UNIVERSITÀ
POLITECNICA
DELLE MARCHE

FACULTY OF ENGINEERING

Master's Degree in Biomedical Engineering

DEVELOPMENT OF A NON-INVASIVE ULTRASONIC SENSOR NETWORK FOR THE MEASUREMENT OF ACTIVITIES OF PEOPLE IN INDOOR ENVIRONMENT TOWARDS OPTIMIZED PERSONALIZED COMFORT

Supervisor:

Prof. Lorenzo Scalise

Candidate:

Francesco Carbonari

Co-supervisors:

Prof. Sara Casaccia

Doc. Ilaria Ciuffreda

Academic Year 2023/2024

INDEX OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Identifying the level of human activities is fundamental to take metabolic rate into account when assessing comfort in indoor environments. This master thesis studies the feasibility of employing multi-ultrasonic sensors integrated with environmental sensors on a multidomain monitoring platform to investigate model for personalized comfort assessment.

For the experiment was considered a living environment for office use and 10 healthy volunteer subjects were monitored. The test includes 5 typical office activities (writing on paper, typing on PC, talking in the phone, standing and walking within the room) alternated by resting phases and surveys that deals with perceived comfort in different domains.

After data acquisition and storage, a data calibration and filtering processes were performed to extract meaningful temporal features, to detect office activities using temporal markers, such as the Root Mean Square (RMS), the Crest Factor (CF) and the Shape Factor (FF). Subsequently the markers were processed into time series, leading to the construction of a dataset used to train different Machine Learning (ML) algorithms selected according to the literature. These algorithms include traditional techniques like Random Forest (RF), Support Vector Machine (SVM) and k-Nearest Neighbours (k-NN), used to distinguish between activity and non-activity phases. Then were applied more advanced techniques such as Convolutional Neural Network (CNN) and Long Short-Term Memory Network (LSTM) to differentiate the different activities. Results confirm the feasibility of integrating ultrasonic sensors in a monitoring platform to capture meaningful movement patterns to discern various office activities. The different ML methods, also used in combination, have reached discrete accuracies in predicting the activities that go from 64 to 87%. In addition, results show that from each activity it's possible to calculate different Predictive Mean Vote (PMV) values. Findings also reveals that activity discrimination has an impact of 213% on the estimated PMV values. This information can be integrated in personal comfort models (PCMs) to optimize the occupants' well-being as well as thermoregulation of the built environment and, hence, the building energy consumption.

1. INTRODUCTION

1.1 Indoor Comfort Environment

In developed countries majority of people spend more than 90% of their time indoors. Indoor conditions have therefore far-reaching implications for their health, general well-being and performance.

These conditions are profoundly affected by their geographical location, season and age, and especially by job requirement, which results to be the primary parameters since the office workers spend at least 40 h a week in their workplace. Thus, an efficient and conducive working environment include properties of office spaces layout such as interior design, size of personal workspace, workstation equipment and furniture ergonomics: all these factors determine how comfortable employees feel in the office because it's a vital and fundamental requirement for occupants to work efficiently and productively [1], [2].

However, over the last decades, research has shown that, in addition to the physical environment of office spaces, indoor environmental quality (IEQ) has a significant effect on the comfort, health and productivity of occupants. IEQ has become a critical consideration as it significantly impacts occupant the well-being, health, comfort and productivity of workers in office buildings. IEQ includes a combination of elements such as thermal, acoustic and visual conditions and indoor air quality.

These four domains are interconnected, and their combined effect contribute to the overall indoor environment. In order to satisfy acceptable ranges of these parameters, codes and standards for each of these factors have been established and technologies and systems are being engineered in an efficient energy manner. Even though the requirements of these standards are met, not all building occupants are satisfied with the indoor environment also depending more specifically on their location and routines. In addition, the same indoor conditions may lead to contrasting subjective responses. One obvious reason is that human requirements are different and therefore not all are satisfied by the same conditions [3], [4].

1.2 Thermal Comfort

Thermal comfort can be defined as “that condition of mind that expresses satisfaction with the thermal environment” and so describe the occupant's subjective perception of the thermal environment. The definition applies to the thermal comfort of an individual. When thermally comfortable, a building user will wish to feel neither warmer nor cooler, if asked about thermal state and preference. In buildings, however, a person usually shares the built environment with other occupants. Thermal equilibrium is obtained when an occupant’s internal heat production is the same as its heat loss. The heat balance of an individual can be influenced by levels of physical activity, clothing insulation, as well as the parameters of the thermal environment. For example, thermal sensation is generally perceived as better when occupants of a space have control over indoor temperature, as it helps to alleviate high occupant thermal expectations on a mechanical ventilation system [4], [5], [6].

1.2.1 Thermal Comfort Models

The evaluation of thermal comfort can be carried out using Fanger’s indices: predicted mean vote (PMV) and predicted percentage dissatisfied (PPD) which make it possible to predict the mean thermal sensation and mean satisfaction with thermal conditions of a group of people. These approaches are recognized according ASHRAE 55 and ISO 7730 standards evaluating indoor environments [7].

PMV model was firstly introduced by Ole Fanger in the 1960s and is based on heat-balance equations and empirical studies related to skin temperature.
PMV is an index that aims to predict the mean value of votes of a group of occupants on a seven-point thermal sensation scale. In fact the thermal comfort can be influenced by a combination of factors that can be grouped in three categories:

- personal factors, that are individual-specific characteristics of the occupant, including the metabolic rate and the clothing insulation;

- environmental factors, that encompass conditions of the thermal environment which include air temperature, mean radiant temperature, air velocity and relative humidity;
- physiological factors, parameters that reflects the human body response, the sweat rate and the heart rate [8], [9].

PMV calculation is based on a combination of physical and personal parameters. Within the PMV index, +3 translates as too hot, while -3 translates as too cold, as depicted below in Figure 1.

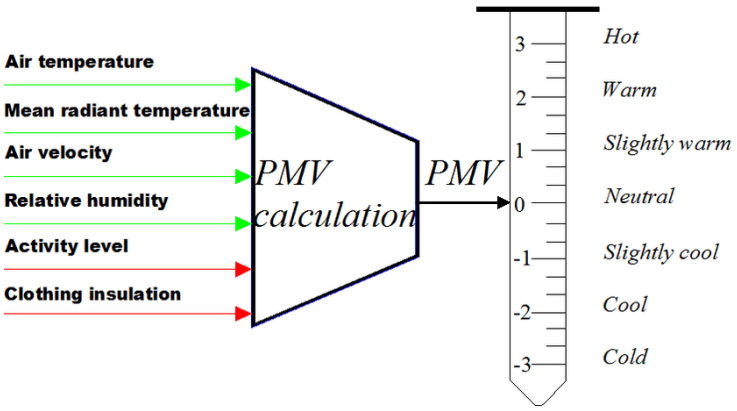

Figure 1. Intuitive representation of how PMV model works. Given the parameters shown on the left, the PMV calculation can compute a number which represent the occupant's thermal sensation.

The adaptive comfort model was developed by De Dear & Brager through extensive empirical and experimental investigations. This model is based on hypothesis predicting that contextual factors and past thermal history modify building occupants' thermal expectations and preferences. The model is intended to compute the comfortable air temperature in primarily naturally ventilated buildings by considering thermal adaptations, including behavioural and psychological adaptations [8], [9].

1.3 Visual Comfort

Visual comfort is a subjective response to the quantity and quality of light within a space. It characterizes the connection between the human needs and the light environment. A number of physical properties of the visual environment are defined and used to evaluate its quality in an objective way. Visual conditions are characterized by such parameters as luminance distribution, illuminance and its uniformity, glare, colour of light, colour rendering, flicker rate and amount of daylight. Visual comfort is considered an essential factor for the safety of older people in living environments [1], [4].

1.4 Acoustic Comfort

Acoustic comfort in the living environment is an increasingly important topic for improving productivity and reducing anxiety for regular users in indoor environments. Providing a good acoustic environment is necessary to minimize the noise and support the satisfaction of the inhabitants, preventing the occurrence of annoyance, sleep disturbance, irritability and also long-term health effects. Older people are more vulnerable to noise because they are more sensitive to disturbance. The quality of the sound environment is linked to numerous physical parameters, which include both the physical properties of sound itself and of a room. Sound is characterized by the sound pressure level in a short-term and long-term period and by sound frequency. The acoustic environment is influenced by such physical room properties as sound insulation, absorption and reverberation time [4], [10], [11].

1.5 Indoor Air Quality Comfort

Introduced more recently the term comfort is not commonly used in relation to indoor air quality (IAQ), and it is mainly linked with the lack of discomfort due to odour and sensory irritation. Acceptable air quality is when there are no specific pollutants in harmful

concentrations, according to the criteria established by the competent authorities and with at least 80% of the occupants exposed do not express dissatisfaction. Consequently, most of the standards providing the requirements for IAQ define the conditions by providing the minimum percentage of persons dissatisfied with air quality. They are mainly based on the discomfort and annoyance caused for visitors to indoor spaces. Recently, some standards also deal with the requirements for occupants [12], [13].

1.6 Personal Comfort Model (PCM)

Both PMV and adaptive comfort models are designed to compute the average thermal comfort for a group of occupants rather than considering individual differences and dynamics in environmental perceptions. In fact, many studies have demonstrated the limitations of these traditional comfort models that discriminate personal differences, such as personal preferences, gender, age and thermal history. Furthermore, thermal comfort in transient scenarios, like coming indoors from a hot outdoor environment, is not accounted in the PMV and adaptive comfort models, which are only observable in the measurement of occupant's physiological signals, such as skin temperature, metabolic rate and body postures.

To face these points, researchers aim to capture occupant comfort via personal comfort models (PCM) addressing the diversity of occupant environmental preferences. PCM employs a data-driven approach using various modelling algorithms and it focuses on capturing the unique comfort characteristics of each occupant rather than predicting the average comfort for a large population.

Indeed, even under the same thermal conditions, individuals may have a different perception of thermal comfort due to physiological factors, which are influenced by human activities, as well as by the personal thermal history. Therefore, the integration of systems for human activities recognition (HAR) becomes critical in the development of PCMs, since HAR contributes to define the metabolic rate. In this context both sensing and digital technologies

play relevant roles since the non-intrusiveness of the adopted sensors can enhance their acceptability by diverse types of users and their accuracy clearly impacts on the result. This is important to considered even when employing Artificial Intelligence (AI) and Machine Learning (ML) algorithms for classification purposes, since the input uncertainty contributions propagate through the entire measurement process, determining a certain confidence interval of the output [1], [9], [14].

1.7 Ultrasound for Activity Recognition

In the last decades, significant work has been done in this area to identify different set of activities in HAR and a wide range of sensors have been used. Mainly there exist two different sensing approaches.

The first approach used is based on wearable sensors. Wearable sensors refer to sensors that are positioned or directly attached to the body resulting at a certain extent intrusive. The sensors used in this case include mobile sensors, accelerometer, gyroscope, proximity sensors etc. Though wearable sensors have a certain ease of use, given their easy deployment and data acquisition method, they still have some limitations. In fact, these solutions reduce the applicability of the smart space to only those that own wearable devices, and to times when the users wear those devices. Moreover, in residential smart spaces, wearable devices can cause discomfort, also because they are usually not seen as a part of the regular residential life. Wearables are also limited in terms of measurement accuracy and susceptibility to motion artifacts.

The other kind of approach considers using environment placed sensors, such as cameras, pyro-electric (PIR), acoustic sensor, ultrasound sensors and thermal sensors, etc. In this approach, the sensors are embedded in the environment, which makes it more suitable to create intelligent applications such as a smart environment. An advantage of using environment over wearable sensors is that they are not directly applied to the individuals.

Anyway, the combination of both the types of sensors offers significant advantages, being able to provide a comprehensive view of the entire ecosystem, including subjects and environment and, hence, considering both environmental and physiological domains [15], [16], [17].

Considering sensors integrated into the environment, it can be noticed that several studies perform HAR mainly based on image features. However, video processing is highly complex o as well as computational expensive. Moreover, there is a concern that individuals may be uncomfortable with being constantly followed and monitored by an intrusive system, fearing the disclosure of sensitive information and potential privacy loss.

There are few studies that utilize ultrasonic (US) sensors to detect and recognize human activities.

In this study, the use of ultrasound technology is employed to detect the occupant activities within an indoor living environment with the final aim of evaluating their impact on personalized comfort through a multidomain monitoring platform. Activity levels can be assigned to different office activities enabling an accurate estimation of metabolic rate, that contributes to PMV model.

By discriminating among different tasks, it's possible to better assess personal comfort through these estimations.

This data can be combined with data coming from physiological and environmental sensors, to widely depict the well-being state of an individual in the living environment; consequently, this information can be exploited for optimizing thermoregulation of the environment and, hence, the related energy consumption. The innovational aspect of this research lies in the application of US sensors for HAR in a sensors network.

To the best of the authors' knowledge, there is no existing literature studies that employ US for HAR integrating the system in a multidomain monitoring platform focused on the personalized comfort assessment [15], [18], [19].

1.8 Aim Of The Study

Therefore, this study main contributions is the innovative use of US sensors for HAR in the broader context of personalized comfort assessment, with the goal to develop a comprehensive monitoring platform, that integrates environmental monitoring stations and a network of US sensors strategically positioned to detect activities and perform HAR.
Moreover, a data calibration and filtering methodology aimed at extracting features (temporal markers) to detect office activities is proposed in order to consider physiological variability and specific patterns.
Then is proposed a methodology that recognize the activity through the implementation of ML algorithms. The performance metrics of the employed ML algorithms are computed to verify which is the most reliable for activities recognition.
Finally, once the activities have been classified, it's assigned a metabolic rate to each one and it's calculated the PMV to obtain a personalized model.

This study is within the framework of the WEPOP project. The WEPOP project (Prot. 2022RKLB3J) "WEarable Platform for OptImised Personal comfort" is co-funded by the Italian Ministry of Research within the PRIN 2022 program.
WEPOP adopts an innovative approach that contributes to enhance the personalized well-being measurement considering the human activity influences on comfort. This approach provides useful information also for the thermo-hygrometric control with a view of energy efficiency in living environments.
The main objective of WEPOP is to develop a fully integrated multi sensing platform which includes wearable sensors, environmental sensors and AI algorithms for real-time personal comfort measurement and control. The project exploits the experience created in the NEXT.COM project [20] in the field of multidomain comfort measurement and modelling. The focus is on designing and optimizing the environmental sensor network to acquire parameters related to thermal, acoustic, visual and IAQ domains.

2. STATE OF ART

2.1 Human Activity Recognition (HAR)

Human activity recognition (HAR) is the process of automated recognition and understanding human actions, that can be individual or group activities. It has become a popular topic in the last decade due to its importance in many real-world applications areas, including health care, assisting living technologies and safety applications. HAR is also emerging as a powerful tool in monitoring systems for general purposes such as monitor physical, functional, and cognitive health of older adults at home. Indeed, essentially, HAR involves identifying and interpreting human actions and interactions with the environment, especially movements of the whole body and limbs. Understanding these actions is crucial for predicting their effects or outcomes, as well as inferring the performer's intention, goal, and mental status.

HAR is the key to build human-centred applications and enables natural interaction between users and smart environments. These environments are instrumented room or space equipped with appropriate sensors and actuators, available in different types, to perceive the physical state or human activities within this space.

HAR aims to recognize human activities in both controlled and uncontrolled settings. HAR algorithms face many challenges such as complexity and variety of daily activities, intra-subject and inter-subject variability, the trade-off between performance and privacy, computational efficiency in portable devices and difficulty of data annotation [18], [21].

There are various sensing approaches able to provide different and complementary information about human actions. Visual modalities, such as RGB videos, skeleton data, depth data, infrared sequences and point clouds, are more intuitive for representing human actions because they are more similar to the human visual system's functioning. In contrast, non-visual modalities, such as accelerometer, gyroscope, magnetometer, radar, and Wi-Fi can be used for privacy-sensitive scenarios or when visual data are insufficient or unavailable. Data fusion techniques improve recognition models by integrating information from multiple modalities, enhancing understanding of human actions, and addressing individual modalities' limitations.

Data fusion can be implemented at the feature, decision, or model levels, increasing the accuracy of HAR models. Further aspects relevant to HAR include the action prediction, which aims to anticipate future actions based on observed actions, and narrative understanding, which focuses on identifying the agent's identity or social role in the context of an action. Additionally, transfer learning and co-learning across different modalities can improve the robustness and generalizability of HAR models, allowing them to adapt to new scenarios and handle diverse data sources effectively [18].

