

The use of off-the-shelf wearable sensors to analyze
daily-living activities and emotional state of a person at the
Smart Condo

by

Dillam Jossue Diaz Romero

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

University of Alberta

Abstract

The Smart CondoTM is a model condo embedded with a wireless sensor network, developed by an interdisciplinary team including researchers from Occupational Therapy, Industrial Design, Pharmacy, and Computing Science. The Smart CondoTM aims to support older adults, including those with physical and cognitive disabilities, to live independently longer [1]. Older adults with complex needs are often limited in their ability to perform necessary daily activities and may require task-specific supports. Continuous health-monitoring systems have the potential to enhance one's quality of life and help older adults live safely in their home. This thesis describes two studies in the use of off-the-shelf wearable sensors in order to investigate the feasibility of using wearable devices and usefulness in the Smart CondoTM.

In the first study, twenty-six participants spent a single two-hour session in the one-bedroom living environment, either alone or in pairs, and performed a scripted protocol of activities of daily living. Twelve of these participants wore the commercial smart eyewear device JINS MEME, which collected electrooculography (EOG), accelerometer and gyroscope data throughout their sessions. This study used an offline classification method to predict the participants' activities. The approach showed that this method yields equal or better results with a variety of activities compared to approaches that involve more restrictive wearable device setups. The results demonstrate the suitability of JINS MEME for recognition of activities of daily living and identify limitations associated with the current model.

In the second study, twenty-one participants viewed a sequence of images from the International Affective Picture System (IAPS) database. Using wearable sensor devices, we collected electroencephalography (EEG), electrooculography (EOG), and kinematic motion data as participants viewed the images; the participants also characterized their own emotional responses to the images. Participants then played the serious game "Whack-a-Mole," wearing the sensor devices, and played three levels of the game that required varying amounts of cognitive effort. This study describes the method for emotion recognition (Ensemble Classifier and Random Forest) in participants as they played the serious game. This approach showed that emotional state during a task can be determined accurately using data collected from wearable sensor devices, with and without self-reported measures.

Preface

This thesis is an original work by Dillam Jossue Diaz Romero. These studies received research ethics approval from the University of Alberta Research Ethics Board. The first study is named *Recognition of Daily Activities Using Environment Based Sensing* (No. Pro00073382) and the second study is named *Measuring engagement, emotional states and cognitive function with wearable sensors* (No. Pro00076184)

The first study with the paper named *The Activity Classification in Independent Living Environment with JINS MEME Eyewear* presented in Chapters 2. The relative work and conclusion of “Activity Classification in Independent Living Environment with JINS MEME Eyewear” paper were written with the assistance of Nicholas Yee. Also, Yee contributed to the ethics propose and the labelling of the data.

- **Diaz, D.**, Yee, N. Daum, C. Stroulia, E. and Liu, L. “Activity Classification in Independent Living Environment with JINS MEME Eyewear”. 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom) (2018): 1-9.

The related work presented in Chapter 3, as well as the conceptual framework presented in Chapters 3.1 , and 3.1.3 the evaluation of our emotion recognition framework in Chapter 3.5, will be submitted as a journal paper.

Acknowledgments

First of all, I am very grateful to my supervisor Dr. Eleni Stroulia, for the guide and for supporting my research with helpful and motivating discussions. Without Eleni Stroulia's support, I would not be able to meet outstanding researchers from all over the world and to have the experience to do research in another country. This research was supported by the AGE-WELL NCE, through funding of Work Package 6.

I am very thankful to Antonio Miguel-Cruz, Adriana Rios, and Lili Liu for being great mentors, and for all the guidance and discussions we had in our meetings. I would also like to thank everyone who made my stay in Edmonton memorable.

Finally, I would like to express my deepest gratitude to my parents (Raul Diaz and Rafaela Segunda Romero) who have shown unconditional love and support at every step in my life. I am especially thankful to my mother for her moral support and for making some of my harder times a lot easier, and my brother and sister for their never-ending encouragement and support. I also would like to thank my lab team (Kalvin Eng, Victor Fernandez, Diego Serrano and Christoph Sydora), Felipe Niño, Nicholas Yee, Cesar Tavera, Funda Arslan, Pedro Rivera, and Houston Quintero for their ongoing support in all respects of my life.

Contents

List of Tables	vii
List of Figures	viii
List of Symbols	ix
1 Introduction	1
2 Recognizing ADLs	5
2.1 Related Work	7
2.2 Protocol	9
2.3 Data Collection & Analysis	10
2.3.1 Sensor-Data Collection	11
2.3.2 Feature Extraction	12
2.3.3 Class Balancing	15
2.3.4 Classification	16
2.4 Findings	17
2.5 Summary	19
3 Measuring Arousal & Valence during Gameplay	21
3.1 Related Work	23
3.1.1 Emotional State, Arousal & Valence	23
3.1.2 Recognizing Emotion from Biosignals through Wearables	23
3.1.3 Emotion During Gameplay	27
3.2 Research Questions	28
3.3 Protocol	29
3.3.1 Study Procedure	29
3.3.2 Image Selection Procedure	31

3.3.3	Ground Truth for Arousal & Valence	32
3.3.4	Arousal & Valence From PANAS Questionnaire	33
3.4	Data Collection & Analysis	34
3.4.1	Data-Stream Synchronization	35
3.4.2	Feature Extraction	36
3.4.3	Learning a Classifier Model	40
3.5	Findings	41
3.5.1	Recognizing the Self-Reported Emotional States Through Biosignals	41
3.5.2	Correlation between Gameplay Difficulty and Self-Reported Emotional State	43
3.5.3	Predicting Player’s Arousal & Valence Through The Pre-trained Machine Learning Models	45
3.6	Summary	47
4	Conclusion and Future Work	48
4.1	Conclusion	48
4.2	Future Work	49
	Bibliography	51
A	Appendix Activity Detection	61
A.1	Activity Classification	61
A.1.1	Manual Label-Activity Mode	62
A.2	Classifies models – Parameters	63
A.3	Data Analysis Framework	63
A.4	Data-Stream Synchronization - MATLAB	63
A.5	Extracted Attributes	69
A.5.1	EOG Attributes	69
A.5.2	Accelerometer Attributes	70
A.5.3	Head Movement Features	71
A.6	Euler Angle Features	71
B	Appendix Questionnaires	73
B.1	Questionnaires	73

List of Tables

2.1	Activities of daily living included in this study.	10
2.2	Average Accuracy, Recall, and Precision from all participants by activity.	17
2.3	Accuracy related work.	18
3.1	Image selection procedure and used formulas for the selection of 4 images from each of the 9 categories.	32
3.2	Relationship between Circumplex and PANA models of classifying human emotions.	34
3.3	Feature extraction per signals.	37
3.4	30 Features selected for emotion recognition based on Random Forest. (The common features between Valence and Arousal are underlining and bold).	39
3.5	Comparison of mean performance across 10-fold user's average of the classifiers Logistic Regression, Random Forest, polynomial liner SVC, and Ensemble Learning.	42
3.6	Comparison of Valence and Arousal for the signals EEG, EOG, kinematic motion and their sum using the 30 most important features and SMOTE.	43
3.7	Accuracy between proposed method and similar works using Databases (DB).	44
3.8	Obtained results of arousal and valence predicted by playing the game Whack-a-Mole.	46
3.9	Accuracy between proposed method and similar works using different games.	46

List of Figures

2.1	Illustration of the pipeline for activity recognition.	11
2.2	The JINS MEME eyewear and embedded sensors [2].	11
2.3	The effect of a blink (1 - blue), upward saccade (2 - red) and downward saccade (3 - magenta) on EOG signal in the vertical dimension as detected by JINS MEME.	14
3.1	Two approaches to classifying human emotions. Circumplex model and Positive Activation - Negative Activation (PANA) model.	24
3.2	Open-source brain-computer interface(OpenBCI) is an affordable bio-sensing system, this device can collect electrical activity of human body such as EEG, EMG, EOG and ECG.[3]	29
3.3	Visual explanation of Binary Classification.	33
3.4	Flowchart of the approach for training the model using IAPS images and 10-fold cross validation.	35
3.5	Flowchart of the proposed approach to predict the emotional state of the participants while playing the game.	35
3.6	Mapping of PANAS as 2-D Russell diagrams for all participants. The y-axis presents the overall participants emotional states. For each participant, scores for valence and arousal for each level based on their self-reported emotional response (high = 1 and Low = -1) were assigned.	45
A.1	List of Activities	62
A.2	List of Posture and Motion Characteristics	63
B.1	Self-assessment of Emotion Response form	74
B.2	Self-assessment of PANAS gameplay experience [4]	75

List of Symbols

A	Arousal
ABS	Absolute
ACC	Acceleration
ADL	Activities of Daily Living
AVS	Arousal and Valence Space
BADL	Basic Activities of Daily Living
BL	Baseline
BLE	Bluetooth Low Energy
DB	Database
DC	Direct Current
EDA	Electrodermal Activity
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
FFT	Fast Fourier Transform
FI	Feature Importance
G-ACC	Glass Acceleration

G-GYRO	Glasses Gyroscope
GSR	Galvanic Skin Response
GYRO	Gyroscope
H-ACC	Head Acceleration
Hz	Hertz
IADL	Instrumental Activities of Daily Living
IAPS	International Affective Picture System
IBk	Instance-Based method based on k-neighbours
ID	Identification
JM	JINS MEME
LDA	Linear Discriminant Analysis
M	Male
NB	Naive Bayes
NH	Null Hypothesis
OBCI	OpenBCI
OOB	Out-Of-Bag
PANA	Positive Activation – Negative Activatio
PANAS	Positive and Negative Affect Schedule
RESP	Respiration
PIR	Passive Infrared
PLR	Pupil Light Reflex
PPG	Photoplethysmogram
PSD	Power Spectrum Density
RFID	Radio-Frequency Identification

SMOTE	Synthetic Minority Over-sampling Technique
SMO	Sequential Minimal Optimization
Std	Standard Deviation
STFT	Short-Time Fourier transform
SVM	Support Vector Machine
V	Valence
VR	Virtual Reality

Introduction

Due to rising life expectancy and declining fertility rates in many countries, the population of older adults relative to the total population is increasing rapidly. The number of people aged 65 years and over in the world is expected to rise from 602 million in 2015 to 1.51 billion in 2050 [5]. Older adults with complex needs are often limited in their ability to perform basic daily activities and may require task-specific support. The detection of activity recognition, daily patterns, and emotional states in an older adult can provide rich information on their capacities and functions. This thesis presents two studies about the use of off-the-shelf wearable sensors in order to investigate the feasibility of wearable devices in healthy participants at the Smart CondoTM for recognizing daily activities, and measuring of emotional states.

