaluation aims to provide a thorough understanding of the methodologies, techniques, and algorithms employed in tremor monitoring using inertial sensors. By achieving this objective, we can gain insights into the advancements made in this field and identify potential areas for further improvement and research.

1.1 Disease overview

Parkinson's disease (PD) and essential tremor (ET) are debilitating neurological disorders that affect a significant number of individuals worldwide. More specifically, is the prevailing movement disorder and exhibits a significant rise in frequency and prevalence as individuals age. Approximately 4% of individuals over 65 years old experience ET, while around 1% of the population above 50 suffers from PD [2]. Parkinson's disease is characterized by motor symptoms such as tremors, rigidity, and bradykinesia, while essential tremor primarily manifests as involuntary trembling movements. Monitoring and assessing tremor patterns are crucial for understanding disease progression, evaluating the effectiveness of drug therapy, and identifying activities of daily living (ADL) that trigger or exacerbate tremor symptoms.

The disease load associated with PD is not only tied up with the disease itself, but also with the progressive disability that patients experience during its course. Deficits in motor function cause mobility problems and interfere with daily activities by causing balance and walking problems that can lead to falls, injuries, and the inability to perform even the most basic tasks. And this can also be seen from an economic perspective. according to the study conducted in the

United Kingdom, by Findley the economic impact of this disease exceeds £400 million per year, and this data can only increase as the average age of population in industrialized country continues to grow [3]. This trend is clear in this graph where it's reported the number of deaths caused by PD per 100,000 people per year in the United States, form 1999 until 2017 (figure 1).

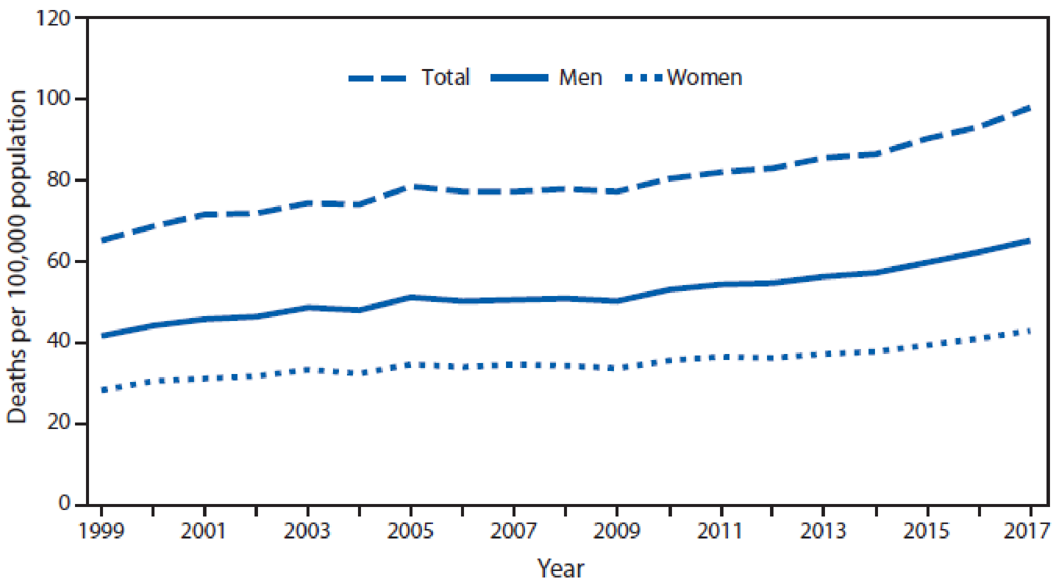


Figure 1: Death Rates for Parkinson Disease Among Adults Aged ≥65 Years, US, 2019.

Traditionally, tremor assessment has relied on subjective observation and manual scoring by healthcare professionals. However, these methods are often limited in their accuracy, objectivity, and the ability to capture the full extent of tremor characteristics. Furthermore, the intermittent nature of tremor episodes and the variability in tremor severity throughout the day make it challenging to obtain a comprehensive understanding of tremor patterns using conventional assessment methods [4].

In addition to the hallmark motor symptoms of tremors, rigidity, and bradykinesia, Parkinson's disease (PD) is characterized by a wide array of non-motor symptoms that significantly impact patients' quality of life and functional abilities. These non-motor symptoms often go unrecognized or untreated, yet they can be just as debilitating as the motor manifestations of the disease. For example, anxiety affects up to 60% of PD patients and is more common in female patients, patients whose disease started at an early age, and patients whose disease is advanced. Anxiety levels rise in response to motor fluctuations, which are linked to low dopamine periods, as well as the start of off periods or freezing while depression, which is clinically relevant in 35% of patients, is typically milder than depression seen in people without PD, and it frequently contains apathy and anhedonia. [5]

Both PD and ET present unique challenges to patients and healthcare providers. While the first one is a progressive condition with a wide range of motor and non-motor symptoms, the second primarily involves rhythmic shaking of the hands and may be more manageable with lifestyle adjustments and medication. Generally, there is currently no cure for PD, although various treatment options are available to manage its symptoms, and monitoring, and precise assessment are crucial for both conditions to tailor treatment and improve the quality of life for those affected [6].

1.2 Application fields of activity recognition

This section aims to critically explore the various application contexts of activity recognition.

There hasn't yet been a single ontological definition given for the idea or concept of "activity." Indeed, human activities lack a standard definition, vocabulary, or structure that would enable us to define a precise and shared problem statement (such as which activity must be identified, how a particular activity is described, etc.) [7]. Despite of this, Ranasinghe et al. state that is possible to assume that certain activities, like "making coffee," are hierarchically structured into actions, like "entering the kitchen," "filling the water container," and so on. These actions are further composed of operations, which are understood as discrete steps that carry out the action, such as "pushing the door handle", "opening the water tap." Activities are thought of as several actions, which are thought of as a number of atomic operations once more. An activity is meant to be executed in a specific amount of time, which is determined by adding up the durations of the individual subunits which make up the activity [8].

The field of activity recognition has demonstrated relevance and impact across different sectors, warranting a detailed analysis to fully comprehend its practical applications. Activity recognition can be very demanding due to the complexity of human motion and to do this activity models have been constructed using a variety of probability-based techniques. Among the most widely used modelling approaches are the Conditional Random Field (CRF) and the Hidden Markov Model (HMM) [9]. Hidden Markov Models (HMMs) are used to understand complex or unfamiliar activities by observing their effects. They gradually build a model of the activity based on these observations. HMMs consist of hidden states representing activities and observable variables like sensor data. HMMs make two key assumptions: first, that future states depend only on the current state, and second, that observable variables depend only on the current hidden state. HMMs find the most probable sequence of hidden states based on observed data by maximizing joint probabilities of transition and observation probabilities. Training an HMM improves accuracy, and multiple HMMs can be combined for more complex activities. However, as Kim et al. wrote in this article, Hidden Markov Models (HMMs) have limitations in representing multiple interacting activities and capturing long-range dependencies due to strict independence assumptions and Conditional Random Fields (CRFs) are more flexible alternatives. CRFs model the relationship between hidden and observed variables without strict independence assumptions. They use potential functions instead of joint probability functions and allow for arbitrary relationships between variables. CRFs are modelled as undirected acyclic graphs and are used to find conditional probabilities rather than joint probabilities. They offer more flexibility in modelling complex activities, such as those with non-deterministic or concurrent steps. [9].

Going deeper into this field, firstly, it will be examined the human activity recognition (HAR) sector and most of its application, where the ability to monitor daily activities through inertial data has become crucial for enhancing the well-being monitoring of elderly individuals and patients with specific pathologies. Secondly, the sector of motor anomaly detection, that is living an arising interest by researchers, due to recent improvements.

1.2.1 Human activity recognition

HAR is a field within the realm of computer vision and artificial intelligence that focuses on the development of techniques to automatically identify and classify human activities based on sensor data. This is a critical component in various applications, including healthcare, surveillance, sports analysis, and human-computer interaction. The advent of wearable devices and the proliferation of sensor technologies have facilitated the collection of diverse and rich datasets, enabling the development of sophisticated HAR algorithms. More specifically, it can be involved in:

- Healthcare and Rehabilitation: in the healthcare sector, HAR finds extensive applications in monitoring patients' activities and behaviours, facilitating remote patient monitoring, and assisting in rehabilitation programs. By accurately recognizing and analysing activities such as walking, sitting, or exercising, HAR systems can provide valuable insights into patients' health status, adherence to treatment plans, and progress in rehabilitation therapy [10]. This facilitates timely interventions, personalized care, and better management of chronic conditions, ultimately improving patient outcomes and reducing healthcare costs. For instance, Schrader et al. explore how these technologies, such as wearable sensors, smart home systems, and computer vision, can monitor and analyse human activities to detect early signs of decline or injury. By recognizing patterns in daily activities, these systems can provide timely interventions and personalized rehabilitation programs, ultimately improving the quality of life for elderly individuals [11].

- Sports and Fitness Monitoring: in the sports and fitness industry, HAR is utilized for tracking athletes' movements, analysing performance metrics, and providing personalized training recommendations. Wearable devices equipped with HAR capabilities can monitor activities such as running, cycling, or weightlifting, allowing athletes and fitness enthusiasts to optimize their training routines, prevent injuries, and achieve their fitness goals more effectively. HAR facilitates real-time feedback, performance analysis, and personalized coaching, enhancing athletes' performance, reducing the risk of injuries, and maximizing training efficiency. [12]
- Smart Homes and Ambient Assisted Living: HAR technologies are increasingly integrated into smart home environments and ambient assisted living systems to enhance the quality of life for elderly individuals and individuals with disabilities. And this is a very important aspect since life expectancy is increasing and there are many older people who need assistance, in fact it is estimated that 9% of adults age 65+ and 50% of adults age 85+ need assistance with the activities of daily living (ADLs), so these systems can monitor occupants' activities, detect potential risks or emergencies (e.g., falls), and provide assistance or alerts as needed [13]. By automating home tasks, adjusting environmental settings, and providing personalized assistance based on occupants' behaviours, HAR contributes to increased comfort, safety, and independence, allowing individuals to age in place and reducing the burden on caregivers and healthcare systems. [14]
- Worker Safety and Occupational Health: another significant application of HAR is in ensuring the safety and well-being of workers in various industries. By monitoring workers' activities and environments, HAR systems can identify potential hazards, assess ergonomic risks, and detect unsafe behaviours or conditions in real-time. In their study Lee et al. explores a method for assessing workers' adherence to safety regulations in industrial settings and propose a novel approach based on spatial-temporal graph convolutional networks (ST-GCNs). These networks are designed to analyse both spatial and temporal aspects of data, enabling them to effectively model the complex interactions among workers and their environment over time. This enables proactive interventions,

training programs, and safety improvements to prevent accidents, injuries, and occupational health issues, ultimately fostering a safer and healthier work environment for workers [15].

HAR can use a versatile technology with applications spanning the previous fields and even more. Its importance lies in its ability to automate activity monitoring, provide valuable insights into human behaviours, and facilitate personalized experiences and interventions across various domains. As HAR technologies continue to advance, their impact on technology, healthcare, safety, and quality of life is expected to grow, driving advancements, and shaping the future of human-centric applications and services. In the figure below it is represented a scheme of the steps of HAR (figure 2). However, accurate activity recognition is challenging because of many problems that can occur such as ambiguity in activity patterns, variations in sensor data due to placement and orientation, and the need for real-time processing. Additionally, there are other issues regarding non-scientific problems such as data privacy, user acceptance, and model interpretability.

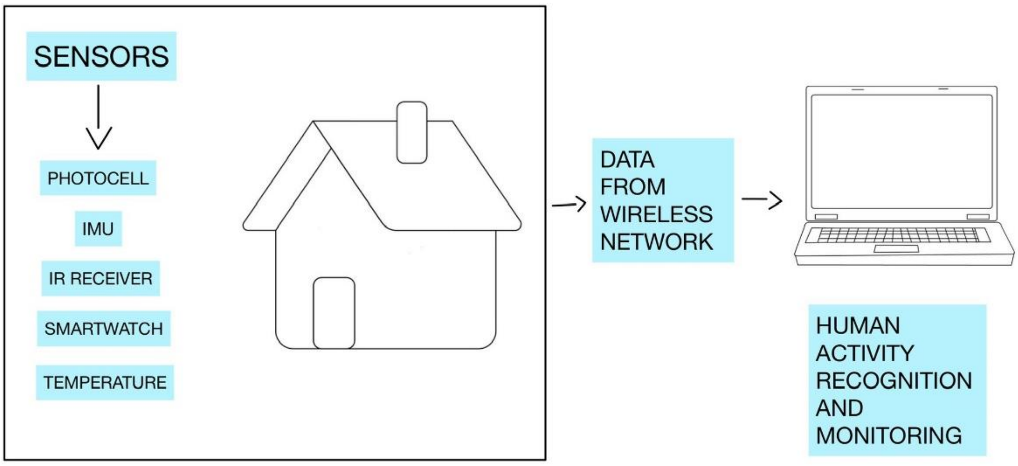


Figure 2: scheme of HAR system in a smart home environment

1.2.2 Motor anomaly detection

There also be the possibility to detect uncommon movements in patients, indeed it is relevant to reveal abnormal biosignals or unusual movements in pathologies such as Parkinson, ET and other kind of motor disease, monitoring people from home, during the daily routine because it can give information to clinician that are crucial to check the current pharmacological treatment and eventually modify it.