Data for training and testing HAR algorithms is typically obtained from two main sources: ambient sensors and embedded sensors. Ambient sensors include environmental devices such as temperature, humidity sensors or video cameras placed at specific locations in the environment. Embedded sensors, on the other hand are integrated into personal devices such as smartphones and smartwatches or are integrated into clothes or other specialized medical equipment. Cameras have been widely used in the HAR applications, however collecting video data presents many issues regarding privacy limitation and computational requirements. While video cameras produce rich contextual information, these privacy issues limitations have led many researchers to work with other ambient and embedded sensors, including depth images as a privacy-preserving alternative. Many approaches to HAR mostly consider recognizing the activities of individuals. There are a limited number of studies that have intended to recognize group activity primarily rely on image-based features.

HAR systems operate by taking large dataset of the measured features from the acquisition devices. The first step of any HAR system is the selection of appropriate sensors and the attributes to be measured. Then, the data coming from the sensors are manipulated and processed in different stages, that generally include the pre-processing, the segmentation, the feature extraction, the dimensionality reduction and, by the end, the classification.

By applying computational algorithms, HAR are able to predict the activities performed by the users. This involves the creation of a feature vector from the raw sensor data, which is then used to train machine learning algorithms to recognize and classify the activities [19], [22].

2.2 Machine Learning (ML) for HAR

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that aims to develop algorithms able to identify and infer patterns from a dataset. These algorithms are primarily categorized in two major classes: supervised and unsupervised learning. Supervised learning has the aim to create a mathematical model based on the relationship between input and output data enabling the prediction of future unseen data points. On the other hand, unsupervised learning focuses on identify patterns in input data without any knowledge of the output. Recently, sensor-based HAR gained popularity as a research topic, with lots of studies exploring different sensing technologies and some methods have been proposed for modelling and recognizing human activities. Research mainly implemented traditional ML processes for HAR tasks, including sensor-based movement data collection, preprocessing, action segmentation, extraction of features, and activity classification. Classical supervised ML techniques such as Naïve Bayes (NB), Decision Tree (DT), Random Forests (RF), Support Vector Machine (SVM) and k-Nearest Neighbours (k-NN), have showed excellent effectiveness in the classification of activities based on data collected from sensors [23], [24].

NB is an extension of the traditional algorithm, maximum likelihood estimation theory (MLE) that allows the classification of testing data based on the previously labelled data. NB is a probabilistic classifiers based on applying Bayes' theorem with the assumption of strong independence between the features. NB classifier is able to develop a probability-based mechanism on input and previous data for classification and prediction [25].

DT are hierarchical data structures that classify data by sorting based on the features values. DT is characterized of nodes and branches: a node represents a feature to be classified, while a branch represents a possible value that the node can assume [26].

RF consists of a combination of multiple decision trees that improves the classification performance of a single tree classifier. It achieves this by combining the bootstrap aggregating (bagging) method and randomization in selecting data partitions at each node during the construction of decision tree. A RF classifier integrates a set of independent DT classifiers each trained on a different subset of data. The final classification decision is made by aggregating

the predictions from all individual trees, using a majority voting mechanism, which helps improve accuracy and robustness and reduce the risk of overfitting [27].

K-NN is a ML algorithm that shows the characteristics of instance-based learning in which the function is just calculated locally until grouping. K-NN is mostly used as a method of classification, being one of the easiest, in which grouping of examples is depending on their coordinates and distance from others on the feature space. In the k-NN classification, the output for an item is a class from which this belongs. K-NN performs a mathematical calculation based on similarity measures to find the distance between the data to makes its classification. The steps of k-NN can be described as follows. After obtaining unclassified data is calculated the distance from the new data to all classified data. Then is initialized the value of k and the distances are sorted. After taking k nearest neighbors is assigned the class to new data by the majority vote of its k nearest neighbors. To measure the distance between two points there are many different ways like Euclidian, Manhattan, Minkowski or Weighted [28].

SVM is a supervised ML method that given labeled training data, outputs a hyperplane which separates and categorizes the new data. SVM is able to perform an analysis of data for classification as well as regression, but it's mostly used for classification. SVM can execute linear classification, but can also do a non-linear classification utilizing the different kernels, by mapping their contributions to high-dimensional element spaces. Kernels characterize a set of mathematical functions. The work of kernel is to require data as input and transform it into the specified frame. There are different types of kernel functions like linear, nonlinear, sigmoid, radial basis function (RBF) and polynomial. The kernel functions return the inner product between two points in an appropriate feature space. Hence by characterizing an idea of similarity, with small computational cost. Even though SVM is used for both supervising and unsupervised, utilizing supervised SVM is normally quicker and has more accuracy [28].

However, all these classical ML techniques face limitations mainly due to dependence on manual feature extraction. In fact, manually extracting features is a laborious and time-consuming process that relies on user experience and expertise to identify relevant statistical and frequency domain features. Furthermore, the characteristics are able to characterize certain basic human actions rather than complex ones, making it difficult for simple ML methods to adapt to novel, complex HAR scenarios [24], [28].

2.3 Deep Learning (DL) for HAR

In recent years, the advent of deep learning (DL) for HAR has led to significant enhancements in recognition accuracy by overcoming many limitations encountered by traditional ML methods and due to their superior performance. The strength of DL is based on the idea of the data representation. DL relies in its ability to automatically extract optimal features from raw input sensor data in a task dependent manner without any human intervention and present the low-level original temporal features with high-level abstract sequences. DL approach to HAR in a way that seeks to reduce on the efforts needed for feature engineering which requires domain knowledge [24]. This ability enables the identification of unknown patterns that otherwise would remain hidden or unknown. However, DL models also present some limitation: they are black-box models, meaning that interpretation is difficult and inherent and they require large datasets for training and have a high computational cost. Due to these limitations, in some areas classical machine learning (CML) methods are preferred, especially when the training dataset is small, or when is necessary a fast training. Some of the most common DL algorithms are: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory Network (LSTM) [24], [29], [30].

CNN is DL technique that has the ability to consider the spatial structure of the input information. Nowadays, the use of CNN is an inevitable tool of the process of recognition, especially in the image processing research community and it's frequently applied to solve the problems of DL and pattern recognition. These systems are prevalent in detecting human activity and in recognizing different classes of body movement and actions. CNNs are inspired by biological processes and their layers have neurons that are grouped by dimension. Each neuron in a layer connects to a small region of the previous layer. CNN is a regularized version of the artificial neural network (ANN). Generally, ANN are completely connected network, where every neuron in a layer is connected to all the neurons in the next layer. Due to the large number of parameters these types of networks tend to overfit the data. To overcome this problem, CNNs employ alternative regularization techniques to include a magnitude estimation of weights to the loss function or to gather progressively complex patterns from smaller and easier patterns. A CNN consists of an input layer, followed by hidden layers and

an output layer at the end. The hidden layers are composed of sequences of convolutional layers that perform convolutions with filters by means of multiplication or other dot product. Convolutional operations reduce the number of training parameters with respect to standard neural networks and this allows an efficient training of deep architectures. The activation function is generally a Rectified Linear Unit (Relu) layer, used to introduce non-linearity, which outputs the input directly if it is positive, otherwise, it outputs zero. Relu is generally trailed by extra convolutions, for example, normalization layers, aimed to normalize the data within the network to improve convergence during training. There are pooling layers that reduce the dimensionality of the data by aggregating the outputs of neuron clusters in one layer (generally max pooling operations). At the end of the network there are fully connected layers, that associate every neuron in one layer to every neuron in another layer fully connected layers are used to determine the class score [28], [30].

Differently from CNN, RNN is specifically known for its ability to label sequences or time series data. RNN has the distinguishing feature of keeping the memory from previous input sequence, which is then used to influence the current output sequence. By using RNNs the network is trained via backpropagation through time, taking advantage of the temporal relationship between sensor readings. Although RNN can capture chronological information from sequential data, it has an issue known as the problem of gradient vanishing. This problem hinders the ability of the network to model between raw sensor data and human activities in a long context window, causing an exploding gradient problem. This happens when large error gradients start accumulating, resulting in significant changes to the neural network model during training, which in effect prevents a model from training with the available data, and causes the trained model to be unstable [30], [31].

LSTM, which is a variety of RNN, has been designed to overcome these issues that arise in RNN architectures. LSTM and RNNs are very similar, with the difference being that hidden layers in LSTMs contain memory blocks with cells instead of non-linear units, which can store information over long timespans. Each cell remembers values of arbitrary time intervals, while the gates regulate the information flow both from and to the cells. These networks are ideal for classification, processing and predicting time-series based data. LSTM has great advantages

in feature extraction of sequence data and is proven to excel in learning, processing and classifying such types of data [29], [30], [31].

2.4 Ultrasonic Sensors in HAR

Indoor activities, especially activities of daily living, such as standing, sitting and falling are the most popular use-cases for using US sensors. Specifically, for recognizing simple indoor activities, pulsed ultrasonic sensors are frequently used to measure distance to the objects involved in the activity, aiding in the accurate detection and recognition of these movements [29].

Ghosh et al. [32] used an heterogeneous ultrasonic sensor grid with two kinds of ultrasonic sensors: 4 HC-SR04 ultrasonic sensors mounted on a grid panel of 70 x 70 cm on the ceiling and a LV-MaxSonar-EZO range finder in the middle to reduce the dead zone. Activities to be predicted include standing, sitting and falling as well as the direction of movement of a single person under the panel. They collected 100 samples from each of the five participants that were asked to perform all those activities in a controlled laboratory environment with uniform spaces. This system has extensive application possibilities beyond the laboratory for monitoring of human activity in hospital, workplace, and home environments. Relying on the distance profile, they performed a classification analysis using multiple machine learning algorithms: SVM, K-NN and DT techniques to classify the targeted activities. Their experimental results show 81% to 90% correct detection of different activities with DT that gave the best results.

They later extended their previous work to recognize these events for a group of multiple people [16], [33]. These experiments were conducted in a supervised laboratory setup collecting test data from multiple occupants. The algorithm employed was Hidden Markov Model (HMM). HMM is a multi-layer probabilistic model that consists of hidden and observable states. It considers a sequence of inputs to predict final observations by constituting hidden states. To identify the final activities the transition probabilities between

the states are calculated, combining spatial and temporal aspects of the activities and integrating them for obtain an overall activity profiling. The experimental results have shown that HMM can detect different activities with accuracy more than 90% in laboratory setting and also improve overall identification accuracy compared to existing works. This developed system can be further evolved into ready-to-deploy smart sensing panels which can be effective for human activity monitoring in an indoor environment.

Patel [17] targeted at a complete new set of activities of daily living including Using Refrigerator, Used Refrigerator, Appeared near burner, and Using burner. It's applied a fusion of sensor networks consisting of Infrared (IR) Breakbeam Sensor, Ultrasonic sensor (HC-SR04) and Passive Infrared sensor (HC-SR501) interfaced with Arduino MKR1000 board. Then Raspberry acts as the bridge between the software components and the lower-level hardware components. The sensors are placed in activity-dense areas, that their location is based on high probabilities of observing activities. Activities and activity transitions are modeled using a Finite State Machine (FSM) representation in which there are states that represent distinct activities and transitions that show the changes between activities. This system deploys a FSM-based activity detection, achieving 96% accuracy in real-time. Afterwards, several ML methods are used to create an effective activity prediction framework. By implementing multiple ML algorithms, including logistic regression, SVM and Neural Network, and apply these methods on a dataset with around 20000 data points collected over three months, it can be achieve up to 98.5% activity prediction accuracy with 4 ms delay. This makes it a perfect real-time system example. This smart and pervasive space implementation doesn't use any intrusive sensor or data acquisition unit such as wearables or camera, so the user is not dependent to it and has no security and privacy issues. Consequently, this non-intrusive smart space implementation can provide highly accurate activity detection and prediction with low computation requirements.

In [34] it's proposed the Echolocation based Activity Detector, that involves a contactless sensor array of four HRLV-Max Sonar 42 kHz pulsed ultrasound sensors, which compute the separation distance to the most proximal reflector with mm resolution. The research demonstrates the capacity of the sensor to distinguish between four common activities like sedentary sitting, typing, writing, and standing performed at a workstation within an office

environment with various classical items to serve as representative background reflectors and facilitate activity performance. To simulate the natural variability with which an individual may perform such activities over the course of a workday, various activity subclasses were performed. Cubic support vector machine (CSVM) classifiers are developed using dispersion-related features computed from the time-series array outputs. Sedentary activities typically performed in an office environment (sitting, typing, writing) were distinguished from standing with 85.7% accuracy. Moreover, an average classification accuracy of 80.2% was observed for distinguishing between the entire set of activities.

Ali et al. [35] suggested an approach for classification detecting motion patterns with NB and CNN algorithm using an array of five US sensors with PIR sensor. The features are extracted from detected motion according to spatio-temporal parameters that are reflected over the blueprint design of the building. The NB algorithm is used to classify the detected motion pattern to be normal or abnormal motion depending on the collected training data. In the final stage, the CNN approach is also used for classification purposes combined with NB for a better accuracy and detection rate, CNN is applied over walking path patters inside the living room. Combining ML algorithms such as NB with CNN is very effective. It allows to have a hybrid algorithm that enhances the accuracy of the detection rate by having another check of the motion subset, which also increases the number of features to better describe the complete or partial motion patterns.

In [36] it's proposed a framework to recognize human actions by ultrasound active sensing. The system employs a tweeter to emit ultrasound and a micro electro-mechanical systems (MEMS) microphone as the receiver. Data collection took place in three different spaces: anechoic chamber, room without furniture and room with furniture. The study included eight fundamental action classes: hand-waving, throwing, kicking, picking-up, walking, lying-down, sitting, and standing. The work focused on validating the potential of two types of features: time-series reflected waves and time-series envelopes of reflected waves. By analyzing the temporal variation of the amplitude of ultrasound reflected waves, feature values were calculated. Classification was then performed using a SVM and a typical deep CNN design with numerous layers called Visual Geometry Group (VGG). The method achieved an accuracy of

97.9% when trained and evaluated on the same person in the same environment and 89.5% accuracy when trained and evaluated on different people.

Venkatesh et al. [37] tested a system composed of 12 HRLV-MaxSonar EZO ultrasonic sensors on a 4 x 4 m room furnished with a kitchen bench. The sensors are placed equally distant on the ceiling in an array of 4 x 3 x 2,5 m to cover the regions under investigation. This type of sensors are considered to be high resolution, high precision and low voltage, with a range that can reach 4,5 m without deterioration in the quality of the signal. The covering area can reach 60 cm in radius for a distance of 3 m. Each sensor outputs its distance from the top of an object or person obtaining the room height when there is nothing between the sensor and the floor. The activities to be predicted are: eating, laying down on the sofa, cooking, sitting on the sofa and walking. An activity is predicted based on the feature's values at that timestamp, allowing for robust real-time activity recognition. They asked three participants to perform their usual daily activities in the room for 6 hours. To investigate the effectiveness of the US sensors' distance values, duration and on/off state in recognizing a person's various activities, they trained a ML classifier to predict activities based on each of those features. In this experiment, was used NB as the classifier and compared accuracy results of the model trained with only the sensor's distance readings (R), the duration of being on at any given time (D), the on/off state (O), or any possible combinations of those three features (DO, RD, RO and ROD) as input attributes. US sensors' readings, duration and on/off states can accurately predict (86%) of the other resident's activities, even when the activities are interleaved with each other.

In [38] it's developed a DL based device free activity recognition framework, named EI, able to remove the environment and subject specific information from the activity data and extract environment subject-independent features shared by the data collected on different subjects under different environments on US signal. The transmitter is an iPad on which an ultrasound generator app is installed, and it can emit an ultrasound signal of nearly 19 KHz, while receiver is a smartphone and with a recorder app to collect the sound waves. 12 subjects were employed to perform 6 different activities (wiping the whiteboard, walking, moving a suitcase, rotating the chair, sitting, as well as standing up and sitting down) in six different rooms. Subjects repeat these six activities in each room for 5 rounds. The proposed EI framework can achieve better performance compared to more classical algorithm like RF. For RF, though was

used Mel-frequency Cepstral Coefficients (MFCCs), a feature commonly used for audible sound-based recognition tasks, as its input data, its accuracy is still lower than that of the deep learning models.

3. MATERIALS AND METHODS

The aim of this work is to develop a non-invasive acquisition system through a specific measurement sensors network, in the perspective of the future integration in a unique multidomain monitoring platform, with focusing on ultrasonic sensors to detect activities and perform HAR in office indoor environment. Machine learning (ML) algorithms has been chosen, according to literature, as the most suitable approach for the prediction of occupants' activity. Hence, in this section, it is possible to find the explanation of the phases that have been implemented during the practical part of this thesis with the aim of providing a proper amount of information to improve the proposed work.