An interdisciplinary team including researches from Industrial Design, Occupational Therapy, Pharmacy, and Computing Science developed the Smart CondoTM, a model condo embedded with a wireless sensor network, and video cameras. The condo is located at the University of Alberta's Edmonton Clinic Health Academy (ECHA), to study how technology could be used to enable older adults, including those with physical and cognitive disabilities, a longer independent living [1].

At the very first beginning, the Smart CondoTM was designed as a living space in an office space, which was equipped with integrated sensors [6]. These primary devices had the potential to provide several services such as controlling heating and air-conditioning, light switches, temperature and light intensity [6]. Also, a software architecture for analyzing the recorded sensor stream, to extract sensor reading patterns corresponding to the occupants' activities, and information of interest to patients, their caregivers, and their health care providers [7] was deployed. All the data obtained by the sensors is transferred into a central cloud-based repository.

Today, the Smart CondoTM is a fully functional apartment with one bedroom, a bathroom, an open kitchen, and living space. Several studies started which required the addition of various sensors to the condo, such as radio-frequency identification (RFID), cameras, BLE stickers on appliances, and wearables. In 2015, Azghandi *et al.* developed a location-and-movement recognition using passive infrared (PIR) and RFID readers in the Smart CondoTM. This approach is limited in terms of efficient activity recognition because the RFID reader is required to be embedded in everyday items.

In 2017, Mohebbi *et al.* [8] equipped the Smart CondoTM with Bluetooth Low Energy (BLE) beacons attached to different objects in the house and a service running in the background on the occupants' smartphones. The study reported problems on the indoor location, the use of BLE-enabled stickers and beacons, and WIFI access points. The BLE beacons were used to transmit and report signal strength measurements to a nearby smartphone. This technique has the advantage to determine the location of two individuals accurately, but it is necessary to use BLE beacons in every day items to detect activity recognition.

The first part of the thesis involves activity recognition through the use of wearable sensors, intending to help older adults live safely in their homes and have an enhanced quality of life [9]. The second part involves measuring emotional response with wearable sensors, helping to build an understanding of the patterns of emotional states and cognitive difficulties in healthy individuals. Wearable sensors allow to collect different types of data (electroencephalography (EEG), electrooculography (EOG), acceleration, and gyroscope), store data on different aspects of human movement under the independent-living conditions, and enable a richer understanding of a person's physical activities and mental state [10].

In our research, low-cost wearable devices JINS MEME and OpenBCI were used for collecting EEG and EOG signals, respectively. JINS MEME is smart eyewear which can detect three types of data. Within the frame of the glasses are three EOG electrodes, an accelerometer, and a gyroscope [2]. This device represents eye movement velocity rather than a position as in conventional EOG signals, but is easily adaptable and it can be possible requirement for older adults. Therefore, the use of this type of technology can easily be integrated into daily life. On the other hand, OpenBCI is an open-source system that collects EEG signal and motion data. Sixteen electrodes contained within the headset collect data following the international 10-20 system, while an accelerometer embedded in the device's microcontroller board collects the motion signals. This device cannot collect the high-frequency component of the brain signal and some extra channels, but the signal is good enough for empirical studies [3]. Both devices are commercially available (JINS MEME is almost EOG, and OpenBCI is almost EEG). Consequently, it was of interest to investigate their balance of usefulness and practicality.

The first part of this thesis, *Recognizing Activities of Daily-Living (ADL)*, focuses on the area of activity recognition. In this study, twenty-six recruited participants had a one to two-hour session in a one-bedroom living environment in the Smart Condo™, either alone or in pairs where they performed a scripted protocol of activities of daily-living [11]. Twelve of these participants were asked to wear JINS MEME. The study results have demonstrated the suitability of the JINS MEME eyewear for activity recognition of daily-living and have identified some of the limitations associated with the current model of the device [11]. Overall, previous research provides a more fine-grained insight into a person's activities.

The second part of this thesis, *Measuring Arousal & Valence during Gameplay*, focuses on the area of emotional response to challenging cognitive activity. We collected data from twenty-one recruited participants, where each was presented with a series of images from the IAPS database. Wearable sensor devices collected EEG, EOG, and kinematic motion as participants viewed the images; the participants also characterized their own emotional responses to the images [12]. Then, the participants were presented with the serious game "Whack-a-Mole". The serious game contains few rules and features a simple touch-screen interface, making it appropriate for cognitively healthy older adults as well as those with up to moderate dementia [13]. During gameplay, successive levels require progressively greater cognitive skills. The software embedded in the game captures gameplay data in an anonymized file.

In this case we used the game "Whack-a-Mole", developed for the assessment and training of visuomotor, cognitive and mental health conditions [14]. While wearing the sensor device, the participants played three levels of the game, each requiring different levels of cognitive effort. A methodology was created to measure the emotional response by integrating diverse types of wearable sensors, pre- & post-processing tools for data analysis, and machine learning algorithms. This study shows that the extraction of indications for arousal and valence can imply the increase of cognitive load.

The results obtained from these studies have the potential to lay the necessary groundwork for identifying the emotional states in people through objective measures, and advancing the activity and emotion recognition state of the art through off-the-shelf wearable sensors. The long-term contribution of this work will be to enhance health monitoring systems for people with cognitive disabilities by the integration of new wearable devices in the Smart Condo™. These devices improve state of the art in some way such as, reducing costs, enhancing accessibility, improving the quality of life, providing an alternative system of evaluation and enabling automated assisted living systems, and providing suitable adaptive care for the caregiver and clinicians.

The remainder of this thesis is organized as follows. In Chapter 2, related work concerning the

JINS MEME device, and activity recognition with and without wearable sensors are summarized (Chapter 2.1). Afterwards, the process of activity classification and the effectiveness of pre- & post-processing techniques are discussed. The results are presented and compared to the literature. In Chapter 3, the method of classifying valence and arousal during gameplay is described, followed by the discussion of background information and related work using wearable sensors to measure and recognize emotional states during gameplay (Chapters 3.1 and 3.1). Further, a comparison with results obtained with and without self-reported measures are included. Finally, Chapter 4 concludes the thesis and discusses future work.

Recognizing ADLs

Older adults with complex needs are often limited in their ability to perform basic daily activities, and they may require task-specific support. With continuous health-monitoring systems, the ability to recognize people's activities in their homes can enable automated assisted living systems, caregivers and clinicians to provide suitable adaptive care. With the advent of miniaturized sensing technology, which can be wearable, it is now possible to collect and store data on different aspects of human movement under realistic independent living conditions. In our Smart CondoTM study, twelve participants wore the commercial smart eyewear device JINS MEME, which collected electrooculography, accelerometer, and gyroscope data throughout their sessions. In this chapter, we describe our method for offline classification of the participants' activities. Also, this chapter defines activity recognition, summarizes the related work, background, and is structured as follows. Section 2.1 reviews similar researches about wearable sensors in combination with environment-based sensors for activity classification. Furthermore, the chapter sets out questions, protocol, data collection and analysis, and reviews findings for the study *Recognition of Daily Activities Using Environment Based Sensing*. Additionally, activities, sensor-data collection, attribute extraction, and proposed signal characteristic attributes are presented.

Older adults typically prefer to stay in their homes rather than enter a healthcare institution – a survey undertaken in the United States indicated that 30% of those over 65 years of age would “rather die” than enter a nursing home [9]. Placement in a care facility, particularly when it occurs against an individual's wishes, has been associated with depression, social isolation, and greater dependency on others for self-care tasks [9].

The rise in the population of older adults has led to an increased burden on healthcare systems,

as the health of elderly persons deteriorates with age. According to the American Geriatrics Society, complex needs are chronic conditions that frequently require services from different healthcare practitioners in multiple settings including frequent hospitalizations [15]. In the United States, the prevalence of older adults with multiple chronic conditions exceeds 60%. Older adults with complex needs are often limited in their ability to perform basic daily activities, due to physical, mental and psychosocial challenges requiring complex continuing care [16]. With continuous health-monitoring systems, the ability to recognize people's activities in their homes can enable automated assisted living systems, caregivers and clinicians to provide suitable adaptive care. This can enable older adults to live independently longer at home and reduce their reliance on caregivers while supporting caregivers in providing better care. With the advent of miniaturized sensing technology, which can be wearable, it is now possible to collect and store data on different aspects of human movement under realistic independent living conditions [10]. These technologies have the potential to provide a cost-efficient approach to enhance one's quality of life and help older adults live safely in their homes [16].

The recognition and classification of activities of daily-living can provide an important context for caregivers, clinicians and assisted living systems to plan suitable methods of providing care. Wearables have the potential to be used in automated activity profiling systems, which can produce a continuous record of activity patterns over extended periods of time. While the vast majority of activity classification systems have used inertial sensors such as accelerometers and gyroscopes [10], activity classification has also been performed using electrooculography (EOG) [17]. Inertial sensors are also often used in combination with ambient sensors placed in a smart home environment [18, 19, 20] or with other wearable sensors [21, 22, 23, 24, 25, 26] in order to increase accuracy, or recognize a greater diversity of activities.

JINS MEME is a commercial eyewear device that contains sensors – three EOG electrodes, an accelerometer, and a gyroscope – embedded in a traditional eyeglasses frame, and transmits data via Bluetooth Low Energy (BLE). The device can be outfitted with prescription lenses, and is intentionally designed to be cosmetically suitable and non-restrictive [27]. Since 92% of people 70 years of age and older already wear glasses [28], this device has a high likelihood of being adopted by older adult populations.

In the here presented Smart CondoTM study, twenty-six participants spent one two-hour session in the one-bedroom living environment, either alone or in pairs, and performed a scripted protocol of activities of daily living. Twelve of these participants wore the commercial smart eyewear device JINS MEME, which collected electrooculography, accelerometer, and gyroscope data throughout their sessions. These data from the JINS MEME was used to classify the activities of daily living. This study aims to develop a method of accurately classifying activities of daily-living that is

practical to set up and not restrictive to the user's movements or social interactions.

The study investigates the benefits and limitations of using the JINS MEME glasses for activity recognition, and identifies techniques and practices in signal processing and machine learning that are most suitable for activity classification using data from this device.