Here there are more specific examples of pathologies that can be supervised analysing motor anomalies found in literature: Dennis Khin et al. in their article focus on the importance of assessing instrumental activities of daily living (IADL) in stroke patients. The article discusses various assessment tools commonly used to measure IADL in stroke survivors, such as the Lawton Instrumental Activities of Daily Living Scale or the Nottingham Extended Activities of Daily Living Scale and emphasizes the need for comprehensive evaluation methods that consider the unique challenges faced by stroke patients, including physical impairments and cognitive deficits. The article highlights the significance of accurately assessing IADL to inform rehabilitation strategies and improve patient outcomes post-stroke [16]. On the other hand, regarding PD, gait analysis is of particular interest because it can provide important data on the evolution of the pathology in the patient [17], in fact another common motor symptom of PD is the freezing of gait (FOG), which presents a rapid and frequent shortening of stride length. The lasting of each episode, which is often not continuous, can range between 2 and 30 seconds. During these episodes is possible to notice that the foot or toe does not come off the ground and an alternating leg tremor appears at frequencies of 3-8 Hz and 11-13 Hz [18]. In the following picture is sown a common laboratory designated for gait analysis (figure 3).

Figure 3: room for gait analysis

1.3 Steps of human motion modelling system

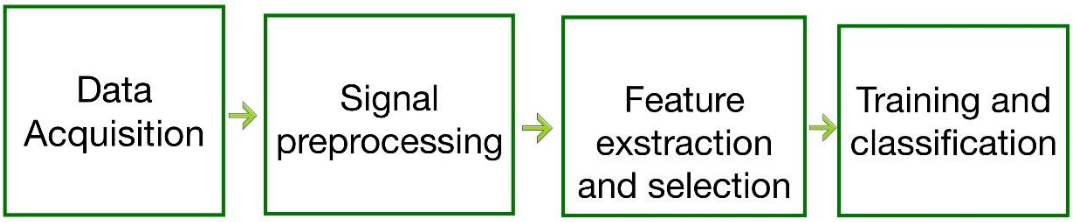


Figure 4: general structure for a human activity classification

The pathway for the classification of human movement has a fairly precise structure that may vary slightly depending on the purpose of its application. A group of studies has developed a typical structure for this pathway by naming it Activity Recognition Chain (ARC) [7]. This structure consists of four main steps:

- Data acquisition: this step involves the collection of data, which can take place via wearable devices such as smartwatches or smartphones or with inertial sensors.
- Signal pre-processing: it includes the analysis of collected signals, noise reduction, possibly window splitting or even the transformation of data into other domains such as frequency.
- Feature extraction and selection: it consists of selecting and obtaining features of signals that can characterize human movement.
- Training and classification: this stage trains the model based on the features obtained in the previous step and performs the classification.

A single technology can conduct many stages in an independent module, or multiple technologies can accomplish the functions carried out by the modules in the ARC in a distinct module.

1.4 Data acquisition: the role of sensors and wearable devices

Human activity recognition can be categorized according to the devices and sensor types utilized, into wearable, video, ambient, and smartphone-based methods. This section provides an overview of the various types of sensors used for data acquisition in human activity classification models. The first ones are inertial sensors, that include accelerometers, gyroscopes, and magnetometers, which are the most common and can be installed in a single IMU or they could be even embedded in smartwatches, smartphones, wrist band, and in these cases are therefore defined as wearable devices.

The rise of wearable devices has provided new opportunities for continuous tremor monitoring and analysis due to their ease of implementation and comfort [19-20]. Inertial sensors can capture fine-grained motion data, allowing for detailed characterization and quantification of tremor episodes. By utilizing these wearable devices, it becomes possible to monitor tremor patterns in real-time and gather objective data on tremor frequency, amplitude, and duration.

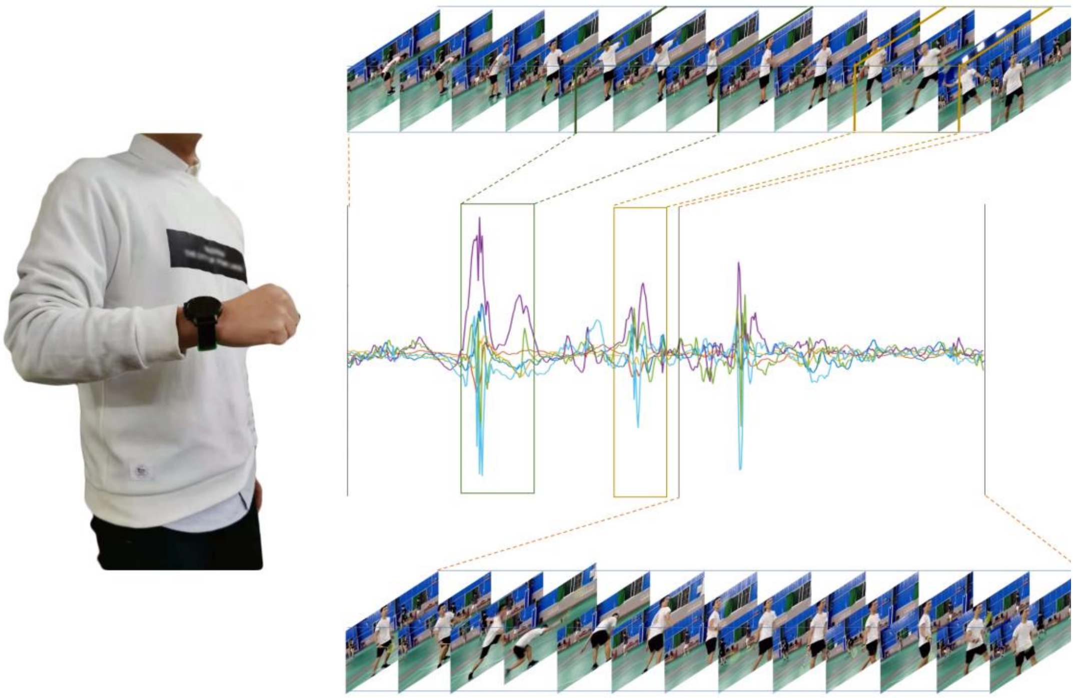


Figure 5: example of IMU signal collected by a smartwatch.

The field of recognizing activities using inertial sensors has experienced significant advancements in recent times, this progress has been mostly driven by the increased availability of wearable devices and the growing demand for personalized healthcare solutions. Activity recognition refers to the process of automatically identifying and categorizing different activities or gestures based on patterns taken from IMU data or other kind of sensors. By combining advanced data processing algorithms with information gathered from inertial sensors and devices that are not uncomfortable for the patient, there is potential for more accurate and sophisticated activity recognition systems. These systems could interpret the motion of patients constantly during the day and allow to record a great amount of data in different conditions, in fact current studies are looking into the possibility of monitoring people with Parkinson's disease at home, who have mentioned how user-friendly the sensors and computer equipment were. Most participants wore the sensors at home for the duration of the trial, indicating that further research on wearable sensors for home-based therapies with continuous activity monitoring in PD patient may be possible [21].

Most of the studies found in the literature that have attempted to achieve this goal use smartwatches, little IMU sensors or smartphone, but there are other kind of methodology, such as video-based methods which utilize sensors that capture images or utilize surveillance camera features to identify daily activities, or ambient approaches that focus on monitoring human-environment interactions by employing sensors attached to smart objects within the environment, including sound, pressure, temperature, and other crucial indicators, particularly beneficial for elderly monitoring. Additionally, the use of smartphone sensors has seen significant attention in recent years. Smartphones, being widely accessible devices, boast an array of sensors such as accelerometers, GPS, gyroscopes, and microphones, facilitating comprehensive health tracking, indoor positioning, and pedestrian navigation [22].

Then, there are physiological sensors, which therefore provide physiological signals, that allow detecting anomalies in body function. Some of these are in standalone devices while others can be installed in wearable devices. Some examples are:

- Temperature sensor: that provides body temperature, usually in Celsius degrees (°C).
- Blood volume pulse (BVP) sensor: it measures cardiovascular dynamics detecting changes in arterial translucency and it can be obtained by Photoplethysmography (PPG). PPG converts the wave-like motion of the blood through the vessels into an electrical signal, in the range of millivolts, which synchronizes with the heartbeat [23].
- Hearth sound sensor: they collect data about noise generated by the hearth beating and the consequent flow of blood through it. The output signals are usually in Volt.
- Electrodermal activity (EDA) sensor: It monitors the variations in conductivity caused by an increase in sweat gland activity in the skin, that is linked with the sudomotor nerve activity and so with the emotional state of the patient [24].
- Electroencephalogram (EEG): is a non-invasive method employed to discern human behaviours by monitoring the neurological responses during cognitive and motor tasks [25].
- Electromyograph (EMG): it measures the electrical activity produced by skeletal muscles at rest and during contraction. This test is commonly used to evaluate muscle and nerve function and can help diagnose conditions such as muscle disorders, nerve disorders, and neuromuscular junction disorders [26-28].

1.5 Signal pre-processing

Each sensor often records and saves signals in different format or stores them in various file-type distribution, in addition all signals usually contain noise or even missing data, so it's fundamental to edit and clean the raw data. The most common techniques are:

- Time interpolation: is widely used in literature to resample signals [1,7,11, 29, 30], and involves estimating values of a signal at points in between known data points, enhancing temporal resolution for accurate analysis. This technique aids in capturing finer details and smoother representations of signals over time.
- Filtering and denoising: used to clean data from noise or unwanted information and improve signals quality, selecting the range of frequencies [30, 26]. However, sometimes it could cause some information loss, so it's not always used [31, 32]. On the other hand, the possibility of losing signal information is precisely exploited according to the study's objectives, as in the case of Gallego et al., who use various filtering parameters to separate voluntary movement components from involuntary ones in patients with essential tremor. For instance, they use a low pass filter with cut frequency of 2 Hz to extract the voluntary action because tremors are reported to occur at higher frequencies, in general 3-12 Hz [33]. (Fig 6)

To evaluate and handle shorter data elements, most of the human motion modelling and recognition systems [1,19-21, 34, 35] have divided the raw data into overlapped windows of few seconds. This method is known as sliding windows. A window is made up of a subset of samples and this sample selection slides over the entire sequence, with or without a defined overlap. An illustration of a 10-second sliding window with a 2-second shift (or step) across a time-domain signal is shown in Figure 7. Every window receives a classification result from the recognition system. since this is a very impactful aspect on the final result and there is no definitive protocol on which to base the choice of windowing parameters, it will be analysed in more detail in section 3.4.

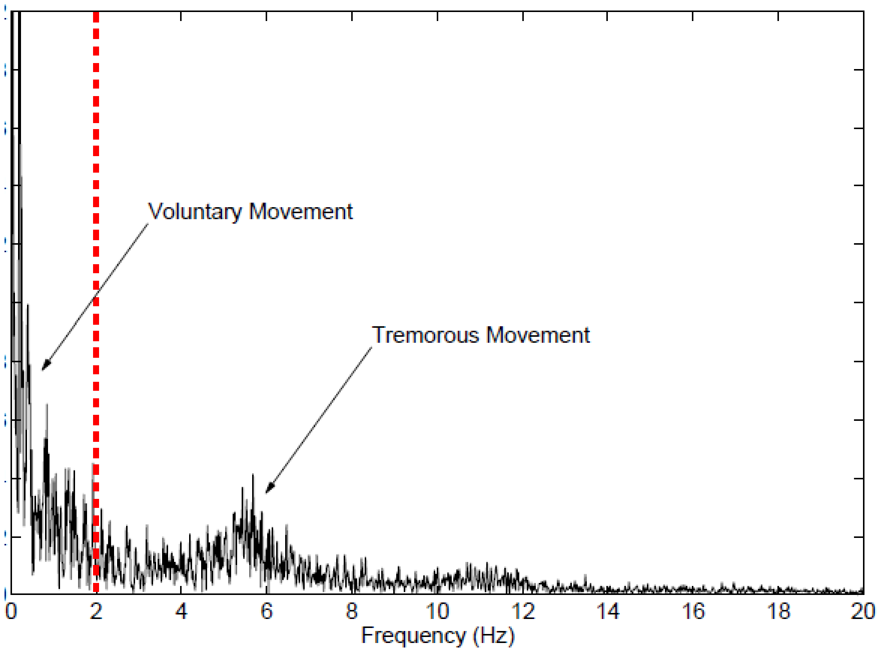


Figure 6: PSD of a signal, the red line discriminates between voluntary movement on the left side and unvoluntary movement on right side.