The sensors that made up the system capable of acquiring the parameters necessary for this purpose are listed in Table 1.

The whole system operates through a code developed in Python language that controls every procedure, from signals acquisition to their processing.

The subject's comfort perception was registered through the use of a questionnaire, administered periodically by using a Google Forms web page.

The acquisition of the environmental parameters was instead carried out using three different sensors capable of measuring multiple parameters inside the room which include temperature, relative humidity, noise level, CO_2 concentration, volatile organic compound, particulate matter with different diameters and ambient light. The signals acquired from US sensors were transmitted to a Raspberry Pi 4 which was used as the data processing unit. Raspberry is a flexible device, chosen due to the possibility in supporting different programming languages, including Python, which was used in the development of the code, being able to interface with our ultrasound's platform.

An office experimental setup with a specific test protocol was developed to conduct tests, involving human participants.

Once all signals have been recorded during the tests, they were processed and significant features were extracted to obtain consistent results. This is done to verify whether the collected parameters and the applied algorithm can accurately predict the office activity. The data analysis step is divided into three phases:

Phase 1: Quantitative analysis. In this phase the mean and the standard deviation of the mean of the optimal temporal features were computed, in order to evaluate the differences between the diverse office activities.

Phase 2: ML/AI analysis. In this phase the collected data was processed into time series, leading to the construction of a dataset used to train different ML algorithms selected according to the literature. The aim of these ML techniques was to predict correctly the activity; the ability of classifiers was evaluated with confusion matrix and various performance metrics.

Phase 3: PMV calculation. To observe how the office activities impact on the subjects' comfort perception, the PMV model was computed, by using all necessary factors that include environmental parameters and different metabolic rates, associated to each activity.

Table 1. List of sensors and technical specifications

Sensor	**Technical specifications**		
	Measured quantity	***Measurement range***	***Precision***
Netatmo (sampling interval: 5 min)	T	0-50 °C	±0.3 °C
	RH	0-100 %	±3 %
	CO_2	0-5000 ppm	
	Noise level	35-120 dB	
Sensibo Element	T	0-55 °C	±0.5 °C
	RH	5-95 %	±5 %
	PM	-	-
	VOC	-	-
	EtOH	-	-
HibouAir (sampling interval: 15 min)	T	0-50 °C	±1 °C
	RH	0-100 %	±3 %
	VOC	400-10000 ppm	±30 ppm
	P	300-110 hPa	±0.6 hPa
	PM	0.3 µg/m^3	±15%
	Ambient light	110 mLux	
EV_MOD_ICU-30201-01	Distance	5-9 m	±0.2 mm

3.1 Ultrasonic Sensors

The ultrasonic sensor 30201-01 (TDK, Tokyo, Japan), was the main component employed in this work. ICU-30201 is a miniature, ultra-low power, ultra-long-range ultrasonic Time-of-Flight (ToF) transceiver. Based on TDK's patented MEMS technology, the ICU-30201 is a system-in-package that integrates a nominally 50kHz PMUT (Piezoelectric Micromachined Ultrasonic Transducer) together with a 2nd generation ultra-low power SoC (System on Chip) in a miniature, reflowable package. The small and thin package with bottom port configuration allows an easy integration. The ICU-30201 provides accurate range measurements to targets at distances up to 9 m. Based on ultrasonic pulse-echo measurements, the sensor works in any lighting condition, including full sunlight, and provides millimeter-accurate range measurements independent of the target's color and optical transparency. The sensor's Field of View (FoV) can be customized up to 180° and enables simultaneous range measurements to multiple objects in the FoV [39], [40]. In this work the US sensor with a FoV at 55° was selected as shown in Figure 2.

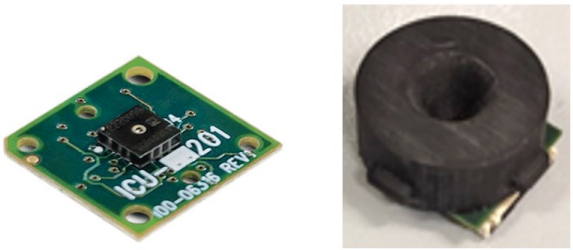

Figure 2. ICU-30201 ultrasonic transceiver without horn and with 55° FoV horn

In this work a commercial kit was used, which includes the DK-x0201, a development platform for ICU family of miniature US transceivers.

The following hardware is needed to the establish connection and set the platform in order to evaluate the sensor.

i) The development/host board - DK-x0201

ii) The daughter/evaluation board - PN100-06351

iii) ICU-20201 module(s) - EV_MOD_ICU-20201-00-0x

iv) Flat Flex Cable(s)

v) USB Cable

The image below (Figure 3) shows the physical connections to run the ICU-20201 EVK. Up to 4 sensors can be attached to the connector on the daughterboard using the flat flex cable (FFC).

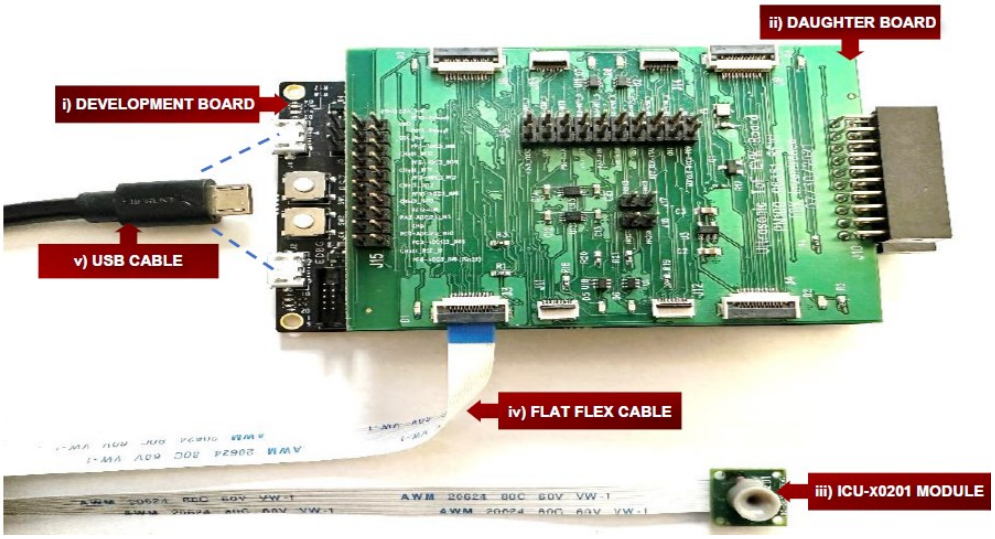

Figure 3. Evaluation platform hardware setup

3.1.1 Ultrasonic sensors: working principle

Ultrasonic sensors are active sensors, which actively transmit and receive signal to remotely perceive its environment. They are a powerful tool in performing distance measurements without making physical contact from the viewpoint of performance and cost.

Ultrasonic spectrum starts from 20 kHz to 200 MHz, that is just above the human audible range.

Depending on the field of application, different kinds of measuring systems and sensors have been used, but the problems related to echo detection are quite the same in every case: attenuation and beam spreading, presence of noise and interference, sensitivity to temperature and humidity and poor resolution.

The US sensor is a piezoelectric transducer, which is able to convert an electrical signal into mechanical vibrations, and mechanical vibrations into an electrical signal.

In many ultrasonic distance measurements, the operating principle is based on the calculation of ToF, estimated once the echo is registered by the US receiver, as shown in Fig. 4. The transducer generates and emits an US pulse which propagates through the medium and is reflected back towards the sensor by an object. The object distance d is correlated to ToF and can be estimated indirectly by the formula showed (Eq. 1), where *v* represents the propagation velocity of the US wave in the medium (through the air is 343.4 m/s at a temperature of 20 °C) and the 2 is the round-trip of the echo signal [41], [42].

$$\mathrm{d} = \frac{v * ToF}{2} \qquad (1)$$

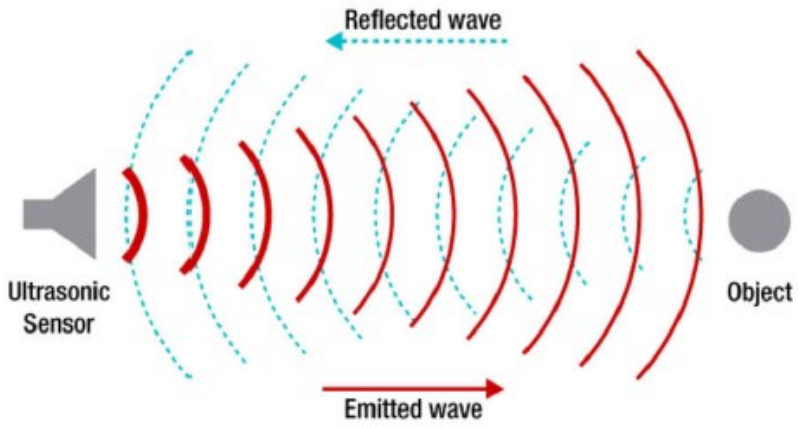

Figure 4. Ultrasonic Time-of-Flight Measurement

Ultrasonic level measurement techniques employ two categories of ultrasonic sensing device: those based on cavity-resonance technique, and those based on sonic-path method.

The first ones derive indirectly the level estimation from the cavity volume; the latter measures directly the fluid level by converting the ToF between pulse emission and echo detection.

As regards the latter strategy, two kinds of sensing are possible, i.e. the continuous wave method and the pulse-echo method, as shown in Fig. 5.

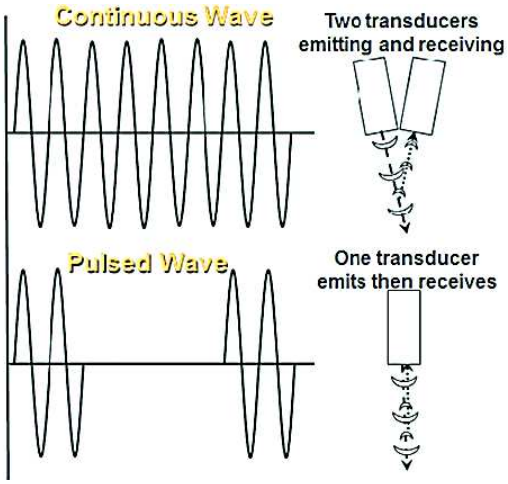

Figure 5.Continuous and pulsed wave methods

The continuous wave method uses two separate transmitting and receiving elements and evaluate the level through a phase shift estimation. This method obtains good performance, but requires a more complex hardware system.

The pulse-echo method, on the other hand, uses only one transducer that operates alternately in transmitting and receiving mode. The pulse-echo method has the advantage that offers a simple and low-cost solution, even if it has poorer results due to the lower time delay measurement accuracy, but they can be improved with software signal processing.

The velocity of an air coupled US echo is influenced by external environmental parameters, such as temperature, humidity, and in-band ambient noise. The sensing range decreases as temperature increases. Although the sensing rate also decreases as humidity increases, this can often be neglected, as the effects are minimal. The rate of attenuation across temperature and humidity is non-linear [41], [42], [43].

3.1.2 Ultrasonic sensors: Integration

US sensors and their platform are interfaced to a Raspberry Pi 4 (Figure 6) for real-time data acquisition and storage.

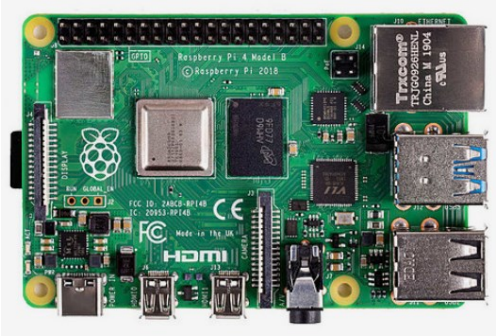

Figure 6. Raspberry Pi 4 model B

Raspberry Pi is a versatile single-board computer (SBC) widely used to develop programming skills and to build hardware projects due to its interface easy to manage. It has a 1.5 GHz 64-bit quad core ARM Cortex-A72 processor and there are available on the market different versions and models of the board. Indeed, the Raspberry Pi model can go from 1 to 5 and its versions can be A, B, or B+. Raspberry is equipped with four USB 2.0 ports that can connect several peripherals like mouse, keyboard, Wi-Fi adapters and all kinds of monitors thanks also to the HDMI ports. Using all these peripherals Raspberry becomes a fully functional computer. Additionally, the board presents an Ethernet port for network connectivity and 40 general-purpose input/output (GPIO) pins for interfacing with electronic components such as physical sensors and other devices (LEDs, modules, boards).

Based on the functionalities and the efficiency that characterize this SBC, Raspberry Pi 4 Model B has been chosen in this work as the board for a correct and reliable integration procedure with the ultrasonic sensors kit.

3.2 Environmental Sensors

Environmental sensors are used to assess the environmental comfort in the following domains: thermal, acoustic, visual and indoor air quality.

It’s important to optimize environmental conditions in the office to improve well-being of occupants to and increment work efficiency. There is the need to obtain accurate data on the

quality of the working environment and the activity of people in the office in order to adapt working environments to user needs.

For this reason, in this study multi-parametric sensors for the measurement of environmental parameters in indoor scenario were used. The parameters of interest were: temperature (T), relative humidity (rh), noise, CO_2, volatile organic compound (VOC), particulate matter (PM) with different diameters and brightness. After a market analysis, 3 types of available low-cost environmental sensors of different brands were selected:

- Netatmo (Boulogne-Billancourt, France), able to measure T, rh, CO_2 concentration, and noise level; it's shown in Figure 7.
- Sensibo (Redwood City, CA, United States), able to measure T, rh, CO_2 concentration, ethanol (EtOH), VOC and PM with a diameter of 2.5 microns (PM2.5); it's shown in Figure 8.
- HibouAir (Sollentuna, Sweden), able to measure T, pressure (P), PM with different diameters (PM2.5 and PM10), VOC, and ambient light; it's shown in Figure 9.

The measurements of environmental parameters were taken at regular intervals during the tests as shown in Table 1. Acquisitions saved the data in an internal database.

Figure 7. Netatmo NWS01

Figure 8. Sensibo Elements

Figure 9. HibouAir Particle Sensor AQI Monitor

3.3 Experimental Set Up

The test population consisted in 10 healthy volunteer subjects, in particular 6 females and 4 male subjects aged 27.4 ± 2.57 (expressed as mean ± standard deviation). All the participants were not subjected to a clothing regime: they were simply asked to wear their everyday clothing to reflect realistic thermal sensations, ensuring the subjectivity of the tests.

To ensure privacy and confidentiality, all data were anonymized and handled in strict accordance with GDPR regulations. Prior to participation, each volunteer provided informed consent by signing a consent form. All the experimental tests were carried out in compliance with the WMA Declaration of Helsinki [44] and the research design was set in accordance with

the ethical standards of the Ethical Committee of Università Politecnica delle Marche (UNIVPM), Ancona, Italy.

Before starting the tests, the full protocol was explained in detail to the participants, with a focus on comfort and well-being related aspects. The tests took place in two days, i.e., February 27th, 2024 and March 4th, 2024 in the room 014 of the Department of Industrial Engineering and Mathematical Sciences (DIISM) of Università Politecnica delle Marche (UNIVPM). The room (Fig. 10) has dimensions of (513×296×300) cm and inside there are positioned two office desks ((180×80×72) cm and (120×80×72) cm, respectively). As the tests involved a single user, only one workstation was used. The environmental sensors (one per type) were strategically located on the subject's desk, precisely at a distance of 85 cm and 132 cm from the southwest corner of the room. Three ultrasonic sensors (US_1, US_2, US_3) were attached side by side to the Netatmo sensor (Fig. 10, 11, 12).