This work contributes to the research area of activity recognition, examining the function of EEG and kinematic motion data. The examination of the data collected using the sensors embedded within the JINS MEME for the classification of activities of daily-living that was carried out. The results showed that these data can be used to accurately classify both, motion-based and visual-based activities. Also, a method for the calculation of information about eye and head movements collected using the JINS MEME was developed, in addition to standard characteristics of EOG and kinematic motion signals, to be used as attributes in the classification process.

2.1 Related Work

The development of wearable physiological sensors has made it possible to collect and store information on different aspects of a person's movement and activity throughout the day [10]. Sensor devices have provided opportunities for researchers to conduct studies that are more realistic than studies conducted in highly controlled environments.

Researchers have investigated the use of wearable sensors in combination with environment-based sensors for activity classification. In contrast to single-sensor systems, using wearable inertial sensors to collect information about a person's behaviour, in combination with sensors embedded into a living environment to collect information about a person's location [18, 20] or their use of specific objects [19, 8], allows researchers to improve the detection of behavioural changes or recognize a greater number of subjects [20].

The effectiveness of wearable sensors in collecting data for activity classification has also been studied by using multiple wearable sensors in combination with each other. Past research has often used additional wearable sensors to increase the accuracy of inertial sensors in collecting data for classifying postures and motion-based activities [21, 22, 26], but studies have also combined sensors with the aim of classifying a more diverse variety of activities of daily-living [23, 24, 27]. While accelerometer data is often used to accurately classify motion-based activities, a key challenge remains in the classification of visual-based activities performed while stationary.

Although the vast majority of activity-classification systems have used inertial sensors such as accelerometers and gyroscopes [10], studies have also proposed the use of wearable devices with an infrared proximity sensor [29] and EOG [17] to address both the limitation of inertial sensors (to

detect motion-based activities) and the limitation of environment-based sensors (to detect location and basic activities). These sensors detect information on blinks to be used in the classification of visual-based activities like reading, writing, and watching videos. While the infrared proximity sensor allows the detection of blink rate, EOG allows additional information to be drawn from eye movements including the duration and amplitude of blinks [17].

Past research on subject and clinician preferences for wearable sensor systems suggests that sensor systems should be compact, embedded, and simple to operate and maintain [30]. Therefore, the use of sensors embedded within a single wireless device worn on the wrist [22, 23, 31] or on the head [25, 29] has been proposed with the goal of increasing the likelihood of user acceptance.

Clinicians and subjects also indicated that wearable sensor systems should not affect daily behaviour [30]. Due to the widespread usage of smartphones, studies have proposed the use of inertial sensors embedded within the smartphones for activity classification systems. It has been suggested that their long battery life and non-obstructive nature will also contribute to increased acceptance among potential users [32, 33]. Past research has achieved high accuracy in the classification of motion-based activities using smartphones placed in the user's hand [32], pocket [32, 33], and belt case [34]. A drawback of systems that use sensors within smartphones is that they are limited to situations where the smartphone remains in the same position relative to the user [35].

The present study proposes the use of JINS MEME, an eyewear device with inertial sensors and three dry electrodes that detect EOG. This combination of sensors collects information that may allow for the detection of a variety of motion-based and visual-based activities of daily-living. The sensors are embedded within a traditional glasses frame that can be outfitted with prescription lenses, thereby increasing the chance of user acceptance among elderly individuals who are likely to already require corrective eyewear. The use of three dry electrodes makes the system less restrictive and more convenient than conventional five-electrode EOG configurations, which typically require adhesion of electrodes to the skin. Since glasses are typically only worn in one manner, the device remains in a consistent position relative to the user for a distinct advantage over smartphone systems. Previous research has used JINS MEME eyewear to accurately classify visual-based tasks [36], drowsiness state during driving [37], and control mode according to the Hollnagel dynamic cognition model [38].

Activity-profiling systems are also dependent on classification algorithms to interpret data from wearable sensor data and identify different activities [10]. Past research has often performed activity classification using support vector machine (SVM) [17, 18, 23, 25, 30, 32, 34], decision tree [19, 20, 21, 22, 26, 31, 32, 34], K-nearest neighbour [20, 24, 32, 33, 34, 35], C4.5 [24, 29], Naive Bayes [20, 33] and Multi-layer Perceptron [20] algorithms.

In this study, the benefits and limitations to classify activities of daily-living using data collected from EOG and the inertial sensors of the JINS MEME device are discussed. Furthermore, this study identifies techniques and practices in signal processing and machine learning that are suitable for activity classification using data from the JINS MEME.

2.2 Protocol

Participants: Twenty-six participants were recruited to spend a one to two-hour session, either alone or in pairs (7 pairs), in the Smart CondoTM. The participants were asked to follow a scripted sequence of activities, i.e., an activity protocol. All participants provided informed consent approved by the Health Research Ethics Board at the University of Alberta. Only participant information required for this study was collected.

Equipment: Video footage was used to determine the ground truth of participant activities. Twelve of these participants wore the JINS MEME eyewear while performing the activities, and no participant-specific calibration or adjustments were made to the device. The data was stored in a local machine at the Smart CondoTM.

Activities: Within the goal of supporting older individuals to live independently longer, this study aims to classify a range of activities of daily-living (ADLs) associated with independent living. Similar to [23], we define two main types of ADLs, basic and instrumental. Basic ADLs (BADLs) are activities which are necessary for self-care, while instrumental ADLs (IADLs) are activities that require slightly more complex skills and are important for independent living. In this context, the Barthel Index [39] is a scale used to measure performance in BADLs and the ability to live independently. The Barthel Index uses ten variables to describe activities and mobility. Our activity protocol includes six of these variables: bathing, dressing, feeding, grooming, toilet use, and walking. Wheelchair transfer, bowel control, and bladder control were excluded because this study included healthy participants only. Stair ascension and descension were also excluded due to the one floor design of the Smart CondoTM. IADLs include cooking, exercise/stretching, low and high intensity housework, typing/writing and watching television (see Appendix A.1). In total, twelve ADLs are included in this study. A list of the ADLs and their descriptions are shown in Table 2.1. The activity protocol was designed to realistically simulate daily-living. Therefore, the tasks were performed for natural amounts of time rather than set, equal amounts of time.

Table 2.1: Activities of daily living included in this study.

Activity	Description
<i>BADLs (Basic Activities of Daily Living)</i>	
Bathing	Sitting down in the shower and pretending to bathe
Dressing	Donning and doffing clothing
Feeding	Eating food and taking medication
Grooming	Washing hands and using sink
Toilet Use	Sitting down on the toilet, retrieving toilet paper, flushing
Walking	Walking on level floor
<i>IADLs (Instrumental Activities of Daily Living)</i>	
Cooking	Preparing a meal using kitchen appliances
Exercise	Stretching
Housework (Low Intensity)	Setting the table and washing dishes
Housework (High Intensity)	Sweeping with a broom, loading the laundry machine and ironing
Typing/Writing	Using a tablet to type and play games
Watching TV	Sitting down and watching TV

2.3 Data Collection & Analysis

This section describes the activity recognition framework that was developed to synchronize the collected signals from the JINS MEME with the video received data from the Smart CondoTM, which is depicted in the following flowchart (see figure 2.1).

Figure 2.1 depicts the approach used for the study *Recognition of Daily Activities Using Environment Based Sensing*. The method contained four main steps to process the data collected to features. The first two steps were used to collect and label the data, and the remaining steps were used for feature extraction and data set construction. As a first step, the EOG and kinematic motion data collected from the JINS MEME, and the video data collected from the Smart CondoTM were inputted. Secondly, the data was labelled to define the activity mode. Then, MATLAB was used to synchronize and extract 214 features from the EOG, ACC, and GRYO data. Finally, after balancing the classes, the machine learning algorithms were evaluated using labelled activities.

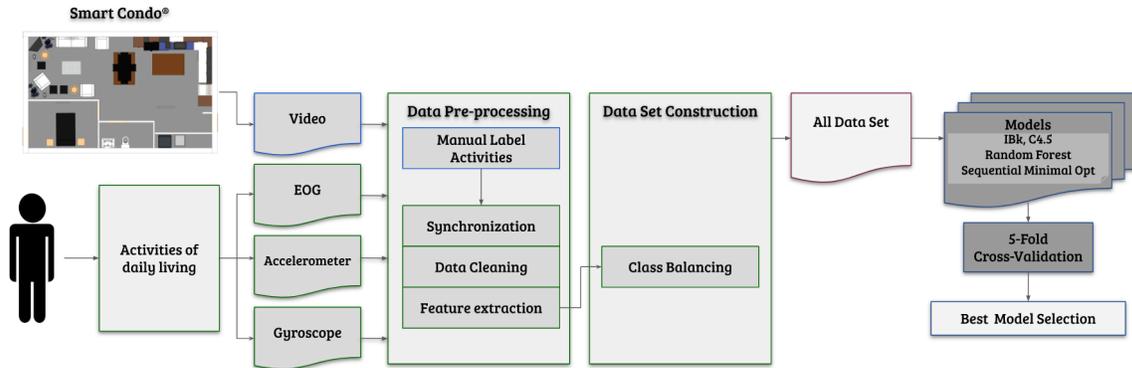


Figure 2.1: Illustration of the pipeline for activity recognition.

2.3.1 Sensor-Data Collection

The JINS MEME eyewear collects EOG and kinematic motion data at 100 Hz and transmits this data to a nearby computer wirelessly via BLE. Three dry electrodes housed within the bridge and nose pads of the glasses collect EOG signals in the horizontal and vertical dimensions. An accelerometer and a gyroscope, housed within one of the arms of the glasses, collect kinematic motion data. Figure 2.2 shows an image of the JINS MEME eyewear.



Figure 2.2: The JINS MEME eyewear and embedded sensors [2].

Because only three electrodes are used, rather than the more conventional use of five electrodes, the EOG signal is calculated in a bipolar method rather than a monopolar method [27]. This means that the signal collected represents the velocity of eye movements rather than eye position.

JINS MEME includes an accelerometer and gyroscope, but it does not include a magnetometer. Therefore, the angular position of the inertial sensor along the user's Euler angles cannot be determined without drift about the longitudinal axis. Instead, we determine the angular velocity of the inertial sensor along the user's Euler angles by combining the accelerometer and gyroscope data and correcting for the sensor orientation relative to the user with a subject-independent offset.

The Smart CondoTM includes multiple video cameras that record footage simultaneously. This

video footage was used to determine the ground truth and label the data.

2.3.2 Feature Extraction

The data collected from the JINS MEME glasses was emitted to a collecting computer in the main room of the Smart Condo™ with a timestamp for each data point. We used the timestamp to synchronize the data with the video collected from the Smart Condo™, since it was found that certain activities and locations in the Condo caused the device to temporarily lose connection with the computer receiving the data. This resulted in some periods of data outage.