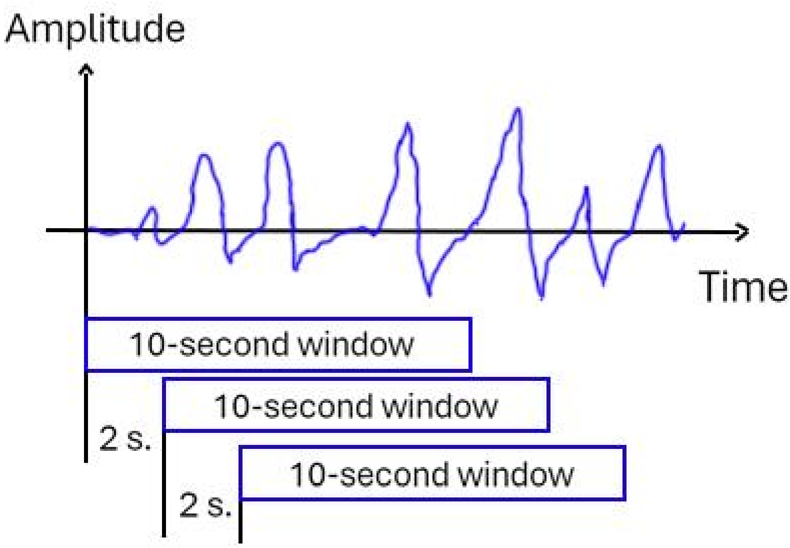


Figure 7: a representation of a 10s overlapping window through a signal.

1.6 Feature extraction and selection

Once each signal is pre-processed and segmented, the next step is to study, select and calculate which features are the most significant and this usually involves both frequency and time domains. The most common features in time domain are:

- mean: signal's DC component.
- median: represents the middle value of the signal's samples when they are arranged in ascending or descending order of their time stamps.
- median absolute deviation: median of the absolute deviations from the median of the signal.
- maximum: signal's highest value.
- minimum: signal's lowest value.
- standard deviation: measure that determines how much the values of a signal differ from the mean value of the signal.
- variance: average of the squared differences from the mean value.
- Empirical Cumulative Distribution Function (ECDF): is a plot that shows the proportion or cumulative probability of observing a value less than or equal to a given data point in a dataset, providing insight into its distribution.
- Signal Magnitude Area (SMA): It quantifies the overall magnitude of a signal over a specific time period by computing the area under the absolute value of the signal curve.
- energy: It is computed by integrating the squared magnitude of the signal over time and refers to the total amount of power contained within a signal.
- Root Mean Square (RMS): square root of the arithmetic mean of the squares of a set of signal values.

The most used features from the frequency domain are the following:

- indexes of specific frequency components.
- mean frequency: weighted average of the frequency components.
- Skewness: measure that indicates asymmetry of frequency distribution comparing to a Gaussian distribution.
- Kurtosis: measure of whether the frequency distribution is light-tailed or heavy-tailed comparing to a Gaussian distribution.
- energy of specific frequency bands.
- Power Spectral Density (PSD): power distribution carried by signal frequencies.

In literature different combination of these features have been used for activity recognition. Some authors, to achieve this, extract just few features, like Serrano et al. who calculate the difference between the final and initial value of the window, the average value, and the standard deviation over the window [1], or Ravi et al. which estimate mean, to have the DC components of the signal, standard deviation to discriminate different activities by the range of acceleration values, such as walking against running, energy to capture the periodicity of the signal and correlation between activities to differentiate between activities that involve translation in just one direction or more [32]. Reiss et al. in addition to these features use power ratio of two frequency bands and peak frequency of PSD to detect cyclic activities [36]. However, there are some authors that decided to consider a much greater number of features, for example, Martins et al. or Vavoulas et al. have extracted all the aforementioned features plus slope, total angular change, fluctuation frequency, waveform length of velocity, spectral centroid and roll-off and many others [29, 35]. In conclusion, most of the studies have good and comparable results unless all these differences in features selection, and due to this great variety of investigation processes this research field is always evolving.

1.7 Training and classification: the role of Artificial Intelligence

As reported in chapter 1.4, there is a great variety of sensor types and the various studies differ greatly in how the data is collected, but what they have in common is the use of artificial intelligence (AI) algorithms in processing this data [1, 7, 10, 12-15]. Al systems attempt to accomplish activities that normally call for human intellect to model and identify particular patterns. The machine learning (ML) methods that underpin this Al's implementation let machines to learn from data. More precisely, deep learning (DL) has been defined by Goodfellow et al. [37] as a special sort of machine learning that achieves significant flexibility by representing the data as a tiered hierarchy of abstract concepts and their link to simple ones. The Venn diagram in the figure below depicts how deep learning is a subset of machina learning, which is used for many but not all AI approaches (figure 8).

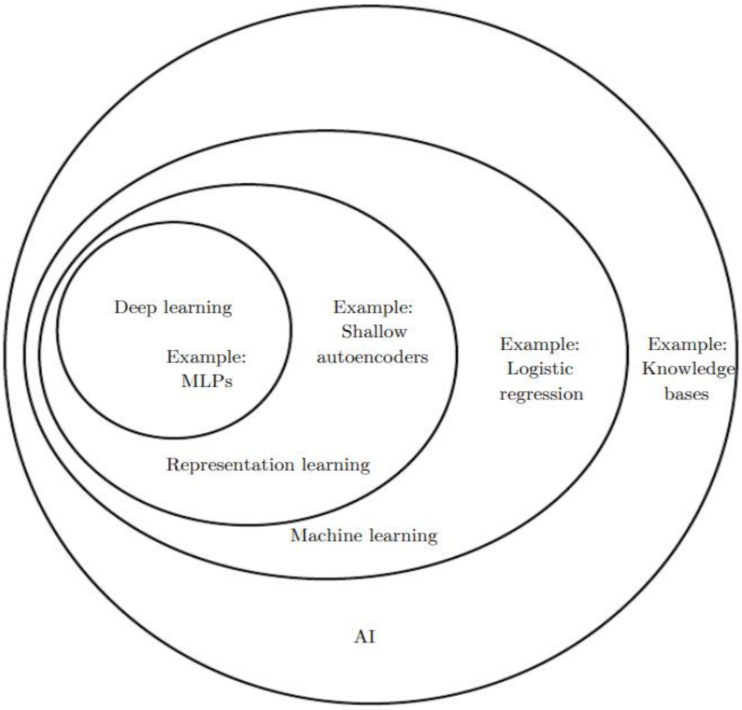


Figure 8: A Venn diagram showing AI concepts hierarchy.

The rise of AI in the last few years has been remarkable and found its way into various aspects of our lives and industries due to:

- Great advancements in machine learning and deep learning algorithms,
- The proliferation of Big Data for training,
- The increase of computing power that accelerated training and deployment of AI models,
- Applications in various fields from healthcare and robotic system to industries applications.

More specifically, the impact of AI in healthcare and research is profound and multifaceted and it is transforming how to approach medical diagnosis, treatment, research, and healthcare delivery, ushering in a new era of precision medicine.

AI algorithms have demonstrated remarkable capabilities in medical imaging interpretation, enabling more accurate and efficient diagnosis of various diseases such as cancer, cardiovascular disorders, and neurological conditions. Machine learning models trained on vast datasets can detect subtle abnormalities in medical images with unprecedented accuracy, facilitating early detection and intervention. For instance, Mitsala et al. affirm that AI tools can at least match or even exceed human performance for colorectal cancer detection and diagnosis [38]. In medical research area, AI has unleashed unprecedented opportunities for accelerating the pace of scientific discovery and translating insights into clinical applications. By harnessing the computational power of AI algorithms, and its ability to analyse great amount of complex data, researchers can analyse vast and heterogeneous datasets with remarkable speed and precision, unlocking valuable insights into disease mechanisms, treatment responses, and therapeutic targets. Furthermore, chemical compound design and optimization are facilitated by predictive modelling, which also helps forecast protein structures and enables the selection of medication candidates with a higher chance of success [39].

Instead, more precisely for this study, the integration of artificial intelligence and machine learning techniques with activity recognition using IMUs has further enhanced the accuracy and efficiency of tremor monitoring systems. These techniques enable the development of algorithms and models that can learn from data, recognize complex patterns, and make accurate predictions. By leveraging the power of AI, it becomes possible to automate the classification and recognition of ADLs and facilitate the analysis of tremor patterns in a more precise and timely manner [21].

HAR systems typically use a machine learning algorithm to learn motion patterns and report classification or detection assessments after the feature extraction process. Conventional system designs take separate modules into account for categorization and feature extraction. Deep learning algorithms, on the other hand, can be used to develop system designs in which both modules are merged in the same algorithm, allowing the learning and classification of motion patterns and pertinent attributes. The figure below shows the different pathway in case of a deep learning algorithm, comparing to figure 4 which consider a general machine learning algorithm (figure 9).

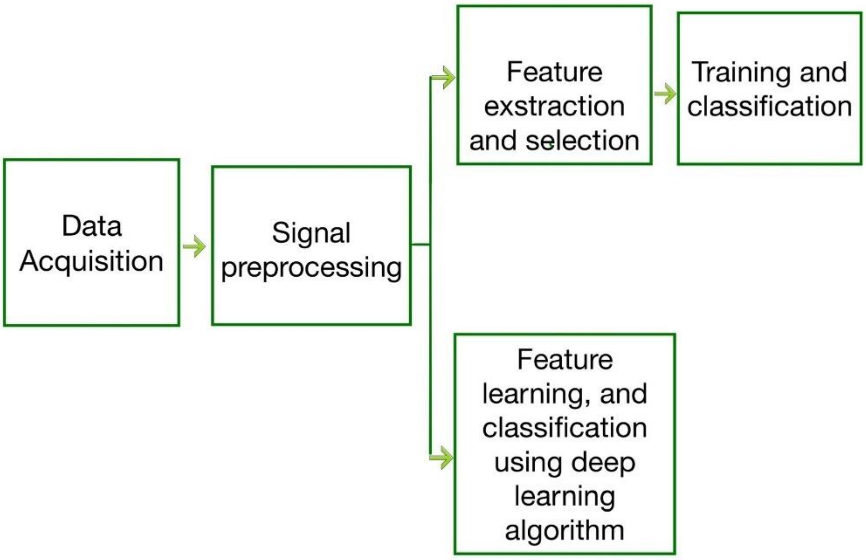


Figure 9: feature extraction and training and classification modalities.

In the subsequent sections will be illustrated different algorithm of ML and DL common algorithm and which one are used by researchers for activity recognition.

1.7.1 Machine learning algorithms

The most common are:

- Decision Trees: These are hierarchical tree-like structures where internal nodes represent features, branches represent decision rules, and each leaf node represents the outcome. They're used for both classification and regression tasks.
- Random Forest (RF): An ensemble learning method that builds multiple decision trees and merges them together to get a more accurate and stable prediction. It's known for its robustness and ability to handle large datasets with high dimensionality.
- Support Vector Machines (SVM): Effective for classification tasks, SVM tries to find the hyperplane that best separates different classes in the feature space while maximizing the margin between them.
- K-Nearest Neighbours (KNN): A non-parametric method used for classification and regression. It assigns a class or value to a new data point based on the majority class or average of its k nearest neighbours in the feature space.

- Naive Bayes: Based on Bayes' theorem, it's a probabilistic classifier that assumes independence between features. Despite its simplicity, it's quite effective for classification tasks, especially in text classification and spam filtering.
- K-Means Clustering: An unsupervised learning algorithm used for clustering similar data points into groups or clusters based on their feature similarity. It partitions the data into k clusters, where each data point belongs to the cluster with the nearest mean.

These are widely used in activity recognition, Leightley et al. analysed both SVM and RF to classify human activity recognition for physical rehabilitation with Kinect video sensors and so tasks like, jumping, walking, squats, arms movement stating that RF better classified six out of this ten activities, but SVM was better in the other four [10]. Reiss et al. compared results of SVM, DT, RF and KNN collecting data with IMU sensors and trying to identify few activities including running, lying and cycling [36]. Serrano et al, classify ten ADL with four IMU sensors with a Naïve Bayes model [1]. Jethanandani et al. developed a decision tree-based model for human activity recognition in smart home environment reaching a high accuracy [14].

1.7.2 Deep learning algorithms

The most common DL algorithms are:

- Artificial Neural Networks (ANN): Inspired by the biological neural networks of the human brain, ANNs consist of interconnected layers of nodes (neurons). They are capable of learning complex patterns and relationships in data through a process called backpropagation.
- Convolutional Neural Networks (CNN): Particularly effective for image recognition and computer vision tasks, CNNs are designed to learn spatial hierarchies of features automatically and adaptively from the input data. They use convolutional layers to detect patterns in local receptive fields of the input images.
- Recurrent Neural Networks (RNN): Suited for sequential data such as time series, speech, and text, RNNs have connections between nodes that form directed cycles, allowing them to exhibit dynamic temporal behaviour. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are popular variations of RNNs that address the vanishing gradient problem and capture long-term dependencies in sequences.