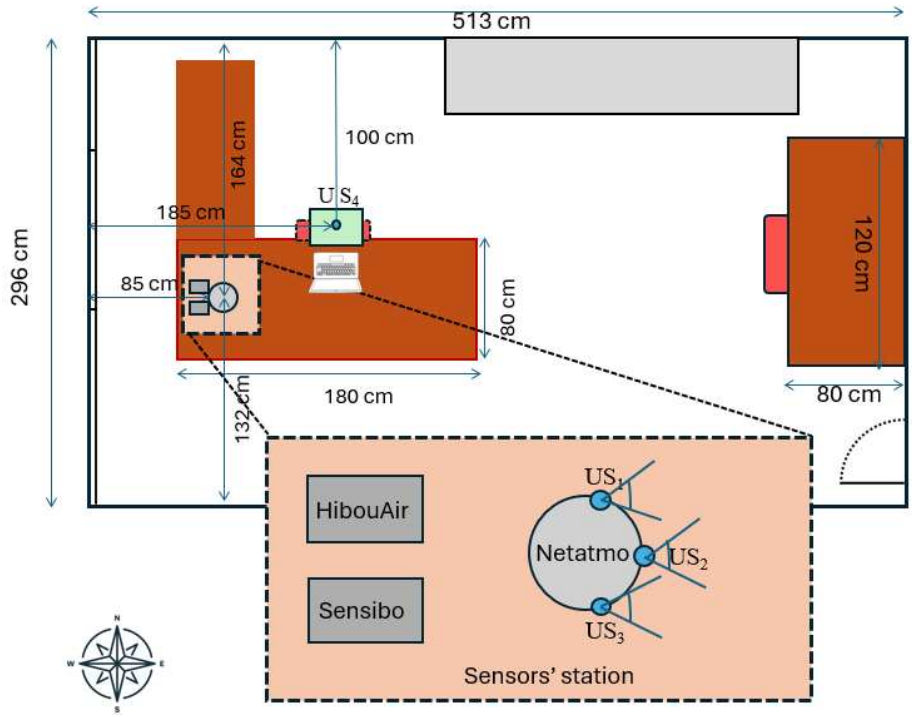

Figure 10. Test room plant. At the bottom of the figure the sensors station used during the test is reported. The environmental sensors are located on one workstation. The US sensors are arranged in a circular configuration on the Netatmo sensor.

Figure 11. Test room arranged for the experiments

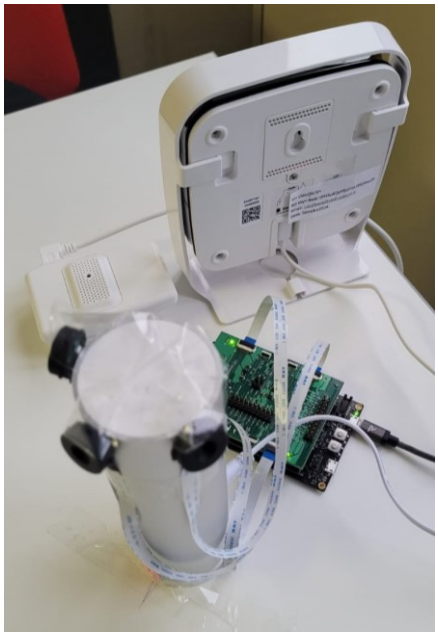

Figure 12. Sensors network strategically positioned over the desk

As the US sensors present a FoV of 55°, their arrangement was considered in order that the US_2 looked straight to the user, while US_1 and US_3 pointed outside the workstation. An additional US sensor (US_4) was placed on the ceiling above the user's workstation, housed in a customized enclosure.

The office scenario activities included a list of five activities that are: writing on paper, typing on PC, talking in the phone, standing, and walking within the room as shown in figure 13. An initial setup phase preceded the start of each session, during which the participant was instructed to sit for 10 s and then stand for other 10 s. Subsequently, the actual test started. Each sequence of activities was repeated three times, with an approximate entire test duration of 25 min.

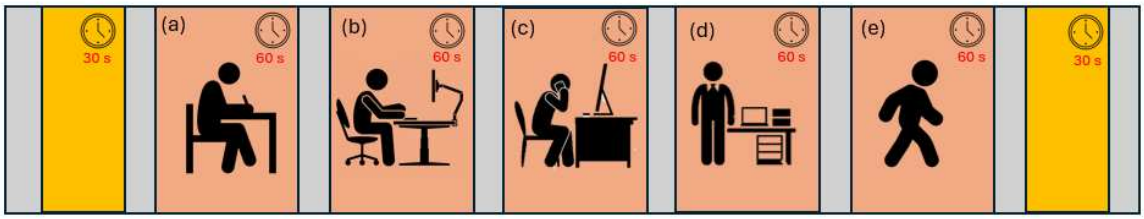

Figure 13. Sequence of activities performed by participants. The grey box indicates the rest period, the yellow box indicates the time in which the questionnaire was administrated to the participants.

The participants executed the following activities sequence:

- Writing on paper. The form was identical for all the participants and it was defined during the design of the test protocol;
- Typing on a provided virtual worksheet;
- Phone conversation reading given sentences to simulate a phone call;
- Standing;
- Walking within the room.

The sequence of activities was shown to the participants by means a web page on a PC located on the desk, an example of which can be seen in Fig. 14. The session began and ended with 30-s rest in sitting position and this item was repeated also between two consecutive activities for displacement signal calibration purposes.

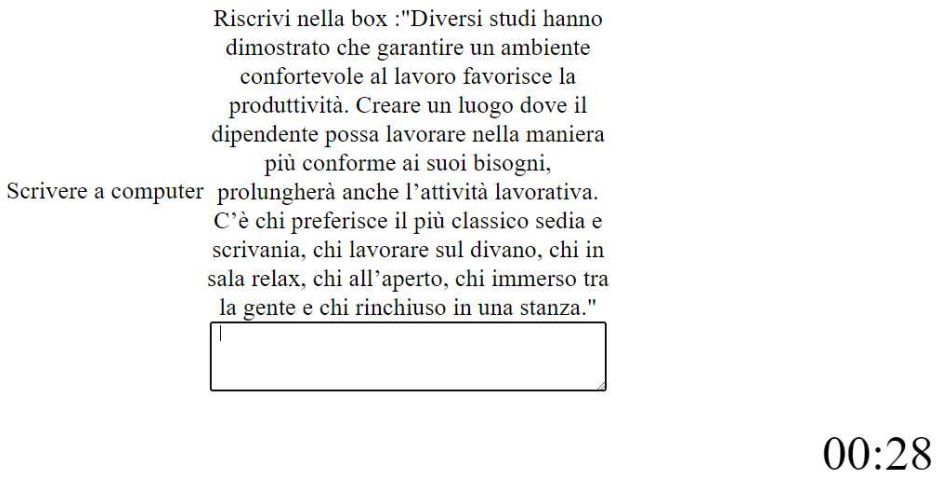

Figure 14. Screenshot of the web page used by subjects during the test. On the bottom right corner there is the time left to perform the showed activity.

During the test, a survey dealing with perceived comfort in different domains (i.e., thermal, acoustic, visual, and air quality) was administered at the beginning of the session and at the end of each sequence. The surveys were administered through a dedicated Google Forms and they lasted 30 s. In particular, the questionnaire consisted of two closed-ended questions, as shown in figure 15, with the following structure:

Q_1: *"Are you currently experiencing a state of comfort?"* Response options included: *"Yes"* (A_1(a)) or *"No"* (A_1(b)).

Q_2: *"If not, which aspect of comfort would you like to improve?"*. Response options included: *"Thermal"* (A_2(a)), *"Acoustic"* (A_2(b)), *"Visual"* (A_2(c)), *"Air quality"* (A_2(d)).

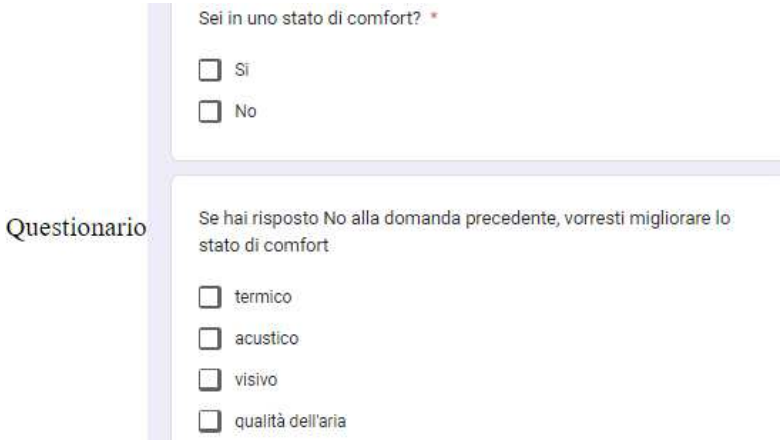

Figure 15. Surveys' layout dealing with perceived comfort in different domains.

3.4 Data Processing

The block scheme of the methodology employed to carry out the entire work is shown in figure 16. In the following subchapters, the various blocks are described. Once all the necessary signals have been acquired through the sensor setup, the amount of data was stored and processed in Python environment in real-time, following a well-determined procedure which is subsequently explained step by step. Since this study focuses on the feasibility of detecting office activities using a multidomain monitoring platform, the data processing methodology mainly involved the US sensor that looked straight to the user, installed on the workstation.

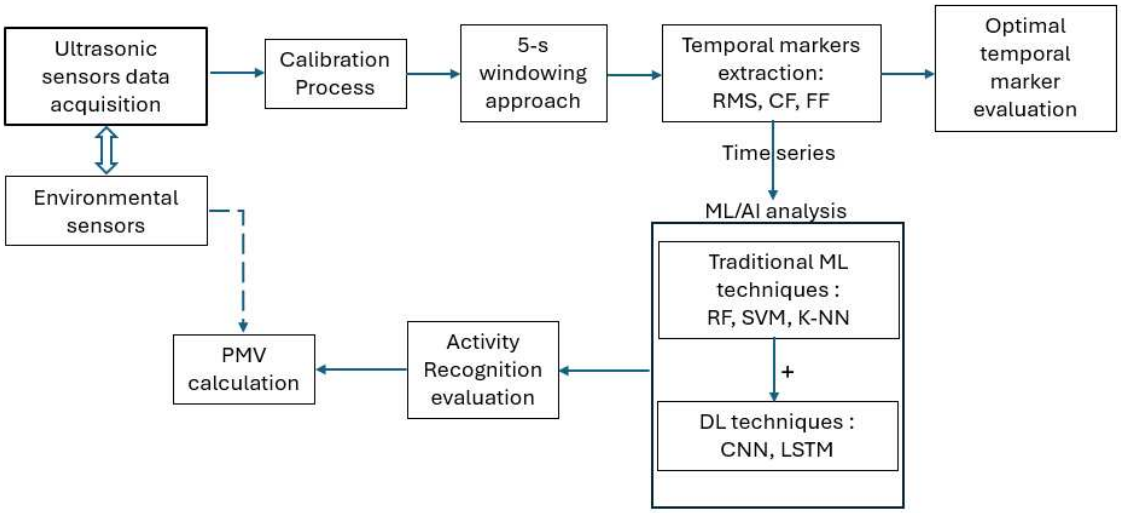

Figure 16. Block scheme of the methodology employed to predict the activities

3.4.1 Calibration Process

Initially, a calibration process was conducted on the ultrasonic signals to eliminate measured distances from furniture, walls, and the user's resting position. Specifically, for the first repetition of the sequence of activities, the average distance value recorded during the 30 s of the resting phase before the user started writing on paper were removed from the US distance data by subtraction. For subsequent repetitions, the rest phase after the user's movement within the room was considered for removal from the US distance data. This calibration process resulted in a US signal where was considered only the user's displacement. This

resulted signal is negative when the testing subject moves towards the US sensors and positive when the subject moves away from the US sensors.

3.4.2 5-second Windowing Approach

After the calibration, a time domain filtering procedure was applied to extract significant features from the US displacement data. Specifically, a 5-second windowing technique was performed. This method created non overlapping windows every 5 seconds, so during the 60 s of an activity are obtained 12 epochs. For each epoch three different temporal markers were calculated: Root Mean Square (RMS), crest factor (CF), and form factor (FF).

3.4.3 Temporal Markers Extraction

The RMS was computed by doing the square root of the mean value of the squares of n points of the signal (x_i) inside each window (Eq. 2).

$$RMS = \sqrt{\frac{1}{n} \sum {x_i}^2} \quad (2)$$

The CF, also called Peak Factor, was computed as the ratio between the maximum of the absolute value and the RMS of each window (Eq. 3).

$$CF = \frac{Peak\ Value}{RMS\ Value} \quad (3)$$

Instead, the FF was computed as the ratio between the RMS value and the absolute mean value of each window (Eq. 4).

$$FF = \frac{RMS\ Value}{Average\ Value} \quad (4)$$

3.4.4 ML/AI Analysis

In the subsequent analysis, that involves ML techniques, were selected all the three temporal markers obtained from the 5-s window approach of the displacement signal. Since they were calculated for each epoch, these features were analyzed in time series, creating a discrete dataset composed of thousands of samples.

The dataset was composed of 10 subjects, with 5 activities per subject and 3 repetitions. Once extracted the 3 temporal markers the dataset consisted in a total of 5400 samples in time series. The time series were given in input to ML models, that gave as output the label of the predicted activity.

The training and testing of the algorithms were done using the LOSO (Leave One Subject Out) technique where the dataset is split by subjects. This procedure consists in carrying out the training phase from the data of 9 of the 10 subjects iteratively and testing on the remaining subject. The models are fed with input-output pairs (time series-labels) from the dataset. The process was repeated by changing the subject used for the test each time to evaluate the capabilities of the algorithms on different individuals.

More in detail, supervised ML algorithms have been used to obtain the predicted activity starting from the data that have been acquired in the previous steps. Specifically, the ML algorithm that have been chosen accordingly to literature, were RF, SVM, K-NN. These traditional techniques, widely used in many researches [24], [28], were applied to compute a first classification. The aim of this procedure was to differentiate the detected patterns between activities present in the protocol and the non-activity phases, which coincides with resting. In this manner, an early binary classification is performed. ML performance metrics such as accuracy, precision, recall and f1 Score are used to analyze the performance of classifiers.

Subsequently were applied more advanced AI technique, such as CNN and LSTM, with the goal of distinguish between the different activities performed by the subject during the test that are writing on paper, typing on pc, phone conversation and standing. These DL techniques were applied in cascade to the precedent classical classifiers, creating a sort of hybrid algorithm possibly able to provide enhancements in recognition accuracy. In literature, as

described before by Ali et al. [35], is applied a similar approach that combines Naive Bayes with CNN. These hybrid algorithms were compared also with the single NN to see possible differences in accuracy.

3.4.5 Activity Recognition Evaluation

The ability of ML classifiers is commonly evaluated using the confusion matrix. The confusion matrix, also known as the error or contingency matrix, able to provide a specific layout for assessing classifier performance. The measures included are true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), where:

TP = True Positive: correctly predicted positive cases.

TN = True Negative: correctly predicted negative cases.

FP = False Positive: incorrectly predicted positive cases.

FN = False Negative: incorrectly predicted negative cases.

These measures help determine the accuracy of the classifier's predictions. The most commonly ML performance metrics used to analyze the performance of classifiers for model evaluation, are Accuracy (Eq. 5), Precision (Eq. 6), Recall (Eq. 7), and F1 Score (Eq. 8).

Accuracy: it measures the overall correctness of the classifier.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Precision: it measures the accuracy of the positive predictions.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall: also known as sensitivity, it measures the ability of the classifier to identify all positive cases.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

F1 score: it's the harmonic mean of precision and recall, providing a balance between the two. The highest possible value of an F1 score is 1, indicating perfect precision and recall, while the lowest possible value is 0, if either the precision or the recall is zero.

$$F1\ Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (8)$$

3.4.6 PMV Calculation

In addition, to observe how the office activities, and thus the metabolic rate, impact on the subjects' comfort perception, the mathematical expression of PMV model proposed by Fanger (Eq. 9), was computed as:

$$PMV = f(ta, rh, va, tr, M, Icl) \quad (9)$$

where t_a (°C) is the indoor mean air temperature, rh (%) is the relative humidity, v_a (m/s) is the air velocity, t_r (°C) is the mean radiant temperature, M (met) is the metabolic rate, I_{cl} (clo) is the clothing insulation level.

To this aim, in Eq. 9 the T_a is 22.5 °C and the rh is 48%. These are the mean values of temperature and humidity extracted from the Netatmo sensor among the testing scenarios of all the subjects in the two days of experiment. The v_a was set to 0.15 m/s [45], the t_r was set to a mean value of 22.5 °C [46] and the I_{cl} was set to 1 [47] which is a reference value for winter clothing. Four different M values were taken from literature [48] as those typical for office activities. In particular were selected M values for the resting activity (1 Met), the phone conversation activity (1.2 Met), writing or typing activities (1.5 Met) and standing still (1.6 Met).

4. RESULTS

4.1 Displacement Graphs

In this section it's possible to see the results following the calibration process (section 3.4.1). Specifically, there are the results concerning the intra-subject displacement signals of the

subject 1 and subject 4. For each activity are shown the three repetitions performed by the subject during the test.

Regarding the subject 1, the activities displayed are resting (Fig. 17), writing on a paper (Fig. 18), typing on pc (Fig. 19) and phone conversation (Fig. 20).