The raw sensor data was analyzed to extract 214 signal attributes (see Appendix A.4, and A.5). To calculate these attributes over the time-dependent signal, the incoming data was sliced into windows; the data in each window was analyzed as a unit, and one set of attributes was computed corresponding to each analysis window.

Previous research has found that a window length of 5.6 seconds is most appropriate for activity recognition in living environments, as it is either optimal or near optimal for many attributes, and also allows for posture recognition [40]. Furthermore, for potential automated applications in real-time, a window length of 5.6 seconds is short enough to enable timely intervention, since a potentially risky activity would be recognized almost as soon as it occurs and the decision to intervene is triggered. The same research also conceded that a disadvantage of such a short analysis window is lower performance on activities with high motion variability, since these activities can be performed in different ways depending on the situation [40]. Furthermore, previous research has found that overlapping the windows is important in order to handle transitions more accurately [41].

In our study, we adopted an analysis window of 5.6 seconds and a sliding increment of 1 second for an overlap of 4.6 seconds. As discussed, data outages occurred when a subject was further away for the receiving computer. These outages caused some analysis windows to contain less than the anticipated 560 data points (5.6 seconds of data at 100 Hz). Windows with at least 60% of this number (336 data points) underwent an initial **interpolation** step to bring their number of data points to the expected 560 before attributes were extracted from the data within those windows.

Signal Characteristic Attributes

The attributes calculated in this study include characteristics of the raw signal, emitted by the JINS MEME device, identified by past research on activity classification using accelerometers [40]. 214 Attributes were calculated in the time and frequency domains: previous research using accelerometer data has identified a relationship between the mean of the accelerometer signal in the time domain and the subject's movement intensity [42], while activities with a similar energy intensity can be

identified in the frequency domain by the period of the accelerometer signal [43] (see Appendix A.5). While [40] only extracted this set attributes from the three axes of accelerometer data, we extracted these attributes from (a) accelerometer data, (b) three Euler angles of the angular-velocity data, calculated as discussed above, and (c) three dimensions of EOG data (vertical, horizontal, reference).

Low-pass filters (20 Hz) and band-pass filters (5-20 Hz) were used in the calculation of signal characteristic attributes. **Low-pass filters** were applied to eliminate signal noise generated by dynamic human motion and preserve the information generated by static human motion or posture information [40]. A **band-pass filter** was applied to reinforce the effect of the low-pass filter and eliminate the static signal component that contains posture information about the orientation of the sensor with respect to the ground [40].

Higher Order Attributes

In addition to the attributes above, we also analyzed the JINS MEME signals to extract higher-order attributes related to eye movements and head movements about Euler angles. A moving average over a duration of 0.05 seconds was applied to the EOG and head angular velocity data to remove signal noise and maintain the overall form of physical movements. Previous research in the identification of eye movements from EOG data used the derivative of the EOG signal to classify the presence of an eye movement as a blink or a saccade [44]. Because the raw signal from the JINS MEME glasses already represents eye movement velocity rather than position as in conventional EOG, it may be used in the same manner as the first derivative in [44].

For the initial detection of eye movement, we used the first derivative (eye acceleration) of the EOG signal from JINS MEME rather than the raw signal because the difference between steady periods of no eye movement and peaks during eye movement was more pronounced. In the first derivative of the EOG signal, all eye movements are characterized by a local maximum and local minimum. The direction of an eye movement in either the vertical or horizontal dimension can be determined by whether the local maximum leads (positive direction) or lags (negative direction) the local minimum.

Since blinks and saccades occur in the same vertical dimension of the EOG signal, there is a need to distinguish between these movements. All eye movements in the horizontal dimension are assumed to be saccades because blinks do not have a significant effect on the signal in this dimension. Blinks typically cause the raw signal to develop local maxima followed immediately by pronounced local minima as shown in Figure 2.3. Upward saccades cause the raw EOG signal to develop local maxima followed by a gradual decrease to zero, while downward saccades cause the

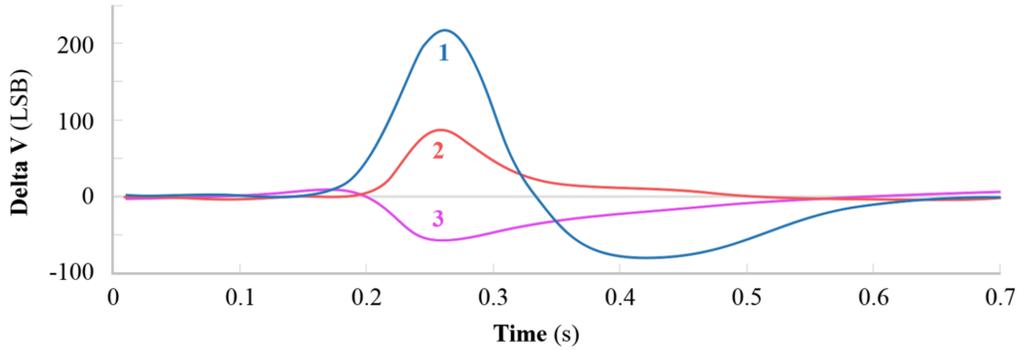


Figure 2.3: The effect of a blink (1 - blue), upward saccade (2 - red) and downward saccade (3 - magenta) on EOG signal in the vertical dimension as detected by JINS MEME.

signal to develop local minima followed by a gradual increase to zero as shown in Figure 2.3.

While downward saccades can be identified from the fact that the local minima occur before the local maxima in the derivative of the signal, there remains a need to distinguish between blinks and upward saccades. Toivanen *et al.* [44] defined a variable, D_v , to distinguish between blinks and saccades based on the differences in fundamental form of the EOG signal between these eye movements. The calculation for this variable is shown in equation (2.1).

$$D_v = \max(\text{raw}_{EOG}) - \min(\text{raw}_{EOG}) - |\max(\text{raw}_{EOG}) + \min(\text{raw}_{EOG})|. \quad (2.1)$$

raw_{EOG} has to be replaced by the corresponding *Horizontal* or *Vertical* raw EOG signal. Using equation (2.1) on eye movements detected in the EOG signal, we expect that D_v for blinks would be large due to the somewhat symmetrical form of the signal. D_v for upward saccades would be close to zero as the minimum is much less pronounced, i.e., the value of \min in equation (2.1) would be a value close to zero. We determined a threshold of D_v for distinguishing between blinks and upward saccades based on trial tests with the JINS MEME eyewear. Baseline drift and static noise during eye movements was removed by subtracting an offset value from local maxima and minima. The offset value was determined by calculating the mean signal value during the time between 0.07 and 0.02 seconds before the start of the detected movement.

We used a similar approach to detect physical head movements about Euler axes using data calculated from accelerometer and gyroscope signals, as discussed above. The head velocity along each Euler angle was used to identify head movements along that angle and determine the directions of those movements.

The average amplitude, duration, and number of eye movements were calculated for left and

right saccades, blinks, and upward and downward saccades. The same properties were calculated for head movements in both directions along each Euler angle. These properties of eye and head movements were used as higher-order attributes for each analysis window. Baseline drift and static noise were eliminated in the calculation of amplitude for eye and head movements as discussed above. Peak-to-peak amplitude was used as the amplitude for blinks.

The JINS MEME eyewear uses dry electrodes that merely contact a user's skin, rather than conventional gel electrodes that adhere to the skin surface. Because of this, the EOG signal tends to exhibit large peaks during facial movements or adjustments of the eyewear that interfere with this contact. Such peaks are significantly larger than those observed during regular eye movements. We calculated the percentage of time covered by these peaks (when they were above a certain threshold) as a measure of contact interference and included this as an attribute for each analysis window. These peaks were ignored in the detection of eye movements.

To evaluate the usefulness of information extracted above, we used an existing technique to balance the instances of each activity, then compared four algorithms for classifying data collected through the JINS MEME eyewear through 5 fold cross-validation.

2.3.3 Class Balancing

Our activity protocol simulates a typical independent living scenario, with some activities (such as typing/writing and exercise/stretching) performed for longer duration than others (such as grooming and toilet use). This difference implies that the classification process must address the class-imbalance problem. This is a phenomenon where uneven instances of classes in training data cause classification results to be less favourable for classes with a smaller number of instances (minority classes). Imbalanced classes can negatively affect the outcomes of activity classification, since many machine-learning algorithms assume a balanced distribution of classes. Due to this assumption, they roughly equate misclassification costs for each class and perform poorly in predicting the minority classes [45, 46]. The implication for our study is that the accuracy of the classifier will be low for activities that occur during a smaller number of analysis windows when included in a data set with other activities that have a far greater number of windows. To mitigate this risk, activities with fewer than 50 analysis windows, i.e., grooming and toilet use, were removed from our evaluation.

For the ten remaining activities, a technique called SMOTE (Synthetic Minority Over-sampling Technique) was used to counteract the class-imbalance problem. Developed by Chalwa et al., [47] SMOTE uses an algorithm to construct new data based on existing data in order to increase the amount of data in underrepresented minority classes while avoiding overfitting. Similarly to [48], the maximum number of samples generated using SMOTE per class was limited to 1,000 as another

caution against overfitting the data.

2.3.4 Classification

In this study, classification was performed using four classifiers that were considered likely to yield accurate results based on the success of past research and the relevance of their strengths to the context of the present study. The parameters can be seen in appendix A.2.

IBk is a K-nearest neighbor classifier that has been shown to be accurate in activity recognition for ubiquitous sensor environments such as the Smart CondoTM [49]. K-nearest neighbor classifiers have been proven to yield classification accuracies greater than 90% for activity classification [24, 31, 32, 33, 34].

C4.5 is a nonparametric classifier that is efficient in dealing with large, complicated datasets without imposing a complicated parametric structure [50]. This rule-learning scheme induces an initial rule set and then refines it through global optimization where individual rules are discarded. PART is an algorithm that infers rules by repeatedly generating partial C4.5 decision trees and avoids the post-processing stage that results in slow performance under the C4.5 method [51]. C4.5 classifiers have been shown to yield classification accuracies over 80% for activity classification [24, 29].

Random Forest (RF) is a decision tree classifier that requires little data pre-processing without normalization of attributes. This algorithm does not require attribute selection and is resistant to overfitting [21]. Past research using JINS MEME for real-time detection of drowsiness during driving used Random Forest for classification with 80% accuracy [36].