Pascual-Valdunciel et al. tested a LSTM network to predict pathological tremor signals collecting data with IMU and a motion capture system, resulting in the

ability to capture tremor variability within the input signals and adapting the output signals to changes in tremor amplitude or phase [40]. Ordóñez et al., also proposed a LSTM model but for human activity recognition, using wearable devices and outperforming by 4% deep non recurrent network used in other work with the same dataset [30]. On the other hand, a CNN based classifier is presented by Chen et al. that reach an accuracy of 93,8 % in recognizing human activity including falling, jumping, walking [41].
Pascual-Valdunciel, et al. also worked on a comparison between deep learning LSTM model and different machine learning (KNN, RF, SVM) in order to provide a binary (Tremor; No Tremor) classification of kinematic and electromyography signals recorded from ET patients and healthy subjects. ML models reached a high classification performance but in general LSTM gave better results [27]. However, is not always the more complex is a model the more efficient it is, in the decision of which classifier use other parameters are involved, such as computational costs, final goal of the study and more.

In summary, in recent years, there has been notable progress in classifying the daily activities of individuals who do not experience tremors, including recognizing posture changes and activities related to motion [33] and all these factors offer a promising avenue for the continuous monitoring and analysis of tremor in PD and ET patients. Despite this progress, the challenge of applying these methods to individuals with movement disorders, who often exhibit pronounced tremors, remains a significant obstacle, especially if we move to real-time classification.

Chapter 2
Methodology

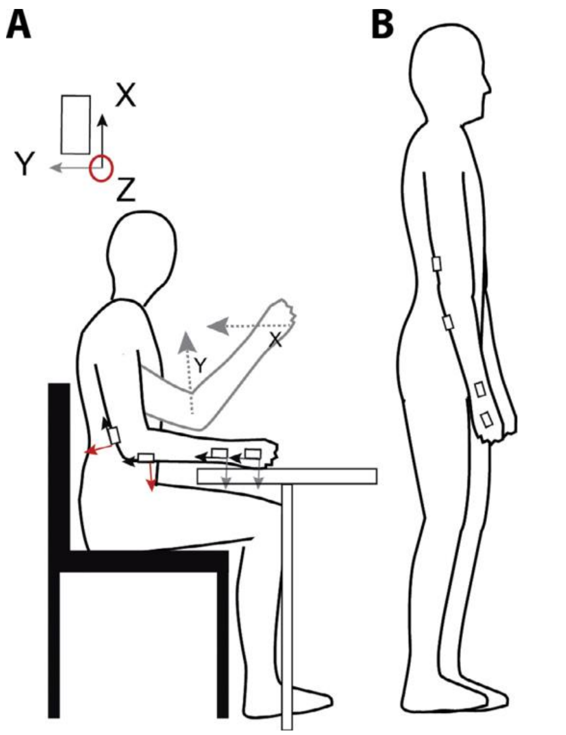


Figure 10: Location of IMU sensor for data collection.

The first part of this thesis work is based on the study of the recordings made in a previous research by Serrano et al. [1]. They collected data from four IMUs along the arm and classified a series of fine and gross movement. More specifically these IMUs were placed to the dominant arm of each participant to track upper limb movement: two IMUs on the back of the hand and the distal third of the forearm to measure wrist movement, two additional IMUs on the dorsal side of the forearm and on the distal part of the arm, to track elbow movement as shown in the picture on the left (figure 10).

Moreover, this thesis proposes a new method to try to improve the accuracy of the activity classification despite the utilization of data derived from just one of the four IMUs (the one on the distal forearm), in order to make a step forward through the continuous monitoring of tremorous movement in PD and ET patients.

2.1 Dataset

The dataset provided contains acceleration and angular velocity records in all three axes (x, y, z) of 16 patients whose gender and age were not provided, nor specified how many were ET and how many were PD patients. Each signal was acquired during a protocol that included the eleven following actions:

- Combing hair (CB)
- Buttoning the buttons of a lab coat (BB)
- Cutting a fake steak (CE)
- Eating the previously cut pieces with a fork (EF)
- Simulate drinking (SD)
- Opening and closing a Tupperware container (OT)
- Turning 3 pages in a book/magazine (TB)
- Printing their name/signing a document (SN)
- Simulate tooth brushing (TB)
- Turning doorknob (TD)
- Resting arms on table (RE)

Each task was performed between 3 and 6 times, besides RE that has been performed once, and each signal has been sampled at 100 Hz and then resampled at 1 kHz [1].

2.2 Preprocessing and filtering

For this purpose, "Python" environment has been used. The data were in ". mat" format and it has been necessary to import them in python and organize them in a data frame ordered by patient, task, trial, and IMU (figure 11).

	id_patient	id_task	trial	timestamp	IMU0_gyro_x	IMU0_gyro_y	IMU0_gyro_z	IMU0_acce_x	IMU0_acce_y	IMU0_acce_z
0	1047583	task_BB	0	0.001	0.010143	0.037228	0.082473	-9.607763	0.137203	-1.550558
1	1047583	task_BB	0	0.002	0.010171	0.037707	0.082041	-9.605139	0.134317	-1.551275
2	1047583	task_BB	0	0.003	0.010201	0.038184	0.081594	-9.602500	0.131408	-1.551969
3	1047583	task_BB	0	0.004	0.010233	0.038659	0.081129	-9.599844	0.128477	-1.552640
4	1047583	task_BB	0	0.005	0.010265	0.039132	0.080648	-9.597173	0.125524	-1.553288
...	...	...	...	...	...	...	...	...	...	...
9511495	944363	task_TD	4	6.496	-0.427350	0.007364	-0.148436	-9.267541	-0.299038	-1.764390
9511496	944363	task_TD	4	6.497	-0.423515	0.007606	-0.150584	-9.272721	-0.293414	-1.766366
9511497	944363	task_TD	4	6.498	-0.419689	0.007843	-0.152701	-9.277904	-0.287778	-1.768340
9511498	944363	task_TD	4	6.499	-0.415873	0.008075	-0.154786	-9.283088	-0.282133	-1.770313
9511499	944363	task_TD	4	6.500	-0.412066	0.008302	-0.156840	-9.288273	-0.276478	-1.772284

9511500 rows × 16 columns

Figure 11: Example of some rows of the data frame.

Going deeper in the signals analysis it was noticed that, calculating the power spectral density (PSD) the higher valuable frequency for every signal was about 3 Hz, so it was possible to state that they have already been filtered with a low pass filter with cut frequency of about 3 Hz. This is a common practice, as it has been already shown in paragraph 2.5, to separate the voluntary component of the movement from the tremorous one. This aspect was also confirmed by Eduardo Rocón, one of the authors of the study.

After discovering that there were some missing files calculated the timestamp of each sample for every signal, it was necessary to search for outliers, or meaningless one. So, it has been analysed firstly, the length of each signal and comparing for all the tasks the time distribution, every signal that was too far from the distribution centre has been discarded, with particular attention to the shorter signals that could represent mistaken recordings. Finally, a total of 12 signals was not taken into account.

Following an empirical and visual analysis of the various signals, the rest task has shown in most of the recordings a little portion of time, at the beginning or at the end, with much greater peak than the average value. These portions of the signal were considered as measurement errors due to a late start of the resting phase by the patient or incorrect movements during the recording phase (fig 12). As these portions were very small compared to the length of the signal and the amplitudes of the accelerations and angular velocities were incompatible with values due to simple tremor phase, it has been decided to cut the signals in these regions.

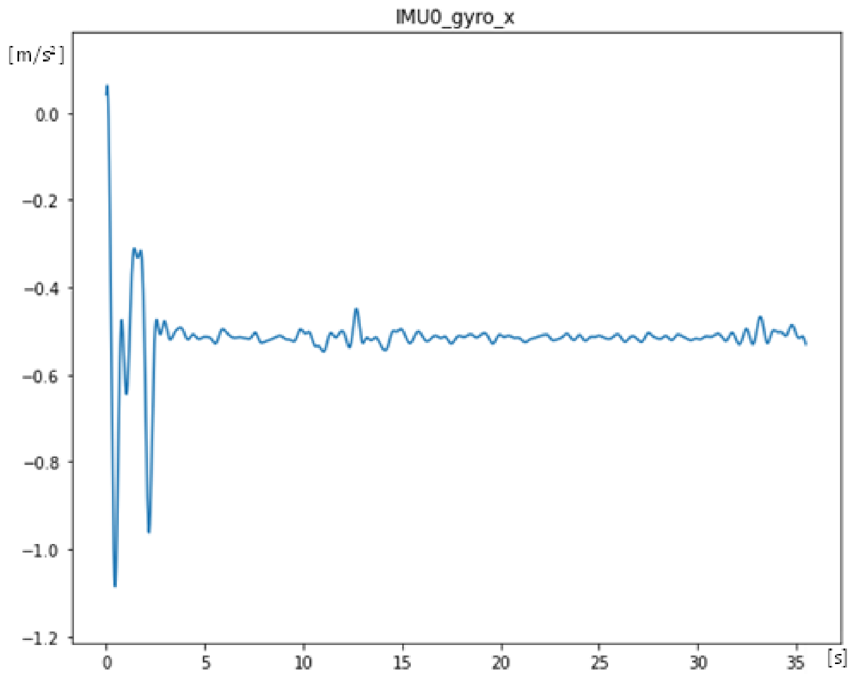


Figure 12: example of a starting error in a signal.

2.3 Feature selection and feature extraction

Since the work has started it was chosen a ML approach for the classification and so the feature extraction became a necessary step. This choice was made in a first instance for lower computational costs, secondly to have more control over the features, study which ones were the most impactful on the final classification.

The choice of which features could be the most useful was first developed by assessing which were the most used in the various studies that are involved in activity recognition using IMU data. So, the following were chosen in the time domain [29, 32, 35, 36]:

- Mean value
- Standard deviation
- Median
- Maximum value
- Minimum value
- Variance
- Root mean square (RMS).

To this set, another feature has been included that calculates the difference between the first and the last value of the window samples to find a slope, a sinusoidal pattern or great decrease either increase inside the window.

Concerning frequency domain, looking at the PSD of the signal, they have shown quite the same behaviour due to the filtering made by the previous study. However, some frequency features have been tested also because of the next step of the classification with new data collection, so the calculated ones are:

- Entropy
- Spectral flatness
- Centroid
- Spectral bandwidth
- Spectral roll-off frequency
- Skewness
- Kurtosis

Therefore, it was assumed that the calculation of features in this domain would not bring additional information and would not improve the classification, so it was decided not to consider them.

The feature set did not undergo feature selection or reduction, resulting in the utilization of all features for each of the classifiers outlined in the subsequent section. The same approach has been done by Reiss et al. too [36], but it is not so common, in fact usually a PCA or other feature selection technique are involved in the process [10, 29, 35].

2.4 Signal segmentation

Analysing the duration of the signals, it has been noticed that each of the signals have different length also between different trials for the same patient doing the same task. In addition, most of the studies that focus on activity recognition, divide the raw signals into overlapped windows, in order to process short fragments of data. This approach consists of sliding a window through the data, calculate all the features and report a classification output for all the windows.

Every work has tried different window sizes based on different sampling frequency: Serrano et al. in their work with this dataset used windows of 0.5 s with no overlap, and since this work is focused on outperforming their result, it has been searched in the literature a different type of segmentation. Hasan et al., for example, uses a 2.56 s window with 50% overlap in a ML algorithm that aims to detect human activity with IMU data [42]. Another option derives from Hussain et al. which segmented their dataset in 10s sliding window with 1s overlap to classify human activity using EEG sensors [25]. As last example reported, Reiss et al used a 5,12s window with 1s shift because they wanted to have, basing on their dataset, at least three segments for each periodic activity [36].

As it possible to see, there is still a lack of defined parameters of the segmentation because they can depend on numerous factors such as the kind of activity to recognize, the kind of raw signals obtained and their number of sample or also the morphology and the type of preprocessing methods used for the data. These and other study are resumed in the table below for a clearer overview (table 1).

STUDY	WINDOWS/OVERLAP	SENSORS	SAMPLING FREQUENCY
Serrano et al. [1]	0.5s/ -	IMU	100 Hz
Hasan et al [42]	2.56s/1.28s	Smartphone	50 Hz
Hussain et al. [25]	10s/1s	EEG	n.s.
Reiss et al. [36]	5,12s/1s	IMU	100 Hz
Martins et al. [29]	0.5s, 1s, 1.5s ,2s/ 80 %	IMU	50 Hz
Dernbach et al. [43]	1s, 2s, 4s, 8s, 12s, 16s / 50 %	Smartphone	80 Hz
Siirtola et al. [44]	7.5s/ -	Smartphone	40 Hz
Ahmadi et al. [45]	10s / -	Accelerometer	30 Hz

Table 1: windows, overlap and respective sampling frequency and sensors used in different studies.

So, this work, in a first instance, tries to analyse different type of segmentation, comparing the results to find the window size that gives the best classification.

The chosen window sizes and their overlap are the following:

- 2.5 seconds with 1 second of overlap
- 5 seconds with 2 second of overlap
- 10 second with 4 second of overlap
- 15 second with 6 second of overlap
- 20 second with 8 second of overlap

Hence, a data frame for each kind of segmentation has been developed, and the features has been calculated for each window. Finally, the value obtained have been normalised with the function “MinMaxScaler” of “Sklearn” library in Python. That’s why different features in a dataset may have different scales, in fact features with larger scales may dominate those with smaller ones, leading to biases in the model. In addition, normalization often results in a better-conditioned optimization problem, allowing the algorithm to reach the optimal solution more quickly, and so have a reduced computational cost.