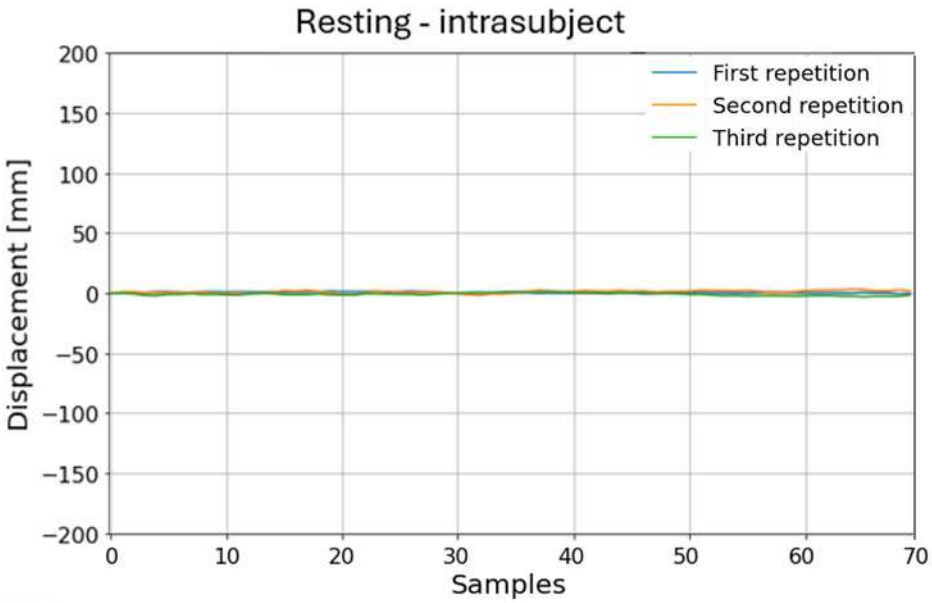

Figure 17. Intra-subject #1 displacement of the resting activity.

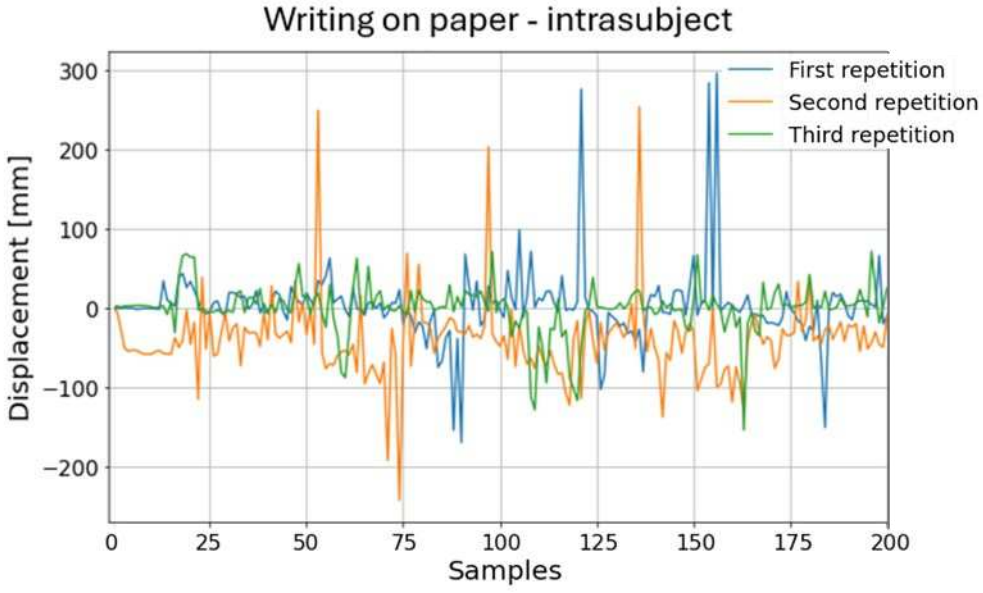

Figure 18. Intra-subject #1 displacement of the writing on paper activity.

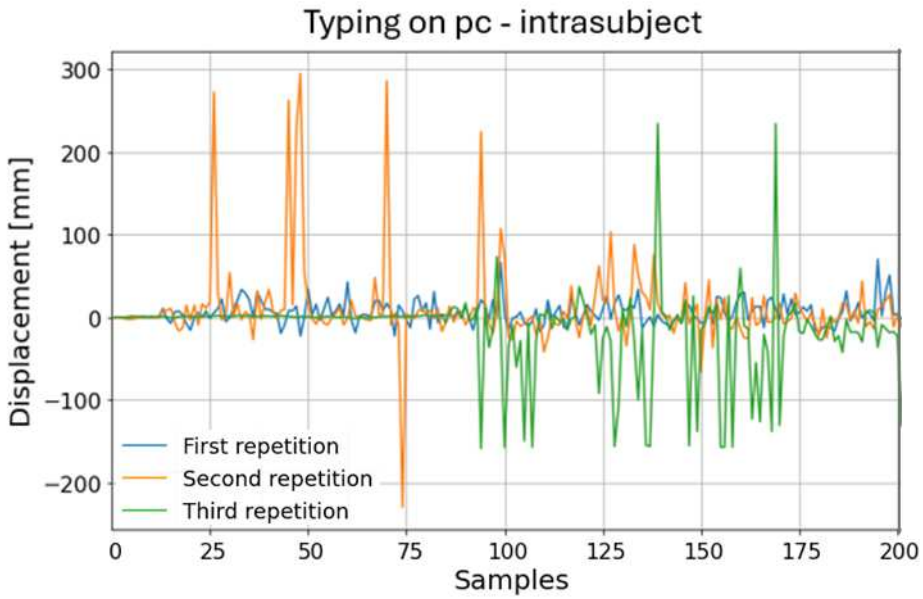

Figure 19. Intra-subject #1 displacement of the typing on pc activity.

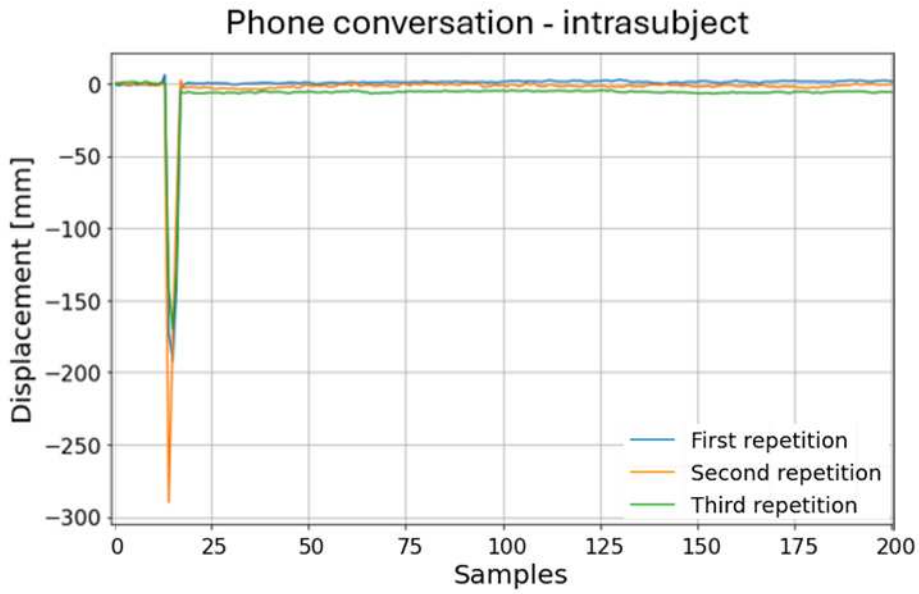

Figure 20. Intra-subject #1 displacement of the phone conversation activity.

Regarding the subject 4, the activities displayed are resting (Fig. 21), writing on a paper (Fig. 22), typing on pc (Fig. 23) and phone conversation (Fig. 24).

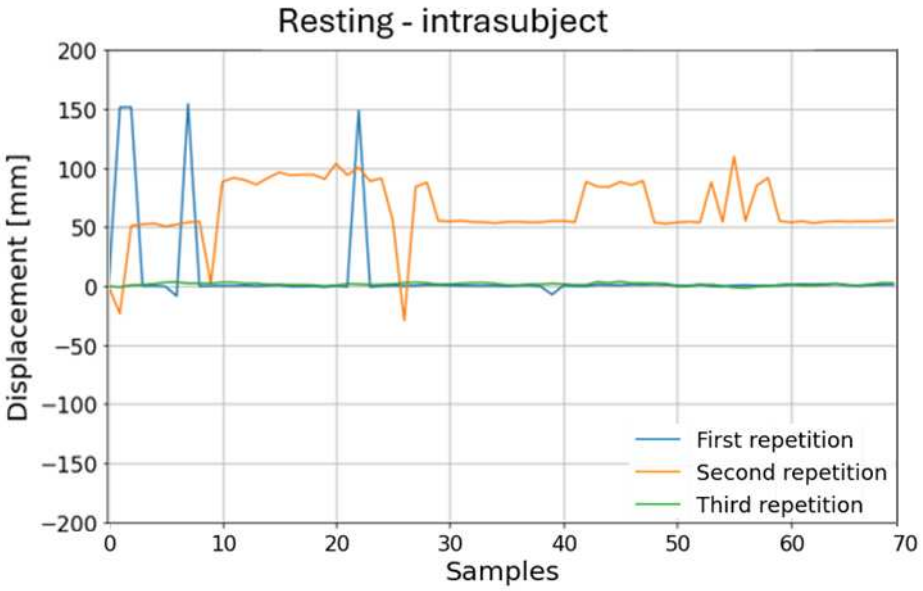

Figure 21. Intra-subject #4 displacement of the resting activity.

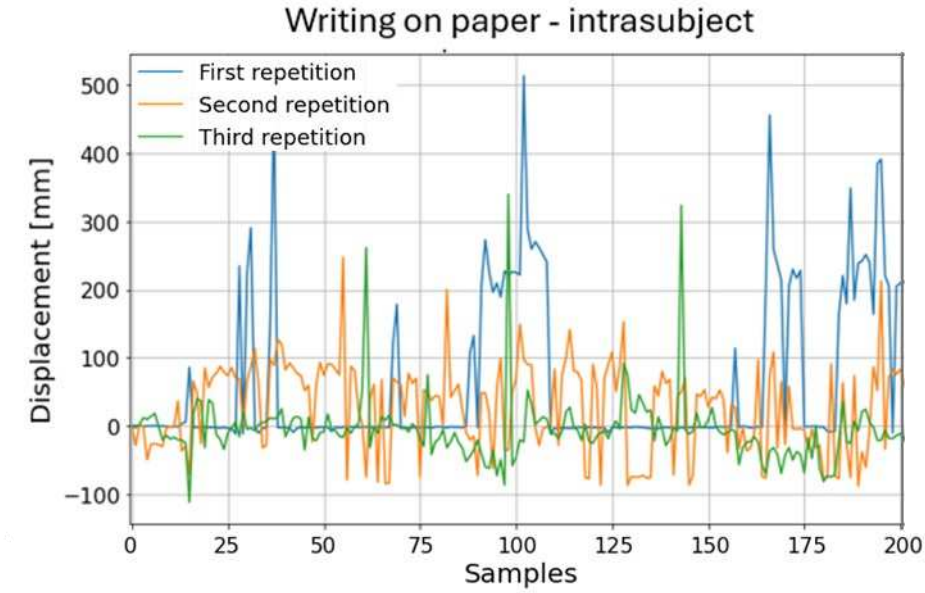

Figure 22. Intra-subject #4 displacement of the writing on paper activity.

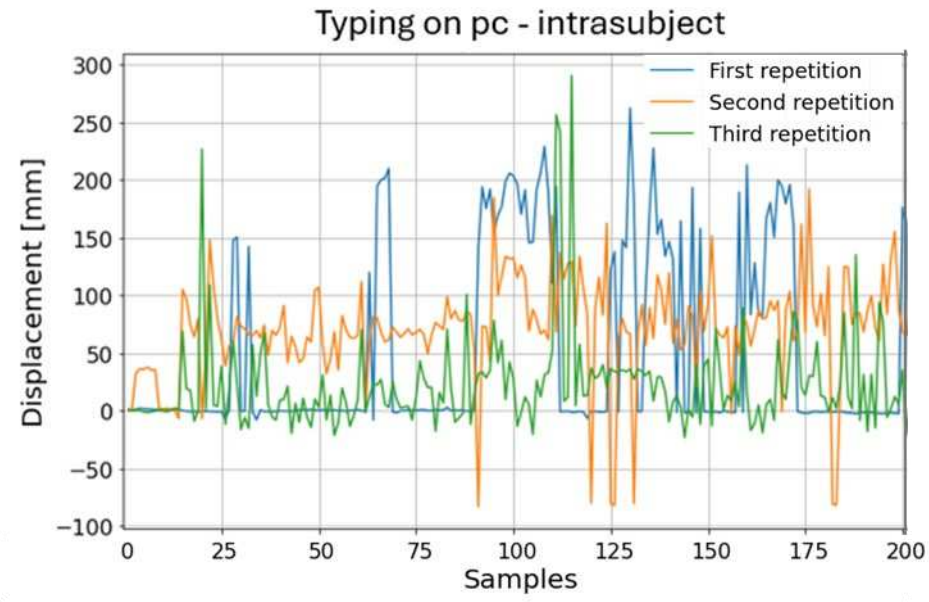

Figure 23. Intra-subject #4 displacement of the typing on pc activity.

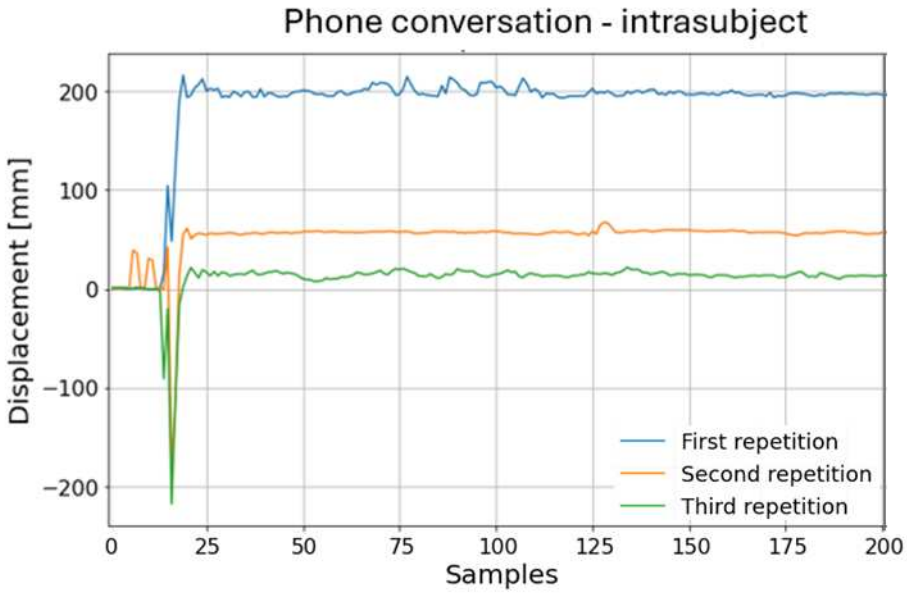

Figure 24. Intra-subject #4 displacement of the phone conversation activity.

4.2 Temporal Markers Computation

In this section it's possible to see the results of the temporal markers extraction (section 3.4.3). Specifically, are shown the results concerning the temporal descriptors values of the calculated using the 5-s windows approach.

Regarding the values of the subject 1 in the following figures are shown: RMS (Fig. 25), FF (Fig. 26), CF (Fig. 27) of four different activities that are resting, writing on a paper, typing on pc and phone conversation.