Sequential Minimal Optimization (SMO) is a simple algorithm designed to quickly solve the SVM quadratic programming problem by decomposing it into smaller subproblems [52]. Past research involving the use of wearable sensors for activity classification has suggested that SVM is an accurate classifier [17, 18, 23, 31, 33].

Classification of the activities was performed in Weka using the four classifiers (IBk, PART, Random Forest, and SMO). While 10-fold cross-validation is a standard way of measuring the error rate of a learning scheme on a particular dataset [53], research on the effect of reduction in cross-validation intervals has indicated that decreasing the number of folds to 5 reduces computation time by half with no loss of power [54]. This suggests that performing cross-validation with 5 folds, instead of 10, may allow the classification to be applied to larger data sets. Due to the large amount of data involved in this study, 5-fold cross-validation was performed for each participant.

2.4 Findings

Of the four algorithms tested, the most accurate classification results were obtained using the Random Forest algorithm. Table 2.2 shows the classification accuracy and precision by activity using Random Forest, with and without applying SMOTE.

After using SMOTE to balance the classes, the classification accuracy improved for all of the activities except typing/writing. For most participants, this activity was performed for the longest amount of time and therefore had more analysis windows. The decrease in classification accuracy of typing/writing could be an indication of the elimination of bias in favour of the majority class. This could also be a consequence of slight overfitting of other classes after applying SMOTE.

Bathing and dressing, activities that involve a high amount of movement, have the highest classification accuracy. Typing/writing and exercise/stretching, activities that involve considerably less movement, show the lowest classification accuracy. These results may indicate that the proposed method is more effective at detecting motion-based activities than visual-based activities.

Table 2.2: Average Accuracy, Recall, and Precision from all participants by activity.

Activity	Imbalanced (%)			With SMOTE (%)		
	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Bathing	80.20	73.86	87.75	98.97	100.00	97.97
Dressing	73.02	88.02	96.44	98.99	99.49	98.50
Feeding	81.78	81.76	81.81	94.08	95.08	93.09
Walking	59.40	39.19	72.01	93.01	93.25	92.79
Cooking	73.02	68.11	78.71	93.03	92.57	93.49
Exercise/Stretching	84.35	84.35	84.36	90.64	91.17	90.11
Housework (Low Intensity)	74.80	71.85	78.00	93.61	95.12	92.15
Housework (High Intensity)	82.15	84.04	80.35	95.42	96.12	94.73
Typing/Writing	89.27	93.20	85.67	90.43	87.15	93.97
Watching TV	59.40	43.63	93.04	92.48	95.80	92.48
Overall	76.37	72.09	83.83	94.23	94.58	93.93

Table 2.3 summarizes the results of our study and compares them to the results reported in relevant previous research. It is important to note that this table is meant primarily as a summary overview of the field and not as a valid comparative evaluation of the reported methods since these methods have not been applied to a single shared data set and experimental scenario.

Based on this table, our method seems superior in accuracy with respect to what has been reported in the literature [21, 25, 36, 23, 17, 35, 18, 24, 22]. This further indicates that the JINS MEME eyewear and the classifier Random Forest can be used to reasonably classify activities in independent living environments. Additionally, the non-obstructive nature and convenient setup of the proposed method may increase user acceptance.

As it can be seen in table 2.3, our results are slightly lower compared to the publications of Ayu *et al.* [23], Lau *et al.* [33], Prekopcsák *et al.* [34], and Hong *et al.* [19]. This is due to the number of sensors, participants, and activities used in this work compared to the literature. Ayu *et al.* and Lau *et al.* [33] predicted five, and Prekopcsák *et al.* [34] predicted ten activities, using only one participant. Overall, they used a significantly smaller sample size than chosen in these studies, which leads to higher accuracies but less conclusive results. Hong *et al.* [19] achieved a 1.66% higher accuracy than our system, which is due to the use of two additional accelerometers, and RFID sensors. Nevertheless, it has to be noted that our approach has the advantage that JINS MEME is a non-restrictive eyewear device, requires fewer sensors, and does not use calibration compared to the RFID sensors.

Table 2.3: Accuracy related work.

Year: Author(s)	Wearable Sensor Location: Sensor Type(s)	# of Subjects	# of Activities	Types of Activities M = Motionbased, V = Visualbased	Accuracy
2017: Pavey <i>et al.</i> [21]	Wrist: Accelerometer Thigh: Accelerometer	21	4	M: Sedentary, Stationary, Walk, Run	90.30 %
2014: Zhan <i>et al.</i> [25]	Head: Accelerometer, Video Camera	2	12	M: Walking, Ascending Stairs, Descending Stairs, Drinking, Standing Up, Sitting Down, Sitting Still, Washing Hands, Switching Water Tap V: Reading, Watching TV, Writing	90.38 %
2014: Ishimaru <i>et al.</i> [29]	Head: Accelerometer, Infrared, Proximity Sensor	8	5	M: Sawing V: Watching Video, Reading, Solving Puzzle, Talking	82.00 %
2013: Chernbumroong <i>et al.</i> [32]	Wrist: Accelerometer Gyroscope, Magnetometer, Altimeter, Temperature	12	9	M: Brush Teeth, Dressing, Feeding, Ironing, Sleeping, Sweeping, Walking, Washing Dishes V: Watching TV	90.23 %
2012: Ayu <i>et al.</i> [23]	Smartphone: Accelerometer	1	5	M: Jogging, Jumping, Sitting, Standing, Walking	98.00 %

2011: Bulling <i>et al.</i> [17]	Body & Head: Electrooculography	8	6	V: Copying Text, Reading, Writing, Watching Video, Browsing Internet, Null	70.50 %
2010: Das <i>et al.</i> [35]	Smartphone: Accelerometer	1	7	M: Idle, Walking, Running, Jumping, Ascending Stairs, Descending Stairs, Phone Detached	93.00 %
2010: Fleury <i>et al.</i> [18]	W (Arm): Accelerometer, Magnetometer, E: Infrared Presence Sensors, Door Contacts, Temperature, Hygrometry sensor, Microphones	1	7	M: Sleeping, Cooking/ Eating, Dressing, Resting, Hygiene, Bowel Movement, Communication	86.20 %
2010: Lau <i>et al.</i> [33]	Smartphone: Accelerometer	1	5	M: Walking, Standing Sitting, Ascending Stairs, Descending Stairs	99.27 %
2010: Hong <i>et al.</i> [19]	W (Arm): Accelerometer W (Waist): Accelerometer W (Leg): Accelerometer E: RFID Sensors	15	10	M: None, Cutting, Brushing Teeth, Taking Picture Shaking Hands, Wiping with Cloth, Putting on an Umbrella, Jumping Rope, Vacuuming Pushing Shopping Cart, Applying Skin Conditioner	94.69 %
2009: Maguire & Frisby [24]	Thigh: Accelerometer, Heart Monitor	6	8	M: Standing, Brushing Teeth, Ascending Stairs, Descending Stairs, Walking, Running, Vacuuming, Sit-ups	90.07 %
2009: Prekopcsák <i>et al.</i> [34]	Smartphone: Accelerometer	1	10	M: Walking, Running, Working (Sitting), Cooking, Vacuuming, Stairs, Elevator, Riding Bus, Lying Down. V: Watching TV	95.10 %
2006: Maurer <i>et al.</i> [22]	W (Wrist): Accelerometer Temperature Sensor, Light Sensor, Microphone	1	6	M: Walking, Running, Standing, Sitting, Ascending Stairs, Descending Stairs	92.80 %

2.5 Summary

This study examines the potential of an off-the-shelf eyewear device, JINS MEME, as a means of sensing and accurately recognizing activities of daily-living and emotion recognition. Past research

into subject and clinician preferences for wearable sensor systems suggested that they should be compact, simple to operate and maintain, and should not affect daily behaviour. JINS MEME's convenient packaging of multiple sensors – three EOG electrodes, an accelerometer and a gyroscope – within the frame of traditional eyeglasses, which the majority of older adults already wear, can potentially offer a practical method of activity recognition to help older adults living independently longer.

We conducted our study on data collected from twelve healthy adults who performed activities of daily-living in the Smart Condo™ was conducted, an independent living suite, while wearing JINS MEME. Video footage was used to determine when participants performed which activities. A set of attributes – including signal characteristics and higher order properties of blinks, saccades and head movements – was extracted from the collected data, calculated over 5.6-second windows. The resulting data set was further processed to eliminate infrequent classes (i.e., activities) and to balance the number of data points of the remaining imbalanced classes (with SMOTE).

Several machine-learning algorithms were used to classify each participant's activities offline using cross-validation; the Random Forest classifier resulted in the highest classification accuracy. The results indicate the removal of bias in favour of windows with many windows of processed data, but may suggest a slight reversal of this bias in favour of minority classes (activities with many windows of data generated using SMOTE). Data with a more balanced distribution among activities is needed for comparison in order to determine if overfitting occurred or if bias persists.

Overall, our method is capable of classifying the data collected with 94.23% accuracy for ten activities of daily-living. These results are comparable to the best results achieved in past research using wearable sensors for activity classification, with the advantage that JINS MEME may have a higher likelihood of user acceptance.

Admittedly, there are some challenges to using the JINS MEME device. The Bluetooth signal used to transmit data from the device to a computer was much less effective when participants were not in the same room as the computer. This resulted in a loss of data from activities performed in other rooms. A stronger Bluetooth signal would be needed to make this device effective for use in a variety of living environments. As well, the inclusion of an accelerometer and gyroscope without the inclusion of a magnetometer prevents the accurate determination of the angular position of the user's head without drift. The inclusion of a magnetometer would increase the effectiveness of the other inertial sensors.

Measuring Arousal & Valence during Gameplay

Chapter 3 presents the emotion recognition framework, data-stream and synchronization, feature extraction, and training, followed by questions/hypotheses, protocol, data collection and analysis, and findings for the study *Measuring engagement, emotional states and cognitive function*. Finally, related work for recognizing emotional states during gameplay by using physiological signals from wearable devices is presented. For individuals with cognitive impairment, the cognitive demand of certain activities of daily living can impose barriers for independent living. Performing activities with increased cognitive demand can also increase postural sway and risk of falling [55, 56]. Improving the cognitive function of individuals with cognitive impairment can enhance their ability to perform activities of daily-living, and, consequently, their ability to live independently [57].

Recent studies have shown that training using video games can improve cognitive function [58, 59]. The premise for this research lies on the idea that the human brain retains considerable plasticity with increasing age [60] that allows it to be trained by playing serious games. These brain training games are expected to improve cognitive functions such as executive function, memory, attention, and processing speed — not only during the performance of the serious games but also during the performance of activities of daily-living.