2.5 Classification approach

Once the features were calculated for the five window types, the values were saved in five different data frames, each for each window size. For classification, has been decided to use two different machine learning algorithm, Random Forest and a Support Vector Machine (Figure 13).

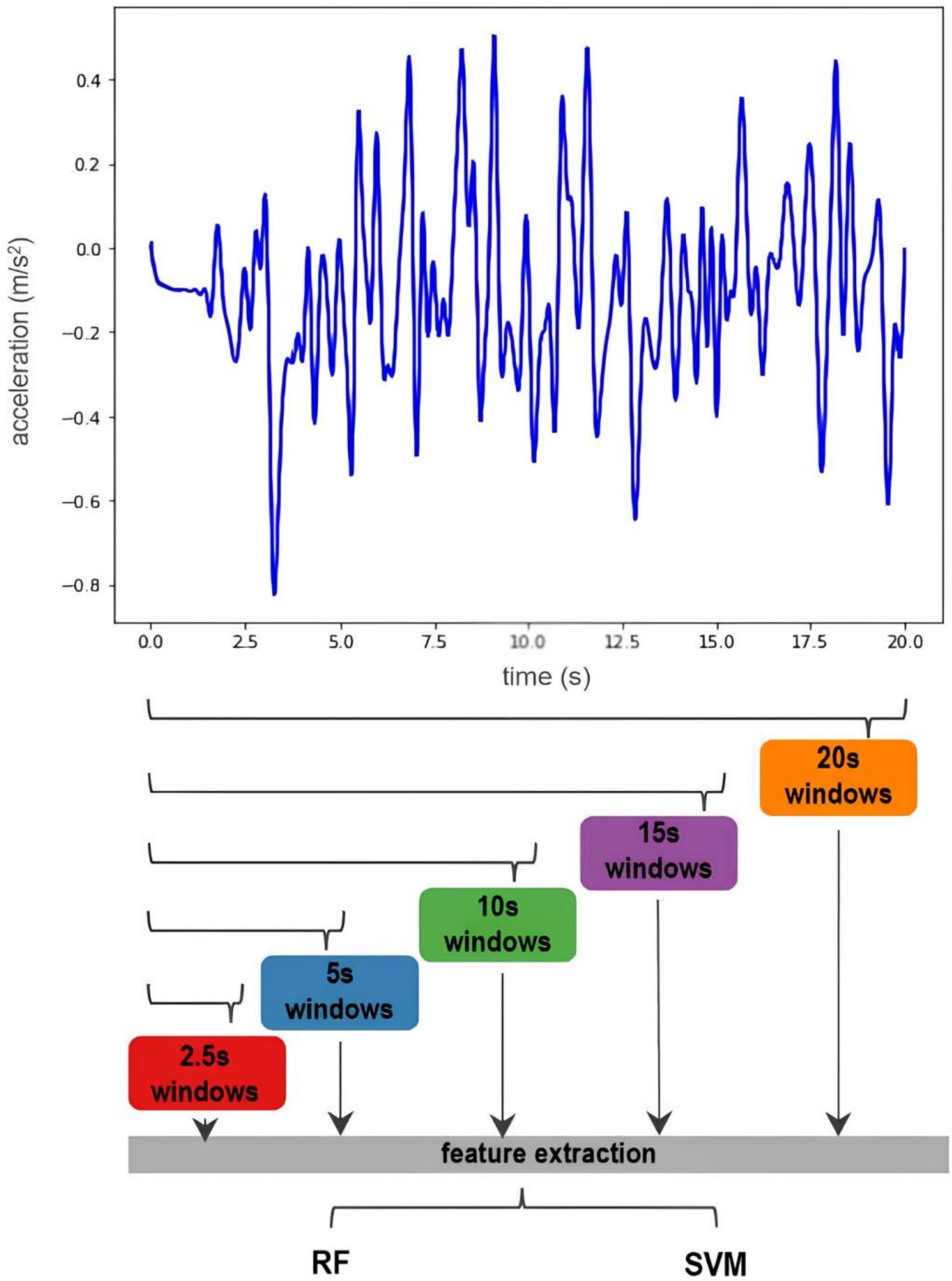


Figure 13: representation of steps described with an example of a signal.

For an approach like this, the splitting of the data set in train set and test set is crucial. The choice of the splitting method can impact the performance and generalization ability of the model and it’s important to examine many aspects, such as:

- Nature of the Data: the characteristics of the dataset and the importance of preserving inherent groupings or structures also after the splitting.

- Goal of the Analysis: and it has been valuated as more important how well the model generalizes to new patients instead of new instances of the same task.
- Data leakage: splitting the data in a way that tasks from the same patient appear in both the training and testing datasets, could lead to data leakage and so to overly optimistic evaluation results.
- Stratification: Since having multiple tasks, it's important to ensure that each one is represented proportionally in both the training and testing datasets, because this helps in avoiding bias towards certain tasks during model training.

Furthermore, it had to be considered that if the data set was separated by task, the network would be trained to search for a recognisable pattern for each patient, and would cause overfitting on the patient itself, rather than on activity recognition, or if it had been divided by trial, data from the same patient would have been present in both train and test set.

Finally, it has been chosen to split up the dataset in test set and train set with a proportion of 30-70% respectively, which corresponded to a division of 5-11 patients. Then all the data has been randomized before proceeding with the training phase. So, at last, has been developed 2 models, each with the 5 different types of windows.

2.6 Results

To evaluate the performance of a classification algorithm is crucial to understanding how well it is performing on every given task. There are several metrics that can be used to assess the classification performance of an algorithm. The terms that will be present in the definition of these metrics are the following:

- True Positive (TP): the number of instances that were correctly predicted as positive by the classifier.
- True Negative (TN): the number of instances that were correctly predicted as negative by the classifier.
- False Positive (FP): the number of instances that were incorrectly predicted as positive by the classifier when they were negative.
- False Negative (FN): the number of instances that were incorrectly predicted as negative by the classifier when, they were positive.
- Total Instances (TI): total number of instances

$$TI = TP + TN + FP + FN$$

The metrics are:

Accuracy: Measures the overall correctness of the classifier. It's the ratio of correctly predicted instances to the total instances.

$$\frac{TP + TN}{TI}$$

Precision (Positive Predictive Value): the ratio of correctly predicted positive observations to the total predicted positives. It measures how many of the predicted positive instances are positive.

$$\frac{TP}{TP + FP}$$

Recall (Sensitivity, True Positive Rate): the ratio of correctly predicted positive observations to all observations in actual class. It measures the ability of the classifier to capture all the positive instances.

$$\frac{TP}{TP + FN}$$

F1 Score: the harmonic mean of precision and recall. It provides a balance between precision and recall.

$$2 \cdot \frac{PRECISION \cdot RECALL}{PRECISION + RECALL}$$

The latter has been selected due to his sensitivity to both false positives and false negatives, which is particularly important to maintain a balance between precision and recall.

After numerous trials, it has been noticed that the best classification results were given when the feature selection wasn't applied, with differences of also 5/6 % between these different methodologies and so, due to this reason has been decided to not apply any selection criteria.

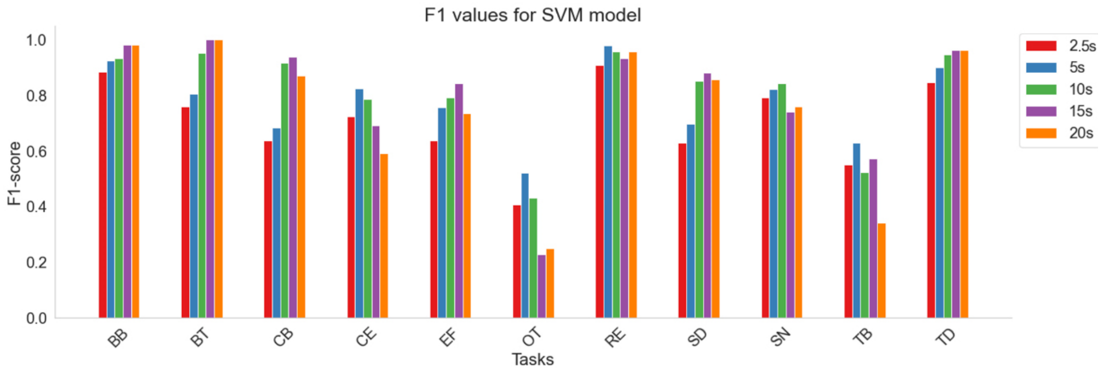
In the following table (table 2) are reported the average results of the classification, based on different metrics, for both RF and SVM for each window type.

Metrics	Algorithm	2.5s	5s	10s	15s	20s
Precision	RF	70,19%	75,74%	81,34%	78,88%	77,92%
	SVM	73,04%	78,94%	80,43%	79,76%	77,76%
Recall	RF	69,57%	76,63%	80,74%	78,59%	77,19%
	SVM	72,14%	77,06%	78,03%	74,16%	74,41%
F1 score	RF	69,05%	75,63%	80,71%	78,38%	76,80%
	SVM	71,42%	77,01%	78,74%	75,82%	75,17%

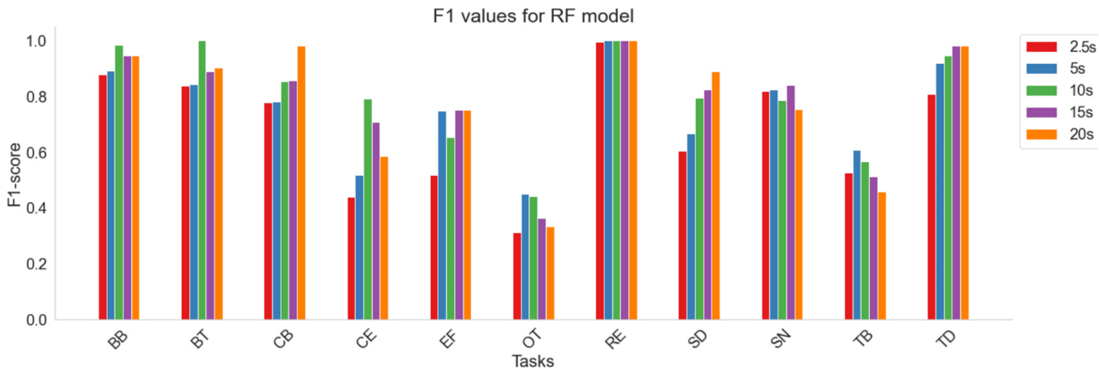
Table 2: average results obtained by the different window sizes for each model.

As can be seen from the table, the results are superimposable for the two approach methods, so it can't be preferred one model over the other. Furthermore, looking at this number, it could be supposed that the 5, 10 and 15 seconds windows are the ones that gives the best results, however it's important to compare the results for each task, as it has been done in the next two graphs (figure 14).

As was to be expected from the results in the table, none of the tasks ranked better using the 2.5 second window than the others. Some activities are recognised more accurately with the 5 seconds window such as OT, TB, and CE in the SVM model. In addition, some tasks such as TD, EF in both models but also, BT and BB in the SVM model are classified better with the longer windows of 15 and 20 seconds. This partly confirms what was written in Chapter 2 which reported that there is no clear consensus on which window size should be preferably used and the most recent advanced techniques for classifying activities of daily living in patients with tremors have not produced definitive evidence so far.



(a)



(b)

Figure 14: a comparison of results of F1 score reached for each task between different window types for both RF (b) and SVM (a).

Ultimately, a comparison has been made between the F1 score achieved by the previous study [1] and this work. In the table below are reported the 10 seconds window, percentage of SVM and RF and the best value of classification.

Tasks	Previous results	10s SVM	10s RF	max SVM	max RF
BB	86,55%	93,15%	97,06%	98,18%	97,06%
BT	85,00%	95,24%	97,78%	100%	97,78%
CB	82,80%	91,53%	88,24%	93,88%	98,04%
CE	52,31%	78,72%	65,96%	82,54%	65,96%
EF	46,40%	79,25%	74,51%	84,44%	78,26%
OT	32,43%	43,24%	36,36%	52,00%	42,31%
RE	-	95,65%	100%	97,92%	100%
SD	83,34%	85,25%	82,14%	88,14%	92,31%
SN	71,43%	84,44%	81,32%	84,44%	81,32%
TB	33,33%	52,46%	62,50%	62,90%	64,00%
TD	92,21%	94,55%	94,74%	96,15%	98,11%

Table 3: a comparison of F1 score between this classification and the previous study.

It is evident that there is an improvement in all the activities and more specifically a much better percentage in tasks as TB, EF and CE which turned out to be among the most difficult to recognise. The results obtained are also explained in the paper published in 2023 with the supervision of Professor Alvaro Gutiérrez from telecommunication department of Madrid polytechnic [46].

Therefore, in view of the encouraging results obtained, and given that an ideal window size had not yet been established, it has been decided to continue the work by devising an algorithm that would select the best window type from time to time. This process is described in the next chapter.