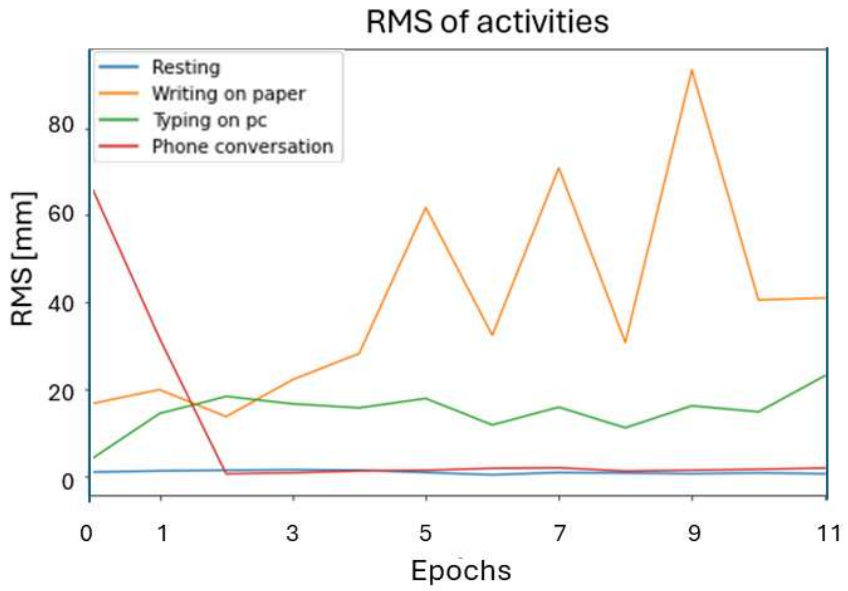

Figure 25. Intra-subject #1 RMS values calculated using the 5-s windows approach

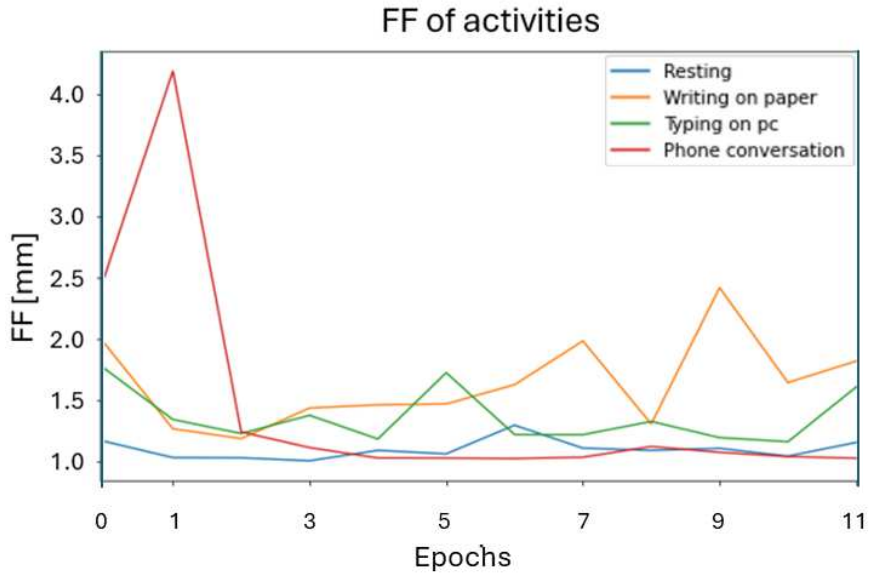

Figure 26. Intra-subject #1 FF values calculated using the 5-s windows approach

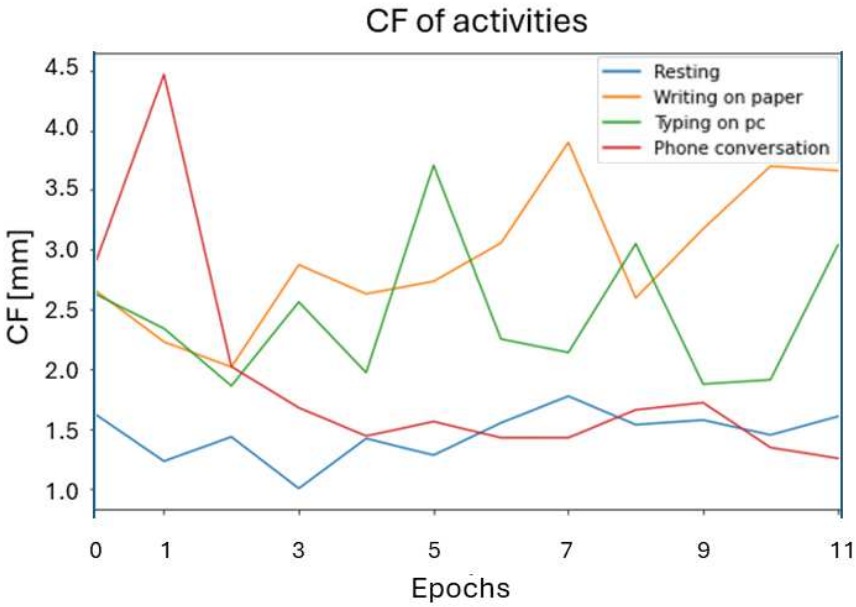

Figure 27. Intra-subject #1 CF values calculated using the 5-s windows approach

Regarding the values of the subject 4 in the following figures are shown: RMS (Fig. 28), FF (Fig. 29), CF (Fig. 30) of four different activities that are resting, writing on a paper, typing on pc and phone conversation.

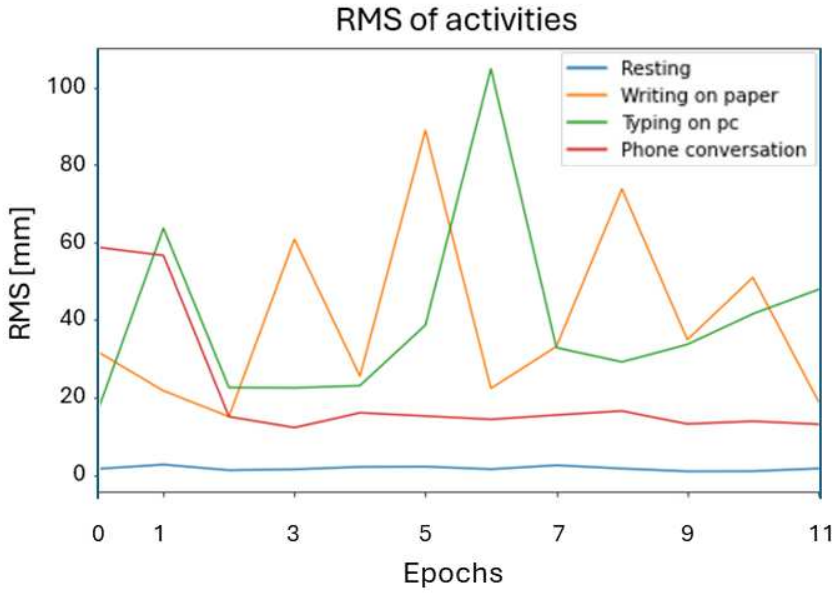

Figure 28. Intra-subject #4 RMS values calculated using the 5-s windows approach

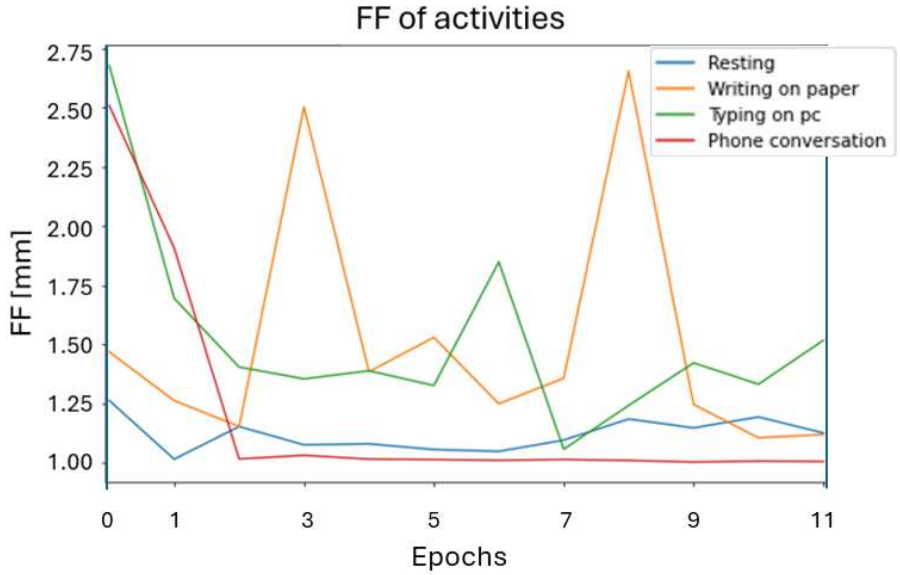

Figure 29. Intra-subject #4 FF values calculated using the 5-s windows approach

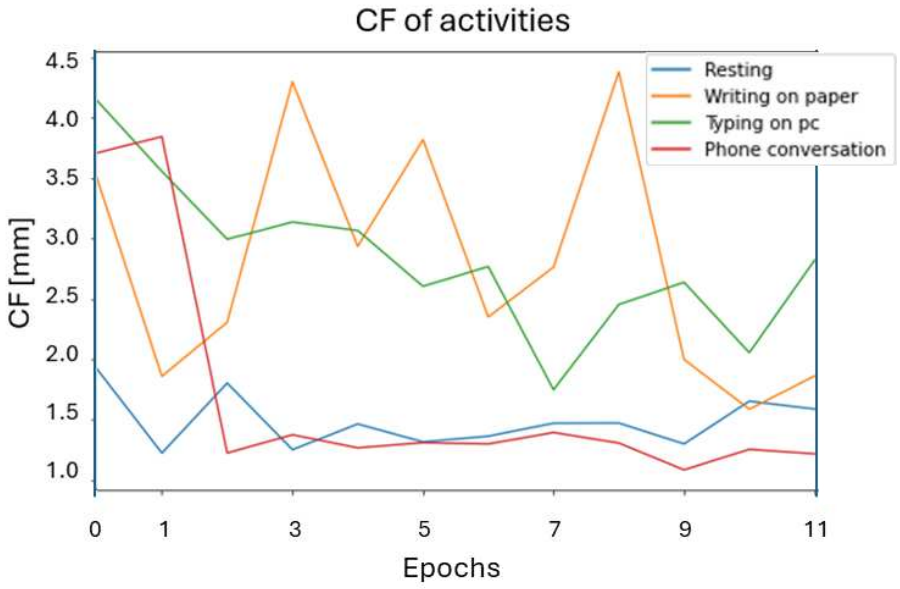

Figure 30. Intra-subject #4 CF values calculated using the 5-s windows approach

4.3 Inter-subject RMS Values

RMS resulted the most performant temporal marker, so its potential in discriminating between the different activities across all subjects is shown in Fig.31, where the mean and the standard deviation of the mean of the selected temporal marker were computed.

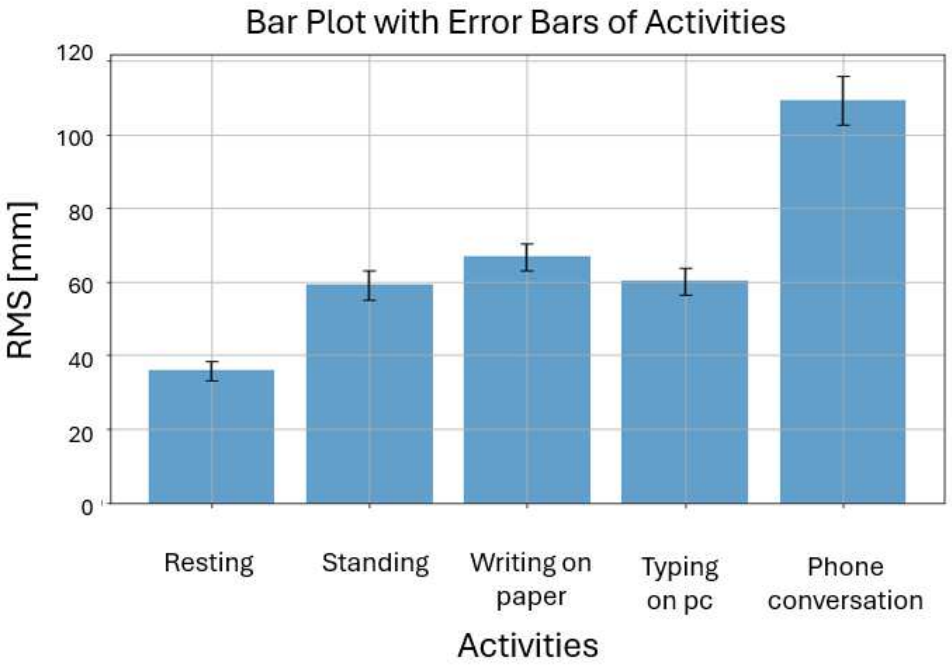

Figure 31. Inter-subject RMS values of the four activities selected (standing, writing on a paper, typing on pc, phone conversation) compared to the resting period.

4.4 ML/AI Classification

In this section are shown the results following the ML/AI analysis (section 3.4.4).

4.4.1 ML Binary Classification

In Table 2 are shown the results of the activity/non-activity classification: performance metrics such as accuracy, precision, recall and F1-score of different ML techniques like RF, SVM and K-

NN were computed. Then are shown the relevant confusion matrices: RF (Fig. 32), SVM (Fig. 33) and K-NN (Fig. 34).

Table 2. Binary Classification Results using traditional ML techniques

ML	Accuracy	Precision	Recall	F1 Score
RF	0,73	0,57	0,58	0,57
SVM	0,87	0,79	0,77	0,77
K-NN	0,85	0,74	0,70	0,70

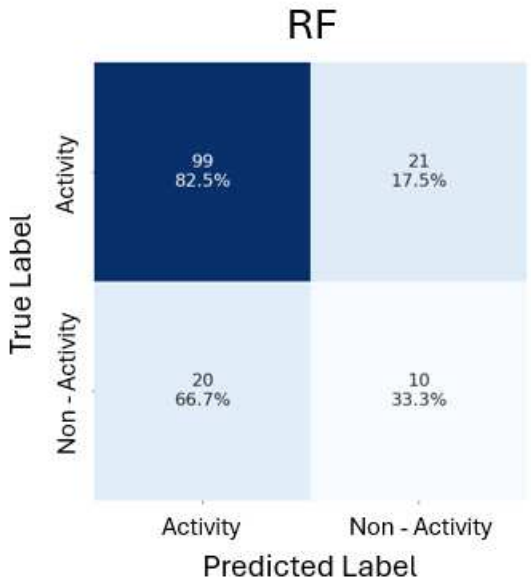

Figure 32. Confusion matrix of binary classification using RF

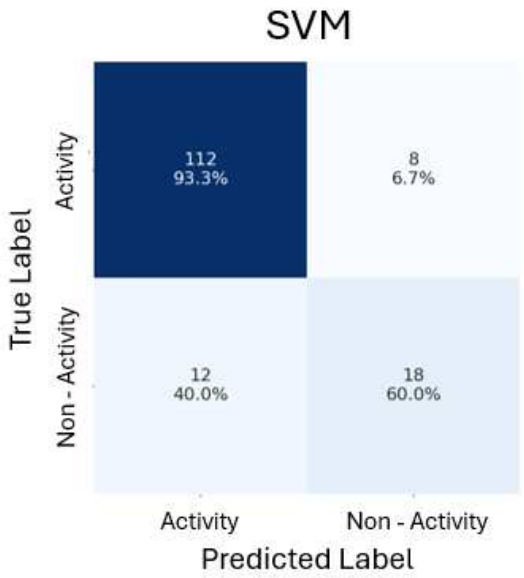

Figure 33. Confusion matrix of binary classification using SVM

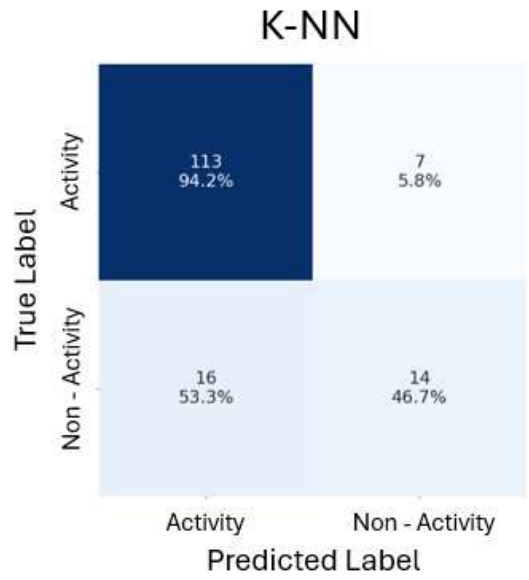

Figure 34. Confusion matrix of binary classification using K-NN

4.4.2 AI Activity Classification

In Table 3 are shown the results of the activities classification: performance metrics such as accuracy, precision, recall and F1-score were computed using CNN techniques. Then are shown

the relevant confusion matrices of multiclass classification of the 4 activities (excluding resting) with CNN technique:

- CNN on all the data considering the 4 activities (Fig. 35).
- CNN on data predicted by RF as activity in cascade (Fig. 36).
- CNN on data predicted by SVM as activity in cascade (Fig. 37).
- CNN on data predicted by K-NN as activity in cascade (Fig. 38).