The proposed study uses Whack-a-Mole [14], a serious game in which a participant must hit all moles shortly after they appear and avoid hitting bunnies. This game was developed for the assessment and intervention of visuomotor, cognitive and mental health conditions [14]. The game's simple rules and touchscreen interface make it appropriate for individuals with cognitive impairment. During gameplay, successive levels require a progressively higher cognitive skill. The software embedded in the game captures gameplay data in an anonymized file. The Whack-a-Mole game has already been pilot-tested with ten older adults with moderate dementia who attended a day program at the Western Ottawa Community Resource Centre at Carleton University [13].

Emotional states are challenging to assess in individuals with cognitive impairment through self-reported measurement tools such as questionnaires since the questions are generally too complex to be understood. The study measures electroencephalography (EEG), electrooculography (EOG) and kinematic motion data using wearable sensors while the participants are playing the game Whack-a-Mole. This data will then be used to assess emotional states during gameplay.

This study aims to answer the following research questions:

- **Q1:** *How accurately can the player's self-reported valence and arousal be predicted based on EEG, EOG, and kinematic motion data?*
- **Q2:** *How does the player's emotional state changes as the game difficulty changes?*
- **Q3:** *How accurately can the players' emotional states be predicted using pre-trained machine learning models with the use of IAPS images during gameplay when the difficulty of the game changes?*

The study involved healthy individuals only. Participants filled out the positive and negative affect schedule (PANAS) questionnaire as self-reported measures of emotional states. These and cognitive difficulty in participants, which was identified by gameplay performance, were used to characterize the features of the biosignals collected during gameplay.

This work contributes to the research area of emotion recognition, examining the function of the EEG, EOG, and kinematic motion as data sources, which can be used to classify emotional states accurately. Further, we developed a method for the extraction of information from eye and head movements using data collected from JINS MEME, a smart eyewear device, which were used as attributes in the classification process. Based on our emotion recognition framework, we created promising methods to predict emotion states during gameplay.

3.1 Related Work

Researchers have been investigating the use of physiological signals for emotional states. In our work, we adopt the two-dimensional valence-arousal model of Russel *et al.* [61], which represents emotional states as points in a two-dimensional space defined by the arousal and valence dimensions of the Circumplex model.

3.1.1 Emotional State, Arousal & Valence

Arousal is the physiological and psychological state of being awake, active, alert, attentive, or excited. From a perspective of dimensional emotions, arousal can be differentiated into emotional states of low arousal (e.g., quiet) and high arousal (e.g., surprised, excited) [13]. Valence represents the intrinsic attractiveness of an event. From a perspective of dimensional emotions, valence differentiates emotional states of pleasure (e.g., happy) with those of displeasure (e.g., sad) [13].

In this study, we considered two different approaches for the classification of human emotions, from Russel and Watson specifically [62, 4]. Russell proposed the Circumplex model, which suggests that emotions are distributed in a circular space with dimensions of valence and arousal. Valence ranges from low (unpleasant/defensive) to high (pleasant/appetitive), and arousal ranges from low (calm, sleepy) to high (excited, tense) [62].

Watson *et al.* proposed the positive activation – negative activation (PANA) model, which is commonly understood as a 45-degree rotation of the Circumplex model. The relationship between the two models is shown in figure 3.1, with neutral arousal and valence at the center of the circular space. Table 3.2 lists the emotional states associated with different positions in this circular space [63].

Positive and negative affect are the emotional components of subjective well-being [64]. The positive and negative affect of participants in this study was measured through the positive and negative affect schedule (PANAS). PANAS asks the respondent about their emotional state in terms of ten positive and ten negative emotions using a five-point Likert scale, which will be discussed in more detail in Section 3.3.4. The sum of the ratings of positive terms and negative terms indicated the participant's level of positive affect score and negative affect score, respectively [65].

3.1.2 Recognizing Emotion from Biosignals through Wearables

Recent studies use wearable sensors to assess the emotional state, which requires external stimuli — such as pictures, sounds, or videos — to induce emotions in participants [66]. Wearable technology

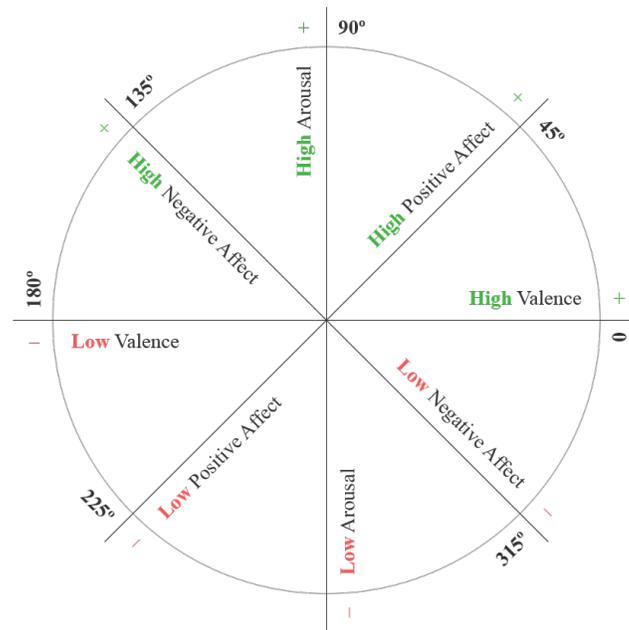


Figure 3.1: Two approaches to classifying human emotions. Circumplex model and Positive Activation - Negative Activation (PANA) model.

has the potential to be used in automated activity profiling systems, which can produce a continuous record of emotion recognition over extended periods. Emotion recognition has been performed using multiple physiological signals such as electroencephalography (EEG), electrooculography (EOG), and inertial sensors such as accelerometers and gyroscopes. The EEG signals are the primary sources of emotions in our body. According to Liu *et al.*, EEG features can be used for positive emotion recognition, but are less efficient for the detection of negative emotions [67]. Hence, to improve the accuracy of the system, it is required to integrate other signals such as EOG and kinematic motion. In negative emotion recognition, EOG features show high accuracy. Following Liu *et al.* using the properties of both, EEG and EOG features, improves the efficiency of emotion recognition [67]. Additionally, gyroscopes are known to enhance valence and head pitch to enhance arousal [68].

Physiological signals can be collected continually and precisely by using wearable sensors. The effectiveness of wearable sensors in collecting data for emotion recognition has also been studied by using multiple wearable sensors in combination with each other. In 2015, Cruz *et al.* [12] developed an automated recognition of facial expression based on EOG signals. In future research, this method can be used for the prediction of emotional states. This research used a traditional four channel EOG electrode to predict six movements of the eyes (namely, up, down, right, left, blink, and frown). The recognition algorithm extracted time and frequency domain features from EOG signals, which

were then classified by a multiclass linear discriminant analysis (LDA) classifier.

In 2017, Wei *et al.* [66] developed a real-time emotion detection system based on EEG signal measurements. An emotion detection headband coupled with printed signal acquisition electrodes, and an open source signal processing software (OpenViBE) were used. The binomial classification of positive and negative emotions was performed using naive bayes (NB) and support vector machine. With the use of linear discriminant analysis (LDA) as a classification algorithm, the highest achieved subject-dependent accuracy was 86.83%, and the highest subject-independent accuracy was 64.73%.

Wiem *et al.* [69] reported in 2017 a classification of emotional statements using an arousal-valence evaluation for peripheral physiological signals, ECG, respiration volume, skin temperature, and GSR specifically. This paper aimed to identify the human emotional states arousal and valence. They defined emotions using three different models: the first and second model utilizing two (in the first model) or three (in the second model) classes using 1-9 discrete self-rating scales, and the third model using 9 emotional keywords to establish the three defined areas in arousal-valence dimensions. Then, 169 features were extracted from the MAHNOB-HCI multimodal tagging database, and the emotional statements were classified by using SVM. The obtained results, 64.23% and 68.75% accuracy for arousal and valence respectively, suggested that SVM could be a promising classifier for emotion recognition.

A recent study from Liu *et al.* [70] in 2018 used a simple wearable system to acquire a single channel EEG signal, a respiration (RESP) signal, an ECG signal, and the body postures (head posture was tracked by a digital gyroscope and accelerometer) to explore the relation between these signals in human emotions as neutral and non-neutral. The study was performed with six participants. Videos provided by the Shanghai Jiaotong University SEED laboratory were used to stimulate the different emotions (positive, neutral, and negative) of each subject [70]. A main difference to our study is, that Liu *et al.* did not define the emotional state based on arousal and valence. They defined emotion as either neutral or non-neutral; non-neutral emotions were further classified positive or negative. To predict emotion, firstly data processing was applied to remove the baseline drift of the signals and a bandpass filter was applied to suppress the high-frequency interference. Then, power spectral density was applied to extract the most important features from the EEG data, standard deviation, average power, mean of the absolute values, and blink frequency (an EOG feature collected as an artifact from the EEG signal). On the other hand, three features from ECG and six from RESP were extracted. Additionally, for head posture stability one feature was extracted. In order to classify emotion, a linear super vector machine (SVM) was used to achieve approximately 87% accuracy.

In 2018, Wang *et al.* [71] proposed an emotion recognition method combining EOG and eye

movement videos by using eight healthy subjects. Firstly, the time-frequency eye movement features were extracted by applying the short-time fourier transform (STFT) to raw multi-channel EOG signals. Then, three emotional states (positive, neutral, and negative) were classified by using SVM with a polynomial kernel function. The average accuracy of the obtained results, around 88% to 89%, showed that eye movement information could adequately reflect the emotional states.

Wang *et al.* [72] experimented in 2018 on the Amigos database for binary classification for valence and arousal. The Amigos database has a total of 40 subjects and 16 short-length emotional videos, where each data has 14 channels for EEG, two channels for electrocardiography (ECG) and one channel of galvanic skin response (GSR) collected from wearable sensors. Wang *et al.* proposed an emotion recognition framework based on physiological signals. This framework has three main blocks. Firstly, a pre-processing block used low- and high-pass filters to clean physiological signals. Secondly, feature extraction methods were implemented per each biosignal. For the GSR signal 32 features, for the ECG signal 77 features, and for the EEG signal 105 features were extracted, with a total of 214 features. The last step proposed a block with entropy domain features and a XGboost classifier [72]. Overall, this method achieved 80% (valence) and 68% (arousal) accuracies of prediction for two affect dimensions. We used this paper as a groundwork for pre-processing EEG signals and extraction features from them. As a difference to previous studies, the present study proposes the use of JINS MEME for the detection of EOG, 281 extracted features (from EEG, EOG, kinematic motion data), feature selection, and classifier models (Random Forest and Ensemble Classifier).