Chapter 3
Algorithm developed

As anticipated in the previous chapter, the best window size for the classification of ADL has not been found yet and especially for this study some sizes give better results for specific task. In this scenario, has been thought to develop an algorithm, which can classify the activities choosing time by time the most appropriate window size and it has been done as for the SVM model, as for RF.

In these chapter two sections are presented, in the first one the algorithm will be described, in the second the results will be shown.

3.1 Algorithm description

When a random forest classifier is working, it makes a prediction based on the probability calculated in the network. For example, for a given window of the signal the probability that the activity is TB is 0.75, TD is 0.45, CE 0.23 and so on, seen this the chosen task for the prediction is TB. This coefficient has been exploited to select the best window step by step.

As it's possible to notice in the figure below (figure 15), the algorithm is divided in two sections:

In the first part, for each type of window, each signal is analysed, and the corresponding probability returned by the RF (or SVM) model is saved for each segment. At this point, several scenarios are possible:

- There is a clear majority of signal segments on the predicted task. that task is selected, and the corresponding average of the probability values is saved. As can be seen in the figure then that there is a majority of green segments, consequently the predicted task type (green) was saved with the corresponding coefficient average value (0.7)
- There is a tie in the number of predicted tasks, so in this case the average of the probability values between the same type of task is performed and

the task with the higher probability is saved. In the pictured case of the 15s window, to discriminate between the green and red signals, the green one was saved as it had a higher value. In contrast, in the case of the 5s window, there is a equality in the number of green and red windows but the red task was selected as it had a higher average value.

In the second part, as input to the algorithm we will have a predicted task for each type of window with the assigned probability value, and the goal is to find the final prediction of the task. To do this, two different paths were possible:

1. Giving priority to the number of predictions: thus, giving the final prediction according to the task predicted the most times.
2. Give precedence to the coefficient: then the final task would be the one with the highest probability.

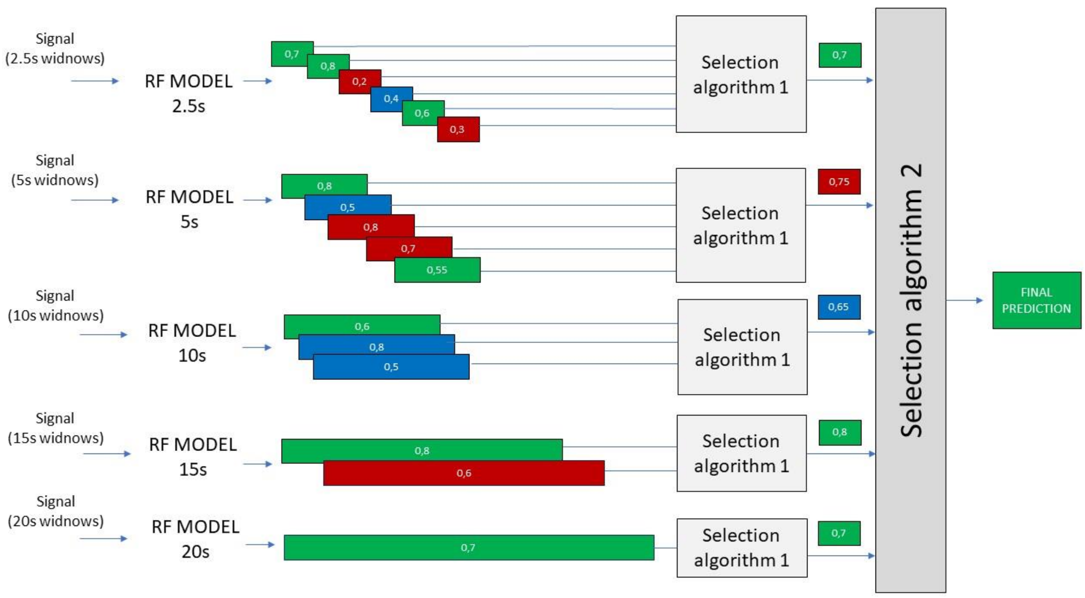


Figure 15: representation of the algorithm developed.

In the end, it was decided to continue with the path outlined in point 1 because one could have presented, for example, the case in which the 2.5 s window would have predicted a task with a high coefficient such as 0.9, and the other four windows all predicted the same task but with a somewhat lower coefficient such as 0.8, and at this point it is much more probable that the prediction of the 2.5 s window is wrong than the prediction of four windows that all give the same output even if with a lower probability so, it has thought that could be the best choice to avoid possible outliers or misclassification

Consequently, given the methodology applied, there can be two cases here as in the first part of the algorithm, and thus in the case of a clear majority that task is chosen, in the case of a tie. There are in turn two scenarios:

- Two windows predict one task and two others another task: at this point, the coefficients are averaged and the task with the highest average is taken as the output.
- All windows predict different tasks: the higher probability coefficient is considered. In the event of a tie between the two highest coefficients (a rare and unlikely but nevertheless possible case), priority is given in order to the window that showed the best results in general in the study carried out in chapter 3. Thus, in order precedence is given to the 10s window then the 15s window, the 20s window, the 5s window and finally the 2.5s window.

3.2 Results

Once the algorithm has been elaborated it had to be tested to compare it with the previous classification. Thus, it has been used with the same dataset and the same splitting. To run the code, it has been used a personal computer with an intel core i7 (8^{th} Gen.) processor which completed the classification and calculation of the results in few seconds, so it does not require high calculation property. In the two figures below the confusion matrixes (with the precision value displayed) for the two algorithms are reported (figure 16 and 17).

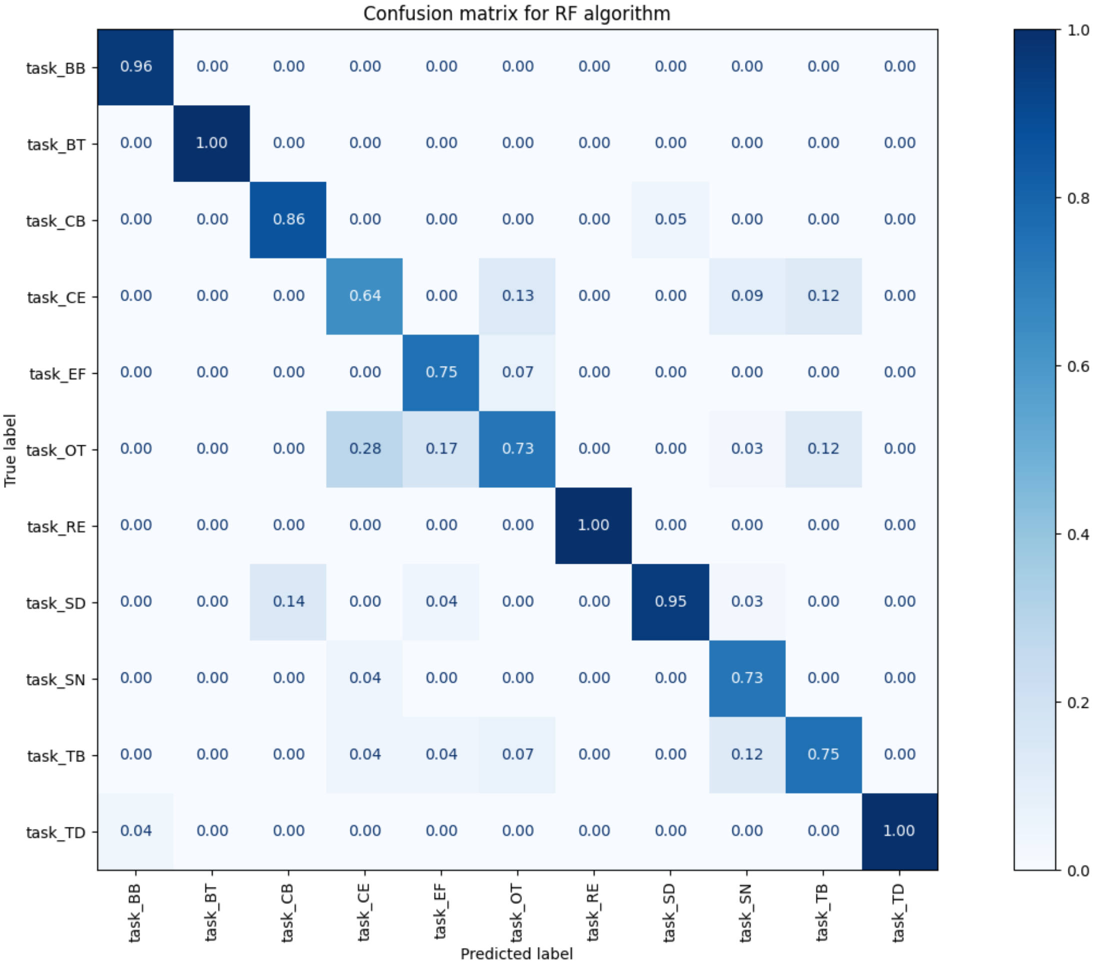


Figure 16: confusion matrix with precision values for RF algorithm.

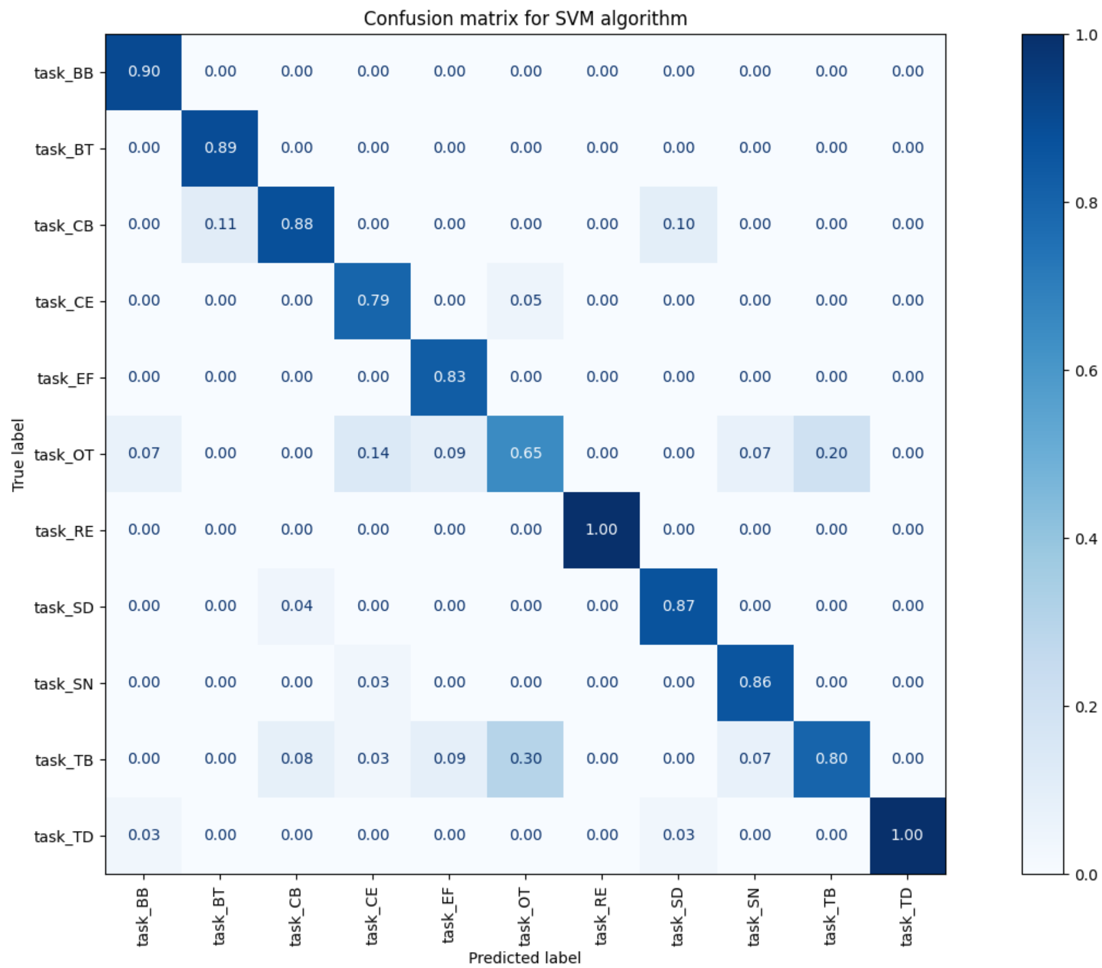


Figure 17: confusion matrix with precision values for SVM algorithm.

The precision values for both models are good and the positive thing is a less variability of the classification, that was one of the first objective. As it was predictable there isn't one model that outperform the other in all the tasks, but there is a balance between them. There is an excellent precision in task RE and TD (100%) for both and quite the same for CB. Then, in one hand, SVM outperform RF in tasks CE, EF, SN and in the other hand RF classifies better BT, OT and SD. Due to this reason, the decision on which classifier adopt for a future mobile app or more in general, recognition system is postponed when these models will be validated with new data.

However, the main objective was to create an algorithm that could choose the best window for each task, time by time, and so it was needed to get comparable results with the best classification rate for each window type. In the next table (table 4) the F1 scores of these models will be compared with the maximum classification results reported in table 3.