Table 3. Activity Classification Results using CNN

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
CNN	0,78	0,78	0,78	0,77
RF+CNN	0,82	0,82	0,83	0,80
SVM+CNN	0,79	0,80	0,79	0,78
K-NN+CNN	0,82	0,84	0,82	0,81

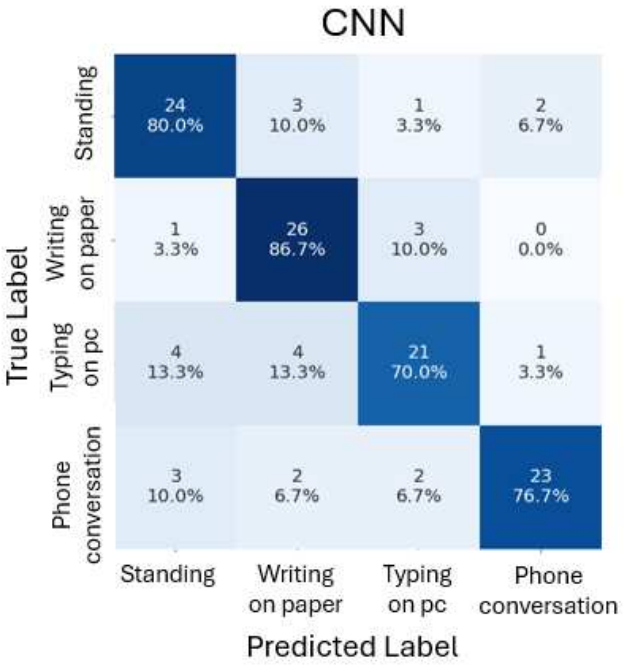

Figure 35. Confusion matrix of activities classification using CNN

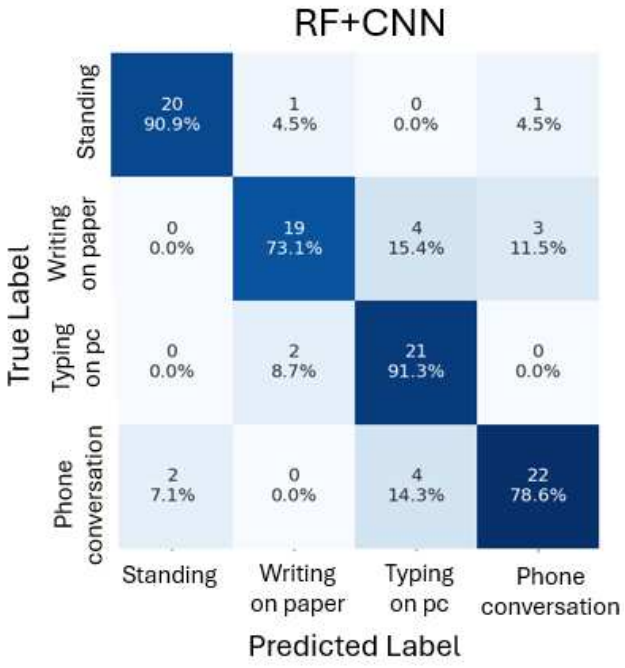

Figure 36. Confusion matrix of activities classification using RF+CNN

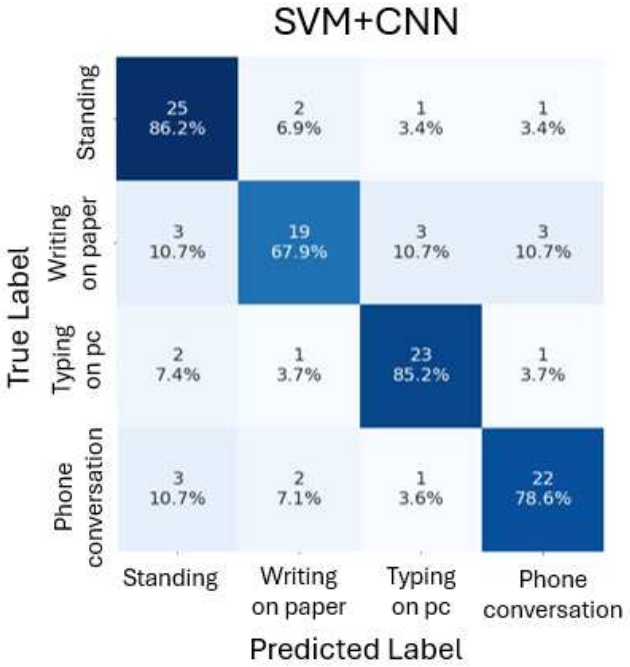

Figure 37. Confusion matrix of activities classification using SVM+CNN

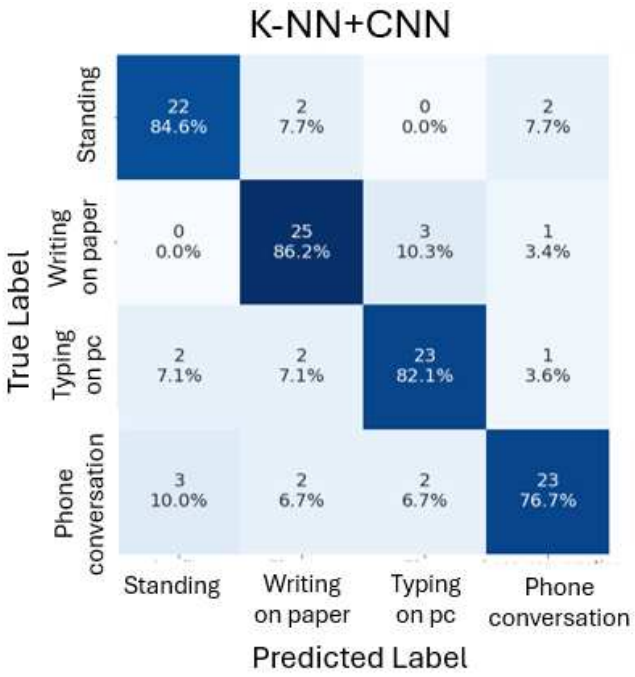

Figure 38. Confusion matrix of activities classification using K-NN+CNN

In Table 4 are shown the results of the activities classification: performance metrics such as accuracy, precision, recall and F1-score were computed using LSTM techniques. Then are shown the relevant confusion matrices of multiclass classification of the 4 activities (excluding resting) with LSTM technique:

- LSTM on all the data considering the 4 activities (Fig. 39).
- LSTM on data predicted by RF as activity in cascade (Fig.40).
- LSTM on data predicted by SVM as activity in cascade (Fig. 41).
- LSTM on data predicted by K-NN as activity in cascade (Fig. 42).

Table 4. Activity Classification Results using LSTM

Model	Accuracy	Precision	Recall	F1 Score
LSTM	0,64	0,68	0,64	0,64
RF+LSTM	0,67	0,70	0,68	0,65
SVM+LSTM	0,66	0,67	0,65	0,63
K-NN+LSTM	0,67	0,70	0,66	0,64

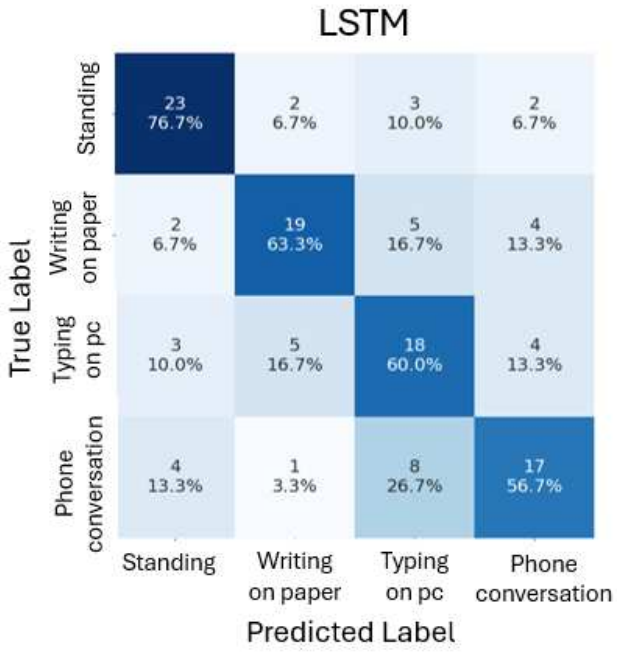

Figure 39. Confusion matrix of activities classification using LSTM

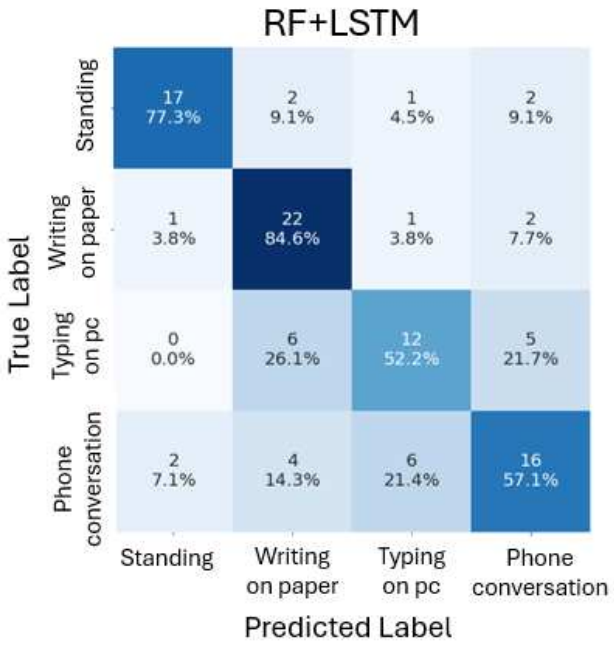

Figure 40. Confusion matrix of activities classification using RF+LSTM

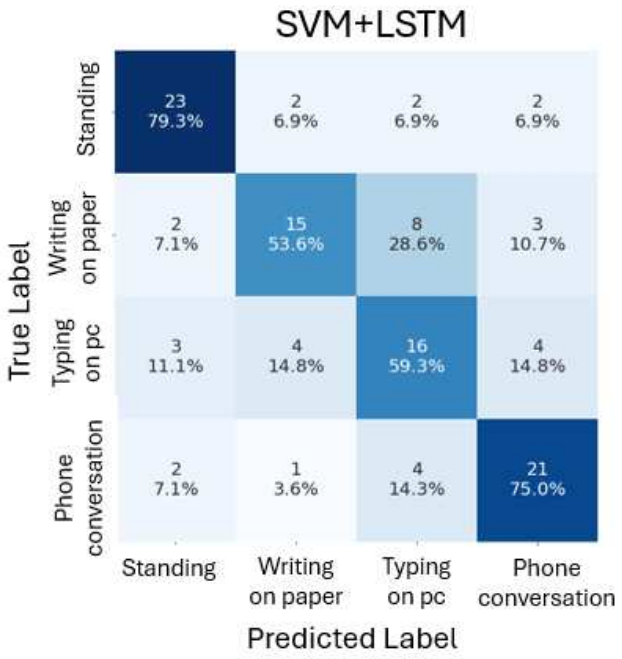

Figure 41. Confusion matrix of activities classification using SVM+LSTM

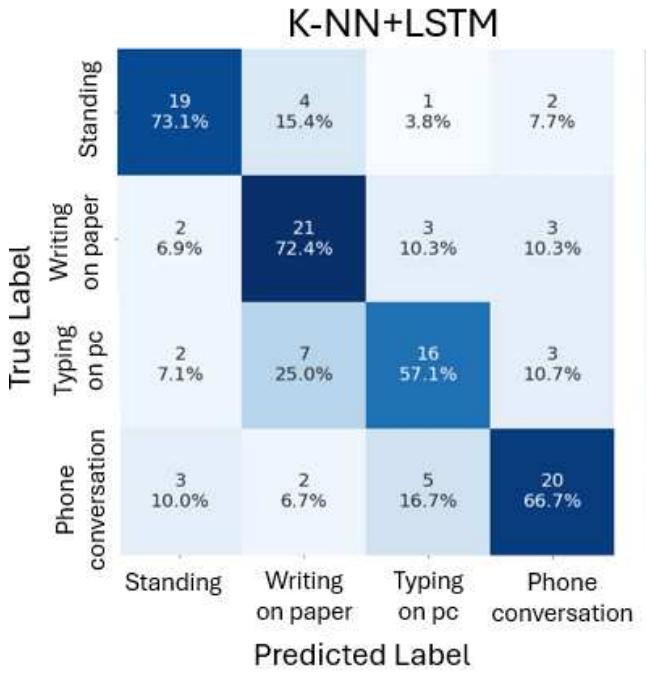

Figure 42. Confusion matrix of activities classification using K-NN+LSTM

4.4.3 PMV Calculation

In Table 5 are shown the results of the PMV calculation, according to Eq. 9 in section 3.4.6, based on the four different M values and considering the parameters taken from Netatmo sensor and from the literature. Each row of the table represents the values of a specific activity, specifically from the top to the bottom: resting activity, phone conversation activity, writing or typing activities and standing activity.

Table 5. PMV results based on the four different M values

	T_a [° C]	rh [%]	v_a [m/s]	t_r [%]	M [Met]	I_{cl} [clo]	PMV
1.	22.5	48	0.15	22.5	1.0	1	-0.38
2.	22.5	48	0.15	22.5	1.2	1	0.09
3.	22.5	48	0.15	22.5	1.5	1	0.50
4.	22.5	48	0.15	22.5	1.6	1	0.61

5. DISCUSSION

In this section the main points that emerged from the results of this thesis are presented.
One of the main findings of this work is that different activities can be associated with different patterns. The displacement signal related to a certain activity has a characteristic pattern with significantly high similarity in some activities, proving to be compatible in both intra- and inter-subject comparisons, as could be seen from Figure 17 to 24. Specifically, it's possible to see the signals of the telephone conversation activity carried out during the tests by the two subjects shown in Fig. 20 and 24: the displacement at the beginning of the test repetitions coincides

with the moment in which the subjects pick up the phone from the workstation, moving towards the US sensors.

To evaluate intra- and inter-subject variability in different activities, the temporal markers, RMS, CF and FF were computed using the 5-s windowing approach, obtaining 12 samples of each one per activity as shown in section 4.2.

The CF and FF temporal markers show limited discriminatory power between the two writing activities, exhibiting similar trends as shown in Fig. 26, 27, 29 and 30.

The RMS temporal descriptor emerged as the most suitable temporal marker among the computed ones. The potential of the RMS in differentiate between activities across all subjects is shown in Fig. 31, where it enables the discrimination between resting periods, telephone conversation activities, and writing, although it was unable to distinguish between the two different writing activities. The importance of discriminating office activities becomes evident when considering the PMV results based on different M values associated to the activities as shown in Table 5.

Activities such as resting and phone conversation have distinct RMS values, suggesting these can be easily discriminated. In particular the resting phase is clearly different from other activities, having significantly lower RMS values in the order of 35 mm, that reflect the minimal movement of the subjects. The same can be said for the phone conversation phases, that exhibit independent RMS values higher than 100 mm. Activities with overlapping mean +/- standard deviation, such as standing, writing on paper, and typing on PC, indicate a comparability among them, therefore with a poor discrimination ability considering the very similar RMS values of about 60 mm, as shown in Fig. 31. So, this approach is not sufficient to distinguish between activities that can have subtle differences. Additional features or methods may be needed for more accurate discrimination for these cases.

Based on these results, ML/AI algorithms are developed for HAR based on displacement signals obtained with US sensors, in particular considering these synthetic features that can be extracted through the signal processing.

Traditional ML techniques were applied (RF, SVM, K-NN) to compute a binary classification with the 2 classes activity and non-activity (resting). ML performance metrics of section 3.4.5 are used to analyze the performance of classifiers.

Results from binary classification in Table 2 show that K-NN and, in particular, SVM achieve a high ability to distinguish between activity and non-activity, reaching considerable accuracies of 85 and 87%. At this point it's applied a multiclass classification of the 4 activities (excluding resting) with CNN and LSTM techniques:

- CNN/LSTM on all the data considering the 4 activities (CNN/LSTM).
- CNN/LSTM on data predicted by SVM as activity in cascade (SVM+CNN/LSTM).
- CNN/LSTM on data predicted by K-NN as activity in cascade (K-NN+CNN/LSTM).

CNN in cascade traditional ML has slightly better performance in predicting activities through different classes respect than CNN applied directly on the activities. In fact, there is an increase of accuracies values from 78 to 82%, like shown in Table 3. It's possible to say the same thing by observing values on results that regards LSTM in Table 4, with accuracies that pass from 64%, of the only LSTM, to 67%, of LSTM in cascade to a classic ML technique.