On the other hand, another recent study showed the use of accelerometer data collected from wearable sensors, such as a SmartWatch, to detect emotion recognition. Quiroz *et al.* [73] performed their investigations as a mixed-design study: within-subjects (emotions happy, sad, and neutral) and between-subjects (stimuli: audio-visual and audio) by using 15 healthy adults. In their procedure, they divided the accelerometer data using sliding window, followed by feature extraction from each window. Then, the majority class was predicted using the classifiers Random Forest (RF), Logistic Regression (LG) and Baseline (BL). The highest achieved median accuracy of these studies was 78% for the binary classification of happiness *vs* sadness. The results confirmed that changes in emotional states and behavioral responses can be detected with a SmartWatch. Hence, accelerometer data can be used for emotion recognition.

Furthermore, the accelerometer data can be used to measure head movements. A recent study [68] investigating the correlation between head movements and self-report measures showed that individuals who displayed more side-to-side head movements gave higher ratings of pleasure. This experiment used a library of immersive virtual reality (VR) clips that were created by paralleling the design used in IAPS. In this study, the participants rated clips on valence and arousal dimensions.

The results showed a significant positive correlation between the standard deviation of head pitch and arousal, suggesting that people who tend to tilt their head upwards while watching immersive VR clips reported being more excited. Furthermore, a positive relationship between the standard deviation of yaw and valence was found.

3.1.3 Emotion During Gameplay

In the context of computer games, engagement has been defined as task involvement and describes the player's subjective perception of a game's reality and their degree of involvement on the task [74].

Affective gaming, the detection of emotion during gameplay, is of great research interest and exploits human emotion for the enhancement of the player's experience during gameplay. Yang *et al.* [75] showed in 2018, that the use of physiological signals can enhance the understanding of emotional states and that they can be applied to affective gaming. In their study, 58 participants with different skill levels played a football simulation game (FIFA 2016) to stimulate emotions. To identify the biosignal response during gameplay, they recorded and created a multimodal database (DAG) that contains biosignals (electrocardiography (ECG), electrodermal activity (EDA), respiration, electromyography (EMG), temperature), facial and screaming recordings, accelerometer signals, and the player's self-reported event-related emotion assessment. To recognize emotion detection during gameplay Yang *et al.* presented a feature extraction from physiological signal segments. Then, each feature per participant was normalized by using standard and min-max normalization. Additionally, a feature selection was applied. To classify the emotions, they used a linear SVM that achieved the best average accuracy with 50.4% for arousal and 50.7% for valence. Overall, they addressed the common challenges for physiological-based affective model such as signal segmentation, feature normalization, and relevant features, but their classifier has a lower performance. This paper shows that the most relevant features for emotion detection derive from EMG and acceleration signals.

Alhargan *et al.* [76] established, that playing the same game with different difficulty levels makes the player feel different emotions. The paper aimed to use eye tracking to recognize emotions (arousal and valence) during the gameplay. Fourteen healthy students played the game Speedboat, where each participant played five affective game levels and completed a questionnaire after each attempt. The methodology used in this paper included pre-processing, feature extraction, feature normalization, and classification. The pre-processing was used to eliminate the noise from the pupillary response, to isolate pupil light reflexes (PLRs) and to remove the artifacts before extracting affective features. Besides, standard features in the time-frequency domain, e.g. mean, standard deviation, mean deviation, spectral entropy, and the power spectral density (PSD), were extracted

from the pupillary response. Max-min normalization was applied to all features for their scaling between 0 and 1. Then, a classification was applied using SVM to recognize up to 76.0% and 61.4% on arousal and valence, respectively. These results show that the use of eye and pupillary responses are highly promising for emotion recognition, but the accuracy for valence is lower compared to previous research. This is due to the fact that they are not using features from blink, saccade and gaze distance which have been shown to associate with valence.

According to Huynh *et al.* [77], the decision to keep playing a game is based on how they feel in the first few plays. In order to enhance the game experience, game designers need to understand emotions and to use them as one of the design factors for triggering positive game experiences. Huynh *et al.* [77] developed a system that evaluates the emotional experience of gamers based on physiological changes. They used 22 participants to explore the relation between the level of difficulty and the player's emotion (excitement and happiness), and the association between emotions and the physiological signals. They demonstrated the use of wearable devices for the detection of physiological signals (electroencephalography (EEG), photoplethysmogram (PPG), and galvanic skin response (GSR)) to identify emotions during gameplay using the game Tank 1990 HD. The emotion recognition in this research had three main steps (pre-processing, feature extraction, and emotion classifier); firstly, a signal processing was applied by computing the beat-to-beat interval series and the heart rate from PPG signals. For the GSR signal, a low-pass filter with a cutoff frequency of 0.4 Hz was applied, and for the EEG a band-pass filter (1 Hz -30 Hz) was applied to eliminate linear trends in signal recording and to cut off high-frequency noise. Secondly, feature extraction was applied to calculate features such as heart rate variability, geometric features, standard deviation, mean value, the fractal dimension of each EEG channel. Last, they ran a correlation-based heuristic search algorithm to select the best features and used the Random Forest and J48 classifier algorithms to create the classification models (one for each emotion). The results show the high potential of using biosignal such as EEG, Random Forest, and wearable devices to predict game experience with classification accuracies of 77.38% and 73.21%, respectively, for excitement happiness.

3.2 Research Questions

The study helps to build an understanding of the patterns of engagement, emotional states and cognitive difficulty of healthy individuals while playing computer games. Therefore, we followed the three main hypothetical questions :

- **Q1:** *How accurately can the player's self-reported valence and arousal be predicted based on EEG, EOG, and kinematic motion data?*

- **Q2:** *How does the player's emotional state changes as the game difficulty changes?*
- **Q3:** *How accurately can the players' emotional states be predicted using pre-trained machine learning models with the use of IAPS images during gameplay when the difficulty of the game changes?*

3.3 Protocol

Participants: 21 healthy participants (11 Females and 10 Males) over the age of 18 were recruited from the University of Alberta. All participants provided informed consent approved by the Health Research Ethics Board at the University of Alberta

Equipment: Two devices were used to collect biosignals. Both devices transmitted data to a nearby computer via Bluetooth. The OpenBCI headset collected EEG and motion data at 125 Hz [3]. 16 electrodes contained within the headset collect data in accordance with the International 10-20 system (Fp1, Fp2, C3, C4, T5, T6, O1, O2, F7, F8, F3, F4, T3, T4, P3, P4), while an accelerometer embedded on the device's micro-controller board data collected about kinematic motion as it can be seen in figure 3.2.

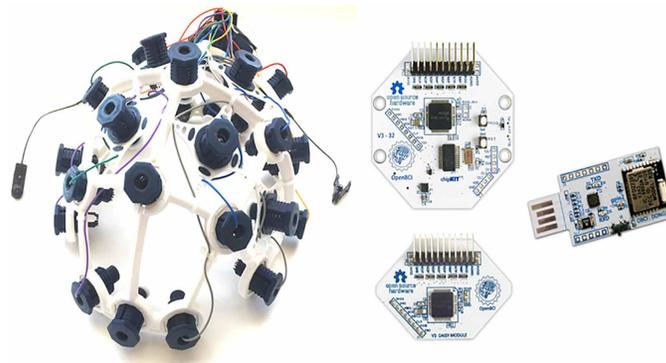


Figure 3.2: Open-source brain-computer interface (OpenBCI) is an affordable bio-sensing system, this device can collect electrical activity of human body such as EEG, EMG, EOG and ECG.[3]

The JINS MEME eyewear collected EOG and kinematic motion data at 100 Hz. Three dry electrodes housed within the bridge and nose pads of the glasses collected EOG signals in the horizontal and vertical dimensions. An accelerometer and a gyroscope, housed within one of the arms of the eyewear, collected kinematic motion data.

3.3.1 Study Procedure

Our study involved three sessions. The first session served as a baseline assessment to recognize distinct levels of individual gameplay performance. The second session was used to learn how

to classify emotion states with EEG, EOG and kinematic motion data. In the last session, the participants played a modified version of the Whack-a-Mole game with the three difficulty levels (*Too Easy*, *Optimal*, and *Too Difficult*) in a random order defined in the first session. While the participants were wearing the OpenBCI and JINS MEME devices during all the sessions for symmetry, collection of sensor data only occurred in the second and third session.

Session 1: Identifying Gameplay Levels of Distinct Difficulty The objective of this session was to associate each participant with three game-play levels of three distinct difficulty levels.

Participants played the Whack-a-Mole game through a sequence of 15 levels of increasing cognitive demand. Each level contained the same number of targets, but displayed them at a speed that increased by 12% each level. Based on each participant's gameplay performance, three levels were defined for each participant: *Too Easy*, *Optimal*, and *Too Difficult*. The very first level of the game was completed with ease by every participant, and was thus defined as *Too Easy* for all participants. *Optimal* was defined individually for each participant as the highest level in which they achieved a perfect score by hitting all moles and avoiding all bunnies. *Too Difficult* was defined as two levels above *Optimal* — approximately 25% faster in speed.

Session 2: Learning to Classify Emotions with EEG, EOG and kinematic motion The objective of this session was to collect self-reported emotional-state data in response to viewing IAPS images, as well as biosignals during viewing these images, in order to evaluate how accurately the participants' self-reported emotional responses can be inferred from the biosignals collected from our chosen wearable devices.

In the second session, sensor data was continuously collected as participants viewed a random sequence of images from the international affective picture system (IAPS). Participants were shown 36 images from the IAPS library that were selected to evoke a range of emotional responses. Participants indicated their emotional response to each image by filling out a questionnaire shown in the figure B.1 in which they assigned scores for valence and arousal on a scale from 1 to 5. Images were shown for 6 seconds after the presentation of a fixation cross for 4 seconds. Each image was followed by a prompt to fill out the questionnaire in the figure B.1 (Appendix B), and participants were given 6 seconds to do so after each image.

Session 3: arousal and valence at Different Game-play Levels The objective of this session was to collect self-reported emotional-state data in response to playing the Whack-a-Mole game at the three distinct difficulty levels, as well as biosignals at the end of each game-play session, in order to analyze how difficulty levels correlate with the player's self-reported emotional state and the emotional state inferred by the biosignal classifier.