Tasks	Previous results	max SVM	max RF	SVM model	RF model
BB	86,55%	98,18%	97,06%	94.74%	98,18%
BT	85,00%	100%	97,78%	94.12%	100%
CB	82,80%	93,88%	98,04%	84,00%	90,91%
CE	52,31%	82,54%	65,96%	86,79%	65,31%
EF	46,40%	84,44%	78,26%	90,48%	83,72%
OT	32,43%	52,00%	42,31%	56,52%	53,66%
RE	-	97,92%	100%	100%	100%
SD	83,34%	88,14%	92,31%	91,23%	85,71%
SN	71,43%	84,44%	81,32%	90,57%	82,76%
TB	33,33%	62,90%	64,00%	60,00%	73,47%
TD	92,21%	96,15%	98,11%	96,15%	98,11%

Table 4: F1 scores of the new models in comparison with the maximum classification value and the previous results.

Analysing task by task the performance of the new model, it is possible to state that:

- Task BB: is very close to the maximum, RF model gives results of almost 100 % even improving the corresponding max RF value by 1% and SVM model returns very similar score.
- Task BT: here there is practically the same situation of the previous one, with RF model returning 100% core and SVM a bit worse but with great result anyway.
- Task CB: is the only activity that has been classified worse than their corresponding max value, however with quite good results and still better than the parallel previous result.
- Task CE: so far, it's one of the most difficult tasks to be predicted, although the RF model reaches the maximum value and SVM model outperforms the maximum by more than 4%.
- Task EF: this activity also results very unpredictable by the previous study. The max values showed that it was possible to enhance its prediction and both new models outperform even the maximum results by more than 6 % respectively.
- Task OT: it absolutely is the most difficult task to be classified, however, the new models have reached a better performance than max SVM and max RF, far exceeding 50 %.
- Task RE: it has been confirmed as the easiest task and it gives 100 % classification rate, which is very important in view of future developments regarding a continuous classification.

- Task SD: for this task RF model does not reach the maximum model but it still better than the other study. However, SVM model outperform the maximum value by more than 3%
- Task SN: both models increase their performance, with SVM exceeding 90 %.
- Task TB: this was the worst classified, as OT, by Serrano et al., in this case the SVM algorithm is almost the same of the maximum, reaching 60 % and the RF increase the result by more than 9 %.
- Task TD: return exactly the same results as their max values.

Given these results, which in general turned out to be excellent both because they reached the maximum classification values obtained by the individual window types and because in many cases, they even exceeded them, it was decided to continue with the validation of the two algorithms by collecting new data.

Chapter 4
Algorithm validation and results

With the aim of validating the algorithm it was necessary to collect new data for testing it. In this chapter will be illustrated the data collection process in the first part, and in the second one the results obtained.

4.1 Data collection

For collecting data, it has been used a Fitbit 2 smartwatch, whose characteristics of interest are resumed in table 5, that was worn by the subjects in their dominant wrist and the axial orientation is shown in figure 16. On this smartwatch and on an android tablet was installed an app, developed by Madrid's CSIC CAR engineers, which allowed to record and collect all the data.

Sampling frequency	30 Hz
Connection	Wi-Fi, NFC, Bluetooth
Sensors	Accelerometer, gyroscope
Bluetooth	5.0

Figure 16: axial orientation.

Since the authors of the previous study did not specify the protocol and how all tasks were made to be performed by the patients, it was decided to draw up a new protocol. This involved each patient being asked to perform the eleven tasks four times consecutively, starting with their hands resting on the table and finishing in the same position. A more accurate description of the task is the following:

- Rest (RE): forearm on the table with palms down for 30s.
- Combing hair (CB): pick up comb from table with the dominant hand, three or four passes over head, leave comb (figure 17).

- Buttoning buttons (BB): while standing up, button and then unbutton a lab coat.
- Cutting a fake steak (CE): out of plasticine, the activity consists of picking up fork and knife, cutting some pieces, putting down fork and knife.
- Eating (EF): pick up the fork, select some of the previously cut pieces, get close to mouth, put down fork.
- Simulate drinking (SD): pick up a bottle of water, bring it to the mouth, hold it to the mouth (with water touching the bottle cap), put the bottle on the table).
- Open and close a Tupperware container (OT): pick up the container, open it, close it, put it down.
- Turn 3 pages in a book/magazine (TB): open the book, turn three pages, close the book.
- Sign your name (SN): pick up the pen, sign with your name, put the pen down (fig. 18).
- Simulate brushing teeth (TB): while standing up, pick up the toothbrush, simulate brushing teeth covering the whole range for some seconds, put toothbrush down.
- Turn doorknob (TD): while standing up, reach for door handle, open door, close door, return to starting position.

This protocol has been performed by 31 volunteer (13 male, 11 female), between 22 and 79 years (average: 29,7y - standard deviation: 13,9y) without any tremor correlated pathology in “Robolabo” laboratory at UPM. In addition, thanks to hospital “Hospital 12 de Octubre” in Madrid, it was possible to record data from 4 PD patients (3 male, 1 female), 4 ET patients (2 male, 2 female). Tremor patients were between 54 and 80 years old (average: 70,6y – standard deviation: 9,9y). All the recordings done in the hospital were videotaped. To conclude, a total of 39 recordings have been added to the starting data set made by 16 patients, so the total amount of the sample is 55 people (average 38,1y – standard deviation: 21,2y). Although, due to a problem with the application, some rest tasks have not been recorded, so in this section all the results will not include this task. However, this does not influence so much the outcomes because rest task was the only one with a 100% precision rate.

Figure 17: a patient wearing the Fitbit smartwatch on the right wrist and combing her hair.

Figure 18: signing task.

4.2 Results

As previously done, all the signals that has been collected have been pre-processed. Firstly, they have been filtered with a Butterworth low-pass filter of 6^{th} order with cutting frequency at 4 Hz, to remove the unvoluntary component of the movement. Then, resampled at 1000 Hz to bring them in the same condition as the signals of the previous study. Finally, each one has been segmented in all the five-window type, and all the features extracted. The dataset has been divided by patients, thus the 30% (17 out of 55) of patients have been included in the test group and the rest formed the train group. As in the previous chapter, both algorithms have been tested and in the next two images, the confusion matrixes, displaying precision rate, are shown (figure 19-20).

Looking at the precision values in both matrixes, it can be confirmed that task CE and OT are the most difficult to predict, but here is a performance enhancement in these cases. Moreover, based on these predictions with this dataset and considering all the parameters set for this study, it seems that RF model gives slightly better results than SVM, there is less variability between the classification of different activities, with just two out of the total which have less than 80 % rate (OT and CE), while SVM has OT less than 70 % and two others less than 80 %, with task SD very close to this threshold. Also, the average precision of RF is better than SVM, 88% versus 83%, even though the difference remains subtle.

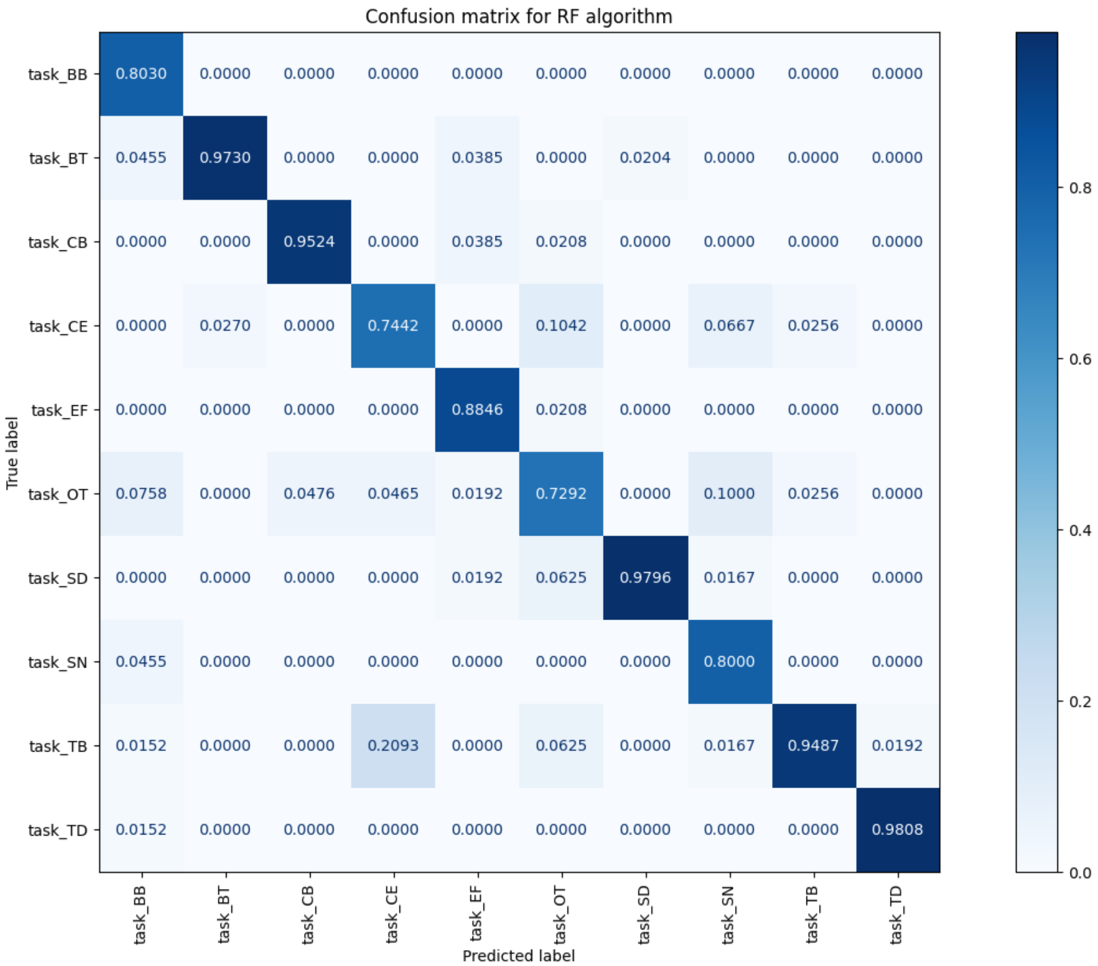


Figure 19: confusion matrix with precision values for RF algorithm.

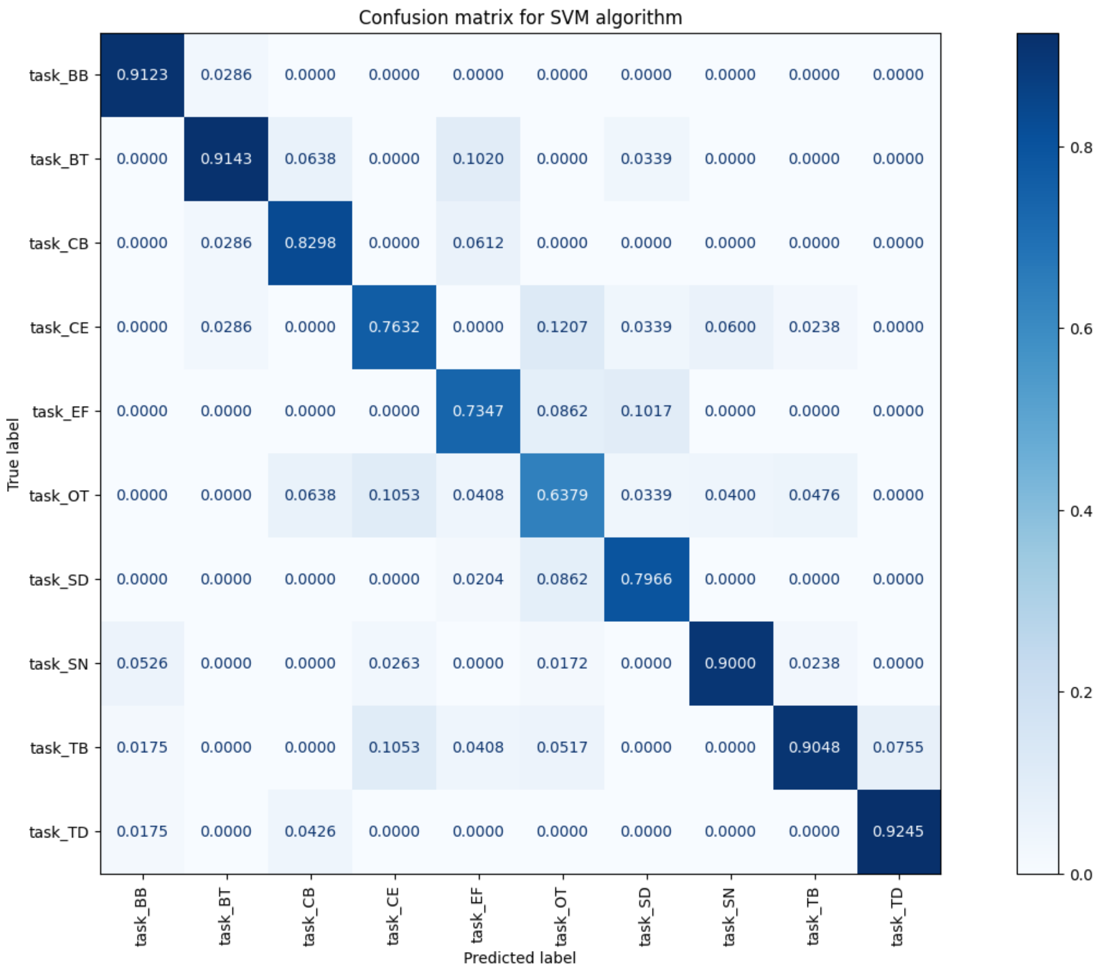


Figure 20: confusion matrix with precision values for SVM algorithm.