Even if major part of activities is corrected predicted some common misclassifications can be observed, especially concerning two similar activities such as writing on a paper and typing on pc, whose predictions can be exchanged. It's possible to see these wrong classifications in the confusion matrices from Fig. 35 to Fig. 42.

It's useful to assign activity levels to the different office activities, because their contribution took part in the assessment of personal comfort through a precise estimation of metabolic rate. In fact, the importance of differentiate office activities becomes visible when considering the PMV results based on various M values (Table 5). The findings reveal the estimation of different PMV values according to each type of office activity, even if the metabolic rates were relatively low and similar. In fact, PMV index, on a scale from -3 to +3, vary from -0.38 to 0.61, remaining almost on neutral levels. These results reflect the activities, generally of low intensity, performed in an office environment, with a low metabolic impact. Therefore, a positive thermal sensation of the occupant is perceived recommending a general comfort state.

By considering a unique PMV defined by the average of PMVs for the four office activities, the findings reveal a 213% increase in PMV when the metabolic rate is adjusted for office activities compared to a metabolic rate defined for resting conditions.

For this reason, by using US sensors to monitor the human activities it's possible to have an optimization on the PMV calculation.

6. CONCLUSIONS

This work provides preliminary valuable insights towards the assessment of personal comfort, with the focus on the use of ultrasonic sensors to predict human activities. In particular using the AI, specifically the SVM for the binary classification and the CNN for the activity prediction, it's possible to reach an accuracy respectively of 87 and 79%. This is a result comparable to those obtained from literature, but with the advantage that our system is non-invasive and it's the first one that involves the integration of HARs in a multidomain service.

The main limitations are that the tests were performed on a small homogenous population, considering only single occupant in an office scenario. Future studies will involve a wide heterogeneous population, expanding the dataset size and capturing physiological variability. This will help in developing PCMs that are more inclusive and representative of different demographic groups.

To better reflect real-world applications will be considered multi-resident scenarios, exploring the analysis of interpersonal interactions in shared spaces.

Beyond the office environment others use cases, such as home and classroom scenarios, will be explored. This expansion will help in developing robust solutions for PCMs, able to ensure a wider applicability and effectiveness.

Results from US sensors, considered preliminary for the future extended experimental campaign within the framework of the WEPOP project, will be fused with data coming from diverse domain sensors, such as physiological and environmental ones. This integration will widely depict the well-being state of dwellers in their living environment; hence, the information can be exploited for optimizing thermoregulation of the environment and the related energy consumption.

REFERENCES

[1] Á. Borsos, E. Zoltán, É. Pozsgai, B. Cakó, G. Medvegy, and J. Girán, 'The Comfort Map—A Possible Tool for Increasing Personal Comfort in Office Workplaces', *Buildings*, vol. 11, no. 6, p. 233, May 2021, doi: 10.3390/buildings11060233.

[2] M. A. Ortiz, S. R. Kurvers, and P. M. Bluyssen, 'A review of comfort, health, and energy use: Understanding daily energy use and wellbeing for the development of a new approach to study comfort', *Energy and Buildings*, vol. 152, pp. 323–335, Oct. 2017, doi: 10.1016/j.enbuild.2017.07.060.

[3] M. Frontczak and P. Wargocki, 'Literature survey on how different factors influence human comfort in indoor environments', *Building and Environment*, vol. 46, no. 4, pp. 922–937, Apr. 2011, doi: 10.1016/j.buildenv.2010.10.021.

[4] S. Casaccia, K. Jokinen, R. Naccarelli, and G. M. Revel, 'Well-being and comfort of ageing people based on indoor environmental conditions: preliminary study on human-coach conversation', in *2022 IEEE International Workshop on Metrology for Living Environment (MetroLivEn)*, Cosenza, Italy: IEEE, May 2022, pp. 164–169. doi: 10.1109/MetroLivEnv54405.2022.9826905.

[5] A. Kaushik, M. Arif, P. Tumula, and O. J. Ebohon, 'Effect of thermal comfort on occupant productivity in office buildings: Response surface analysis', *Building and Environment*, vol. 180, p. 107021, Aug. 2020, doi: 10.1016/j.buildenv.2020.107021.

[6] S. Kawakubo, M. Sugiuchi, and S. Arata, 'Office thermal environment that maximizes workers' thermal comfort and productivity', *Building and Environment*, vol. 233, p. 110092, Apr. 2023, doi: 10.1016/j.buildenv.2023.110092.

[7] A. S. Silva, E. Ghisi, and R. Lamberts, 'Performance evaluation of long-term thermal comfort indices in building simulation according to ASHRAE Standard 55', *Building and Environment*, vol. 102, pp. 95–115, Jun. 2016, doi: 10.1016/j.buildenv.2016.03.004.

[8] F. R. d'Ambrosio Alfano, B. I. Palella, and G. Riccio, 'The role of measurement accuracy on the thermal environment assessment by means of PMV index', *Building and Environment*, vol. 46, no. 7, pp. 1361–1369, Jul. 2011, doi: 10.1016/j.buildenv.2011.01.001.

[9] K. Chen, Q. Xu, B. Leow, and A. Ghahramani, 'Personal thermal comfort models based on physiological measurements – A design of experiments based review', *Building and Environment*, vol. 228, p. 109919, Jan. 2023, doi: 10.1016/j.buildenv.2022.109919.

[10] K. W. Ma, H. M. Wong, and C. M. Mak, 'A systematic review of human perceptual dimensions of sound: Meta-analysis of semantic differential method applications to indoor and outdoor sounds', *Building and Environment*, vol. 133, pp. 123–150, Apr. 2018, doi: 10.1016/j.buildenv.2018.02.021.

[11] P. H. T. Zannin, A. M. C. Ferreira, 'Field measurements of acoustic quality in university classrooms', *Journal of Scientific and Industrial Research*. 68. 1053-1057. 2009.

[12] S. Kim and M. Li, 'Awareness, Understanding, and Action: A Conceptual Framework of User Experiences and Expectations about Indoor Air Quality Visualizations', in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12. doi: 10.1145/3313831.3376521.

[13] J. M. Daisey, W. J. Angell, and M. G. Apte, 'Indoor air quality, ventilation and health symptoms in schools: an analysis of existing information: Indoor air quality, ventilation and health symptoms in schools', *Indoor Air*, vol. 13, no. 1, pp. 53–64, Mar. 2003, doi: 10.1034/j.1600-0668.2003.00153.x.

[14] S. Liu, S. Schiavon, H. P. Das, M. Jin, and C. J. Spanos, 'Personal thermal comfort models with wearable sensors', *Building and Environment*, vol. 162, p. 106281, Sep. 2019, doi: 10.1016/j.buildenv.2019.106281.

[15] I. Ciuffreda et al., 'A non-intrusive ultrasound-based sensing technique for activity detection: proof of concept towards optimized personalized comfort.', *vol. 2024 IEEE International Workshop on Metrology for Living Environment (MetroLivEn)*, Chania, Greece, 2024, under publication.

[16] A. Ghosh, D. Chakraborty, D. Prasad, M. Saha, and S. Saha, 'Can we recognize multiple human group activities using ultrasonic sensors?', in *2018 10th International Conference on Communication Systems & Networks (COMSNETS)*, Bengaluru: IEEE, Jan. 2018, pp. 557–560. doi: 10.1109/COMSNETS.2018.8328272.

[17] A. Patel, C. Prabhudesai, and B. Aksanli, 'Non-Intrusive Activity Detection and Prediction in Smart Residential Spaces', in *2018 9th IEEE Annual Ubiquitous Computing, Electronics*

& Mobile Communication Conference (UEMCON), New York City, NY, USA: IEEE, Nov. 2018, pp. 355–361. doi: 10.1109/UEMCON.2018.8796576.

[18] G. Diraco, G. Rescio, P. Siciliano, and A. Leone, 'Review on Human Action Recognition in Smart Living: Sensing Technology, Multimodality, Real-Time Processing, Interoperability, and Resource-Constrained Processing', *Sensors*, vol. 23, no. 11, p. 5281, Jun. 2023, doi: 10.3390/s23115281.

[19] A. Das Antar, M. Ahmed, and M. A. R. Ahad, 'Challenges in Sensor-based Human Activity Recognition and a Comparative Analysis of Benchmark Datasets: A Review', in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, Spokane, WA, USA: IEEE, May 2019, pp. 134–139. doi: 10.1109/ICIEV.2019.8858508.

[20] F. Vittori, C. Chiatti, I. Pigliautile, and A. L. Pisello, 'The NEXT.ROOM: Design principles and systems trials of a novel test room aimed at deepening our knowledge on human comfort', *Building and Environment*, vol. 211, p. 108744, Mar. 2022, doi: 10.1016/j.buildenv.2021.108744.

[21] B. Fu, 'Sensor Applications for Human Activity Recognition in Smart Environments', 2021, doi: 10.26083/TUPRINTS-00017485.

[22] L. Alpoim, A. F. Da Silva, and C. P. Santos, 'Human Activity Recognition Systems: State of Art', in *2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG)*, Lisbon, Portugal: IEEE, Feb. 2019, pp. 1–4. doi: 10.1109/ENBENG.2019.8692468.

[23] P. Kumar, 'Human Activity Recognition with Deep Learning: Overview, Challenges & Possibilities'. Jun. 02, 2021. doi: 10.20944/preprints202102.0349.v3.

[24] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, 'Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey', *IEEE Access*, vol. 8, pp. 210816–210836, 2020, doi: 10.1109/ACCESS.2020.3037715.

[25] D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli, and I. O. Ellis, 'A "non-parametric" version of the naive Bayes classifier', *Knowledge-Based Systems*, vol. 24, no. 6, pp. 775–784, Aug. 2011, doi: 10.1016/j.knosys.2011.02.014.

[26] V. Podgorelec and M. Zorman, 'Decision Trees', in *Encyclopedia of Complexity and Systems Science*, R. A. Meyers, Ed., New York, NY: Springer New York, 2009, pp. 1826–1845. doi: 10.1007/978-0-387-30440-3_117.

[27] T.-H. Lee, A. Ullah, and R. Wang, 'Bootstrap Aggregating and Random Forest', in *Macroeconomic Forecasting in the Era of Big Data*, vol. 52, P. Fuleky, Ed., in Advanced Studies in Theoretical and Applied Econometrics, vol. 52. , Cham: Springer International Publishing, 2020, pp. 389–429. doi: 10.1007/978-3-030-31150-6_13.

[28] Khimraj, P. K. Shukla, A. Vijayvargiya, and R. Kumar, 'Human Activity Recognition using Accelerometer and Gyroscope Data from Smartphones', in *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, Lakshmangarh, India: IEEE, Feb. 2020, pp. 1–6. doi: 10.1109/ICONC345789.2020.9117456.

[29] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, 'Sensing Technology for Human Activity Recognition: A Comprehensive Survey', *IEEE Access*, vol. 8, pp. 83791–83820, 2020, doi: 10.1109/ACCESS.2020.2991891.

[30] N. S. Khan and M. S. Ghani, 'A Survey of Deep Learning Based Models for Human Activity Recognition', *Wireless Pers Commun*, vol. 120, no. 2, pp. 1593–1635, Sep. 2021, doi: 10.1007/s11277-021-08525-w.

[31] R. Raj and A. Kos, 'An improved human activity recognition technique based on convolutional neural network', *Sci Rep*, vol. 13, no. 1, p. 22581, Dec. 2023, doi: 10.1038/s41598-023-49739-1.

[32] A. Ghosh *et al.*, 'On automatizing recognition of multiple human activities using ultrasonic sensor grid', in *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, Bengaluru, India: IEEE, Jan. 2017, pp. 488–491. doi: 10.1109/COMSNETS.2017.7945440.

[33] A. Ghosh, A. Chakraborty, J. Kumbhakar, M. Saha, and S. Saha, 'HumanSense: a framework for collective human activity identification using heterogeneous sensor grid in multi-inhabitant smart environments', *Pers Ubiquit Comput*, vol. 26, no. 3, pp. 521–540, Jun. 2022, doi: 10.1007/s00779-020-01402-6.

[34] H. Griffith, F. Hajiaghajani, and S. Biswas, 'Office activity classification using first-reflection ultrasonic echolocation', in *2017 39th Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society (EMBC), Seogwipo: IEEE, Jul. 2017, pp. 4451–4454. doi: 10.1109/EMBC.2017.8037844.

[35] A. Ali, W. Samara, D. Alhaddad, A. Ware, and O. A. Saraereh, 'Human Activity and Motion Pattern Recognition within Indoor Environment Using Convolutional Neural Networks Clustering and Naive Bayes Classification Algorithms', *Sensors*, vol. 22, no. 3, p. 1016, Jan. 2022, doi: 10.3390/s22031016.

[36] R. Tanigawa and Y. Ishii, 'hear-your-action: human action recognition by ultrasound active sensing'. arXiv, 2023. doi: 10.48550/ARXIV.2309.08087.

[37] K. Venkatesh, B. Barmada, V. Liesaputra, and G. Ramirez-Prado, 'Robust Features for Activities Recognition', in *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 6237–6239. doi: 10.1109/BigData47090.2019.9006089.

[38] W. Jiang *et al.*, 'Towards Environment Independent Device Free Human Activity Recognition', in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, New Delhi India: ACM, Oct. 2018, pp. 289–304. doi: 10.1145/3241539.3241548.

[39] Klibanov, Lev & Roy, Jason & Boldt, Paul, 'Preliminary analysis of TDK CH101 Ultra-low Power Integrated Ultrasonic Time-of-Flight Range Sensor'. 2020.

[40] S. Kianoush, S. Savazzi, V. Rampa, L. Costa, and D. Tolochenko, 'A Random Forest Approach to Body Motion Detection: Multisensory Fusion and Edge Processing', *IEEE Sensors J.*, vol. 23, no. 4, pp. 3801–3814, Feb. 2023, doi: 10.1109/JSEN.2022.3232085.

[41] Mubina Toa, Akeem Whitehead, 'Application Note on Ultrasonic Sensing Basics'. Texas Instruments.

[42] G. Andria, F. Attivissimo, and N. Giaquinto, 'Digital signal processing techniques for accurate ultrasonic sensor measurement', *Measurement*, vol. 30, no. 2, pp. 105–114, Sep. 2001, doi: 10.1016/S0263-2241(00)00059-2.

[43] P. Li, S. Chen, Y. Cai, J. Chen, and J. Li, 'Accurate TOF measurement of ultrasonic signal echo from the liquid level based on a 2-D image processing method', *Neurocomputing*, vol. 175, pp. 47–54, Jan. 2016, doi: 10.1016/j.neucom.2015.10.014.

[44] ‘WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects – WMA – The World Medical Association’. Accessed: Dec. 09, 2020. [Online]. Available: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

[45] R. Kosonen and F. Tan, ‘Assessment of productivity loss in air-conditioned buildings using PMV index’, *Energy and Buildings*, vol. 36, no. 10, pp. 987–993, Oct. 2004, doi: 10.1016/j.enbuild.2004.06.021.

[46] A. K. Roy Choudhury, P. K. Majumdar, and C. Datta, ‘Factors affecting comfort: human physiology and the role of clothing’, in *Improving Comfort in Clothing*, Elsevier, 2011, pp. 3–60. doi: 10.1533/9780857090645.1.3.

[47] S. Liu and A. Novoselac, ‘Air Diffusion Performance Index (ADPI) of diffusers for heating mode’, *Building and Environment*, vol. 87, pp. 215–223, May 2015, doi: 10.1016/j.buildenv.2015.01.021.

[48] M. Mansoubi *et al.*, ‘Energy expenditure during common sitting and standing tasks: examining the 1.5 MET definition of sedentary behaviour’, *BMC Public Health*, vol. 15, no. 1, p. 516, Dec. 2015, doi: 10.1186/s12889-015-1851-x.

ACKNOWLEDGMENTS

At the end of this master thesis, I would like to mention all those people who have been close to me in this intense course of study and personal and professional growth.
First of all, I would like to express my gratitude to my supervisor, Scalise Lorenzo, for the great availability shown since the choice of the topic.
I would like to thank my co-supervisors, Casaccia Sara and Ciuffreda Ilaria, and Prof. Cosoli Gloria, for the precious advice and indications provided during the realization of this work.
I would like to thank my family, in particular my parents for giving me the opportunity to start this journey and for their constant support.
Thanks also to my sister, Letizia, for his ability to always encourage me to do my best.
I want to thank my lifelong friends, with whom I shared carefree and unforgettable moments and also my study companion Enrico, with whom I shared efforts and laughs during this journey.
I thank my girlfriend Rachel, for always being by my side in these last months of university.
Lastly, to myself, for the tenacity and sacrifices that have allowed me to get this far.