To get a more complete overview, in the following table (table 6), are reported the F1 scores of both models with the beginning data set (illustrated in table 3) and the F1 scores with the new dataset. Looking task by task:

- Task BB: the SVM has approximately the same score, but RF has worse value than the first classification but still more than the previous study.
- Task BT: is the worst predicted task in both algorithm comparing to the first dataset, between 9% and 11% less but with RF that is still more than 91%.
- Task CB: both SVM and RF give better results than their respective previous model, with the last reaching 94%
- Task CE: SVM algorithm gives greatly lower results than the first dataset instead of RF which is much higher F1 than its counterpart.
- Task EF: here the table shows the same behaviour of task CE.
- Task OT: both models enhance their performance with the complete dataset, especially RF that rises to 70 %.

- Task SD: like the last two activities, only RF model gives a better result, exceeding 94 %.
- Task SN: the SVM is lower but only 1% of difference, instead of RF that has a good enhancement of almost 4%.
- Task TB SVM undergoes a great improvement of more than 20%, also RF increased from 73.47% to 81.32%.
- Task TD: RF confirms a great result around 98% and SVM come close to its previous result.

Tasks	Previous results	SVM validation	RF validation	SVM model	RF model
BB	86,55%	94,55%	89,08%	94.74%	98,18%
BT	85,00%	83,12%	91,14%	94.12%	100%
CB	82,80%	86,67%	94,12%	84,00%	90,91%
CE	52,31%	71,60%	74,42%	86,79%	65,31%
EF	46,40%	75,00%	92,93%	90,48%	83,72%
OT	32,43%	67,27%	70,00%	56,52%	53,66%
RE	-	-	-	100%	100%
SD	83,34%	83,93%	94,12%	91,23%	85,71%
SN	71,43%	89,11%	86,49%	90,57%	82,76%
TB	33,33%	80,85%	81,32%	60,00%	73,47%
TD	92,21%	93,33%	98,08%	96,15%	98,11%

Table 6: F1 scores of the new models in comparison with the maximum classification value and the previous results.

Seen these results, has been demonstrated that the algorithm is valid, and has great performances also with different dataset and with a greater amount of data, so the good results obtained in chapter 3 and 4 are not due to a biased classification. In addition, with the complete dataset, SVM model responded positively with less variability and with an enhancement of the two worst classified task, but in general, RF gives a greater performance, and this can be sufficient to prefer it instead of the other.

Chapter 5
Conclusion and future development

The main objective of this thesis was to contribute to the improvements of activity detection for Parkinson and essential tremor patients. This is relevant because, in first instance there is a great amount of people suffering from these disorders and the impact that can have on their lives, secondly in the last years there is an increasing interest in human activity recognition and personalized medicine that can lead to a better quality of life for elders and people with neurodegenerative or motor pathologies.

To achieve this, an important milestone was firstly understanding which kind of approach, regarding artificial intelligence, was the best one. So, it was necessary to choose between machine learning and deep learning, and it has been chosen the first one, and then to stud which kind of window segmentation fitted better with this project. Seen that there was not a clear prevalence through all the results, it has been chosen to work with five different window time by time to keep the best one and this led to the development of the presented algorithm, with the objective of having a more solid model, which classifies activities, not only with the best precision possible, but also with the least variability.

This aspect was important to allow a good classification of a greater quantity of task, and above all, with a view to the future, to be able to add new activities to detect, and obviously reach the best performance possible, to provide medical personnel with a reliable instrument.

Another important aspect which was considered during the course of the project, was the patient comfort. It was important to provide the patient with a device that was not bulky, not heavy and that could be comfortable to wear in a future where a 24-hour real time monitoring system could be developed.

The results obtained in fact support the use of the smartwatch, as opposed to one or more IMU sensors, and in fact the results obtained show how good results can be achieved with this type of device, while also thinking about patient comfort.

As mentioned above, the aim in the future will be to develop a continuous monitoring system, in which the smartwatch detects the patient's activity throughout the day with a reasonable delay. Theoretically, this could allow an enhanced tremor monitoring that is more adaptable to the individual patient, in order to better assess which actions the subject suffers the most and to intervene for personalised pharmacological treatment. To do this, several challenges arise:

- adapt this algorithm to a continuous process that over a much longer window identifies the portion of the signal of interest and analyses it. Identifying the precise moment when a patient switches from rest mode to another activity.
- Connection issues, then understanding how to send all the data collected, both to medical staff and to an engineering staff working on the development of the whole system.
- Ethical and therefore privacy issues regarding data collection and monitoring.
- Hardware issues such as the battery life of the device, which will require cooperation from the patient.

Bibliography

[1] Expert Systems with Applications, "*Identification of activities of daily living in tremorous patients using inertial sensors*" - J. Ignacio Serrano et al. - 2017.

[2] Indawi journal, "*New Perspectives for Computer-Aided Discrimination of Parkinson's Disease and Essential Tremor*" - Povalej Bržan et al. – 2017.

[3] Parkinsonism & related disorders, *"The economic impact of Parkinson's disease"* - Leslie J. Findley - 2007

[4] Frontiers in Neuroscience, "*Measurement tools to assess activities of daily living in patients with Parkinson's disease: A systematic review"* https://doi.org/10.3389/fnins.2022.945398

[5] Nature review Neuroscience, "*Non-motor features of Parkinson disease"* – Schapira et al. – 2017.

[6] Royal college of physicians, "*Parkinson' disease, national clinical guideline for diagnosis and management in primary and secondary care"* ISBN 1 86016 283 5

[7] ACM Computing Surveys, Vol. 46, "*A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors"*- Andreas Bulling et al. – 2014.

[8] Sage journals, International Journal of Distributed Sensor Networks, *"A review on applications of activity recognition systems with regard to performance and evaluation"* – Suneth Ranasinghe et al. – 2016.

[9] IEEE Pervasive Computing, *"Human Activity Recognition and Pattern Discovery"* - Eunju Kim et al. – 2019.

[10] IEEE International Conference on Systems, Man and Cybernetics, *"Human Activity Recognition for Physical Rehabilitation"* – Daniel Leightley et al. – 2013

[11] Springer, Journal of Population Ageing, *"Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People"* – Lisa Schrader et al. -2020.

[12] Sensors journal, *"Sport-Related Human Activity Detection and Recognition Using a Smartwatch"*- Zhuang et al. – 2019.

[13] Chinese Control Conference (CCC), "*Recognizing Human Activity in Smart Home Using Deep Learning algorithm"* – Fang et al. - 2014.

[14] Internet of things vol. 12, *“Multi-label classification based ensemble learning for human activity recognition in smart home”* – Jethanandani et al. – 2020.

[15] Advanced Engineering Informatics vol 56, *“Determination of workers’ compliance to safety regulations using a spatio-temporal graph convolution network”* – Lee et al. 2023.

[16] AHA journals, “*Measurement of Instrumental Activities of Daily Living in Stroke”* – Khin et al. – 1995.

[17] Sensors journal, *“Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications”* -Muro-de-la-Harran et al. – 2014.

[18] IEEE Sensors Journal, *“A Smart Inertial System for 24h Monitoring and Classification of Tremor and Freezing of Gait in Parkinson’s Disease”* – Pierleoni et al. – 2019.

[19] Journal of the Neurological Sciences, *“Smartwatch for the analysis of rest tremor in patients with Parkinson's disease”* - Lopez-Blanco et al. – 2019.

[20] Journal of Neuroscience, *“Essential tremor quantification based on the combined use of smartphone and a smartwatch: The NetMD study”* -Lopez-Blanco et al. – 2018.

[21] Frontiers in Neurology, *“Monitoring Motor Symptoms During Activities of Daily Living in Individuals with Parkinson’s Disease”* – Thorp et al. – 2018.

[22] Information Fusion vol. 46, *“Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions”* – Nweke et al. 2019.

[23] Biomedical Signal Processing and Control vol 80, *“On the use of indexes derived from photoplethysmographic (PPG) signals for postoperative pain assessment: A narrative review”*- Argüello-Prada et al. – 2023.

[24] Sensors Journal, *“Electrodermal Activity Sensor for Classification of Calm/Distress Condition”*- Zangróniz et al. 2017.

[25] Sensors Journal, *“An Explainable EEG-Based Human Activity Recognition Model Using Machine-Learning Approach and LIME”* – Hussain et al. – 2023.

[26] Journal of Neuroscience, *“The Phase Difference Between Neural Drives to Antagonist Muscles in Essential Tremor Is Associated with the Relative Strength of Supraspinal and Afferent Input”* – Gallego et al. – 2015.

[27] Entropy Journal, *"Classification of Kinematic and Electromyographic Signals Associated with Pathological Tremor Using Machine and Deep Learning"* – Pascual-Valdunciel et al. – 2023.

[28] Annals of Clinical and Translational Neurology, *"Essential tremor severity and anatomical changes in brain areas controlling movement sequencing"* – Benito-León et al. – 2018.

[29] Sensors Journal, *"Inertial Data-Based AI Approaches for ADL and Fall Recognition"* – Martins et al. – 2022.

[30] Sensors Journal, *"Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition"* - Javier Ordóñez et al. - 2016.

[31] Sensors Journal, *"Window Size Impact in Human Activity Recognition"* – Banos et al. – 2014.

[32] Conference: Proceedings, The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, *"Activity Recognition from Accelerometer Data"* – Ravi et al. – 2005.

[33] Sensors Journal, *"Real -Time Estimation of Pathological Tremor Parameters from Gyroscope Data"* – Gallego et al. – 2010.

[34] 6th International Conference on Mobile Computing, Applications and Services (MobiCASE), *"Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors"* – Zeng et al. – 2015.

[35] International Conference on Information and Communication Technologies for Ageing Well and e-Health, *"The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones"* – Vavoulas et al. – 2016.

[36] 5th International Conference on Pervasive Technologies Related to Assistive Environments, *"Creating and Benchmarking a New Dataset for Physical Activity Monitoring"* - Reiss et al. – 2012.

[37] *"Deep Learning"* – Goodfellow et al. – 2016 - ISBN-13 978-0262035613.

[38] Current Oncology vol. 28, "*Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era"* – Mitsala et al. 2021. doi: 10.3390/curroncol28030149.

[39] The AAPS Journal, *"Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review"* – Kolluri et al. 2021.

[40] Prediction of Pathological Tremor Signals Using Long Short-Term Memory Neural Networks, *"Prediction of Pathological Tremor Signals Using Long Short-Term Memory Neural Networks"* – Pascual-Valdunciel et al. – 2022.

[41] IEEE International Conference on Systems, Man, and Cybernetics, *"A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer"* – Chen et al. – 2015.

[42] Journal of Healthcare Engineering, *"Employment of Ensemble Machine Learning Methods for Human Activity Recognition"* – Hasan et al. – 2022.

[43] International Conference on Intelligent Environments, *"Simple and Complex Activity Recognition Through Smart Phones"* – Dernbach et al. – 2012.

[44] IEEE Symposium on Computational Intelligence and Data Mining, *"Ready-to-Use Activity Recognition for Smartphones"* – Siirtola et al. – 2013.

[45] Journal of NeuroEngineering and Rehabilitation, *"Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy"* – Ahmadi et al. – 2018.

[46] Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2023), *"The importance of integral time length windows for the classification of activities of daily living based on Machine Learning techniques"* – A. Ruiz, E. Carbone, B. Larraga, E. Rocón, A. Gutiérrez – 2023.

Acknowledgements

I would like to express my profound gratitude to Professor Gastaldi, for her trust in the choice of this project, for always being available, for her insightful feedback and constructive criticism that have been instrumental in shaping this thesis.

I would like to sincerely thank Professor Gutiérrez for immediately placing great trust in me and giving me full freedom in the project. Thanks so much for his professionalism, for all his teachings, for the always constructive discussions and support throughout my time in Madrid and above all for giving me this great opportunity.

Thanks to Professor Larraga for her mentorship, her advice, and especially for her attitude, always ready to help me, and thanks to all the guys of Robolabo, for immediately making me feel part of the group, I wish you all the best.

Thanks to Eduardo Rocón for giving us the first dataset to begin this work, thanks for his guidelines and advice, and thanks to Fernando and Jorge for their great work and professionalism with the application and for putting up with all our requests.

A special thanks to Ainoha who has been the best partner I could have in this project. Working with you has been very easy, you have been a guide for me from the beginning and the understanding we managed to have, was crucial to the success of this work.