The third session started after a short break after the second session. Participants played a version of the Whack-a-Mole game that presented the three levels (*Too Easy*, *Optimal*, and *Too Difficult*) in a random order. Each level was presented for a consistent duration of 60 seconds. This differs from gameplay succession in the first session, in which each level displayed a consistent number of targets. After the end of each level, participants filled out the PANAS questionnaire shown in the figure B.2 (Appendix B).

3.3.2 Image Selection Procedure

The IAPS system assigned average valence (V) and arousal (A) scores to each image based on the average participant response. Separate scores were assigned for male (M) and female (F) participants on a scale from 1 to 9.

For this study, we wanted to present the participants with images that would evoke strong responses and would give us clear, unambiguous emotional-state labels to analyze against the participant self-reported emotional responses.

We overlaid the Circumplex Model of Affect (see figure 3.1) over a grid separated into 9 categories (see table 3.1). Each of these 9 categories represented a region on the Circumplex Model of Affect. This would allow us to use simple formulas to determine which images fit most clearly into each grid section.

First, the following formula was used to normalize the IAPS scores from -1 to 1, placing the origin in the centre of the Circumplex Model of Affect:

$$V = \frac{V_{\text{IAPS}} - 1}{4} - 1 \quad A = \frac{A_{\text{IAPS}} - 1}{4} - 1 \quad (3.1)$$

Each image was ranked according to its value for each of the formulas, and the 4 images with the highest ranking for each formula were placed in the corresponding category. For categories where the selected images overlapped with selected images from other categories, the image was placed in the category where it had a higher ranking. For example, if an image had the fourth highest ranking in category 7 and the second highest ranking in category 4, it would be placed in category 4.

Since participants in this study assigned valence and arousal scores on a scale of 1-5, the IAPS image scores were normalized to the same scale for the sake of comparison.

Table 3.1: Image selection procedure and used formulas for the selection of 4 images from each of the 9 categories.

Category	Position	Valence	Arousal	Formula	Selection
1	Left Top	Low	High	- V + A	Highest Values
2	Centre Top	Neutral	High	- V _{abs} + A	Highest Values
3	Right Top	High	High	V + A	Highest Values
4	Left Centre	Low	Neutral	- V - A _{abs}	Highest Values
5	Centre Centre	Neutral	Neutral	- V _{abs} - A _{abs}	Highest Values
6	Right Centre	High	Neutral	V - A _{abs}	Highest Values
7	Left Bottom	Low	Low	- V - A	Highest Values
8	Centre Bottom	Neutral	Low	- V _{abs} - A	Highest Values
9	Right Bottom	High	Low	V - A	Highest Values

3.3.3 Ground Truth for Arousal & Valence

Following Wang *et al.* [72], in this work, we used two different classification systems: Multi-classification for arousal and valence with three classes (low, neutral, and high) and two binary classification with two classes (low and high).

For the training labels, the Circumplex self-assessments of each subject on two dimensions were used as can be seen in figure B.1 in Appendix B.

As shown in figure 3.3, all original label values from Circumplex self-assessments are between [1,5] and needed to be transformed into binary classes. The way to determine the threshold 3 is intuitive, but the distinction between high and low emotion of one subject is difficult. The rating of high and low values is highly related to personal tendencies; hence, the labelling system decreases in performance [72]. To compensate this imbalance, we determined subject-dependent mean values to define labels for all reactions to 36 images for each of the 21 participants. The usage of mean values instead of median values led to less ambiguity even if the median was labelled positive or negative. The x represents the assigned scores for valence and arousal on a scale from 1 to 5.

Following equation represents an example calculation of the subject-dependent mean value of participant 1 depicted Figure 3.3.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{(2 + 2 + 3 + 3 + 3 + 3 + 4 + 4 + 4 + 5)}{10} = 3.3 \quad (3.2)$$

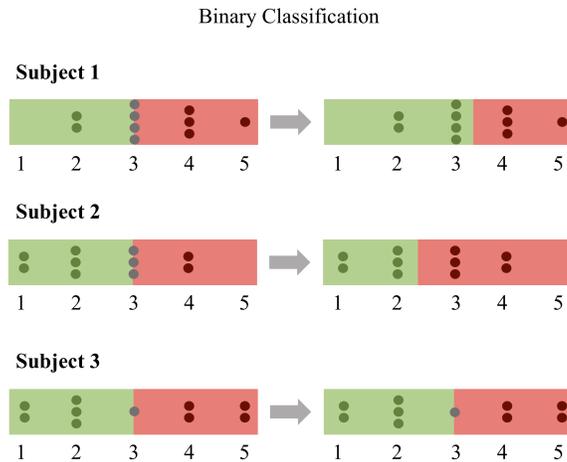


Figure 3.3: Visual explanation of Binary Classification.

The mean values are visualized for each subject 1, 2, and 3 (see figure 3.3). It is to be noted, that figure 3.3 is just an illustrative example and shows the hypothetical reactions of 3 subjects to 10 images.

3.3.4 Arousal & Valence From PANAS Questionnaire

Past research has often used PANAS to assess the subjects' emotional state after completing playing a serious game [4, 78]. As can be seen in figure B.2, the PANAS questionnaire is separated in positive and negative affect. Each of the affected sub-scales contains ten adjectives. Further, each of the ten items was designed to be internally consistent and to have convergent correlations with lengthier measures of the underlying mood factors [4]. In **session 3**, the participants filled out the PANAS questionnaire (see Appendix B.2) as self-reported measures of emotional states after playing each level of the game Whack-a-Mole. The gamers were asked to rate the extent to which they have experienced each particular emotion on a 5-point scale (1 = "Very slightly or not at all" to 5 = "Extremely").

For the calculation of positive and negative affect, further additional scores were required. Hence, the following ten scores were added to calculate positive affect: *Interested, Exited, Strong, Enthusiastic, Proud, Alert, Inspired, Determined, Attentive, and Active*. By rating each score on a 5-point scale, the results (sum of all ratings) can range between 10 to 50, where a higher score indicates a higher level of positive affect. On the other hand, to calculate negative affect the ten scores *Distressed, Upset, Guilty, Hostile, Scared, Hostile, Irritable, Ashamed, Nervous, Jittery, and Afraid* had to be added, where a lower score represents a lower level of negative affect. According to Watson *et al.* [4], the mean scores for momentary positive affect is 29.7 (SD = 7.9) and the mean momentary negative affect score is 14.8 (SD = 5.9). As a result, a positive affect score equal 30

represents *Neutral*, higher than 30 *High*, and lower represents *Low* affect. On the other hand, a negative affect score equal 16 represents *Neutral*, higher than 16 *High*, and lower represents a *Low* affect.

The pre-trained machine learning models in **session 2** were based on the two-dimensional arousal and valence Circumplex model of the IAPS image responses. Therefore, after calculation of the resulting scores, the positive and negative affect was also mapped as two-dimensional arousal and valence Circumplex model (see image 3.1). [4]. Exemplary, a positive affect score of 35 (*High*) and a negative affect score of 15 (*Neutral*) result in *High* valence and a *High* arousal. Table 3.2 lists the emotional states associated with different positions in the Circumplex model shown in figure 3.1.

Table 3.2: Relationship between Circumplex and PANA models of classifying human emotions.

Angle	Position	Valence / Arousal (Russell)	Positive / Negative Affect (Watson & Tellegen)
90°	Top	Neutral Valence, High Arousal	High Positive Affect, High Negative Affect
45°	Top Right	High Valence, High Arousal	High Positive Affect, Neutral Negative Affect
0	Right	High Valence, Neutral Arousal	High Positive Affect, Low Negative Affect
315°	Bottom Right	High Valence, Low Arousal	Neutral Positive Affect, Low Negative Affect
270°	Bottom	Neutral Valence, Low Arousal	Low Positive Affect, Low Negative Affect
225°	Bottom Left	Low Valence, Low Arousal	Low Positive Affect, Neutral Negative Affect
180°	Left	Low Valence, Neutral Arousal	Low Positive Affect, High Negative Affect
135°	Top Left	Low Valence, High Arousal	Neutral Positive Affect, High Negative Affect

3.4 Data Collection & Analysis

Figure 3.4 shows our approach in this study. The emotion recognition starts with the collection of the reactions to the IAPS images using biosignals and the self-assessment questionnaire. As a first step, we input the EEG, EOG, Accelerometer (ACC), and Gyroscope (GYRO) to the synchronized frame. Secondly, the signals were cleaned and pre-processed. We calculated 281 attributes to evaluate the relevance of our input signals. Then, imputation, feature selection, and standardization were applied to enhance the training process.

On the other hand, the self-assessment questionnaire was used as ground truth for arousal and Valence by using subject-dependent mean values. This is explained in more detail in section 3.3.3. The machine learning models were trained corresponding to the ground truth defined previously using Random Forest (RF) and Ensemble Classifier (EC) (Support Vector Machine (SVM), Random

Forest (RF) and Logistic Regression (LR)). Next, an evaluation of our machine learning engine was applied using 10-fold cross-validation to compare the proposed approach to related works.

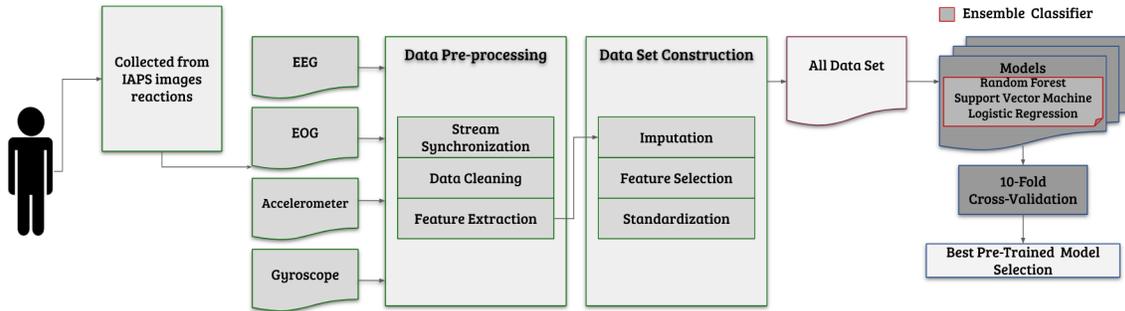


Figure 3.4: Flowchart of the approach for training the model using IAPS images and 10-fold cross validation.

Figure 3.5 shows the use of the pre-trained machine learning models for the emotional state prediction of the participants while playing the game. The new data set (biosignals collected from playing games and PANAS questionnaire) was inputted, as shown in figure 3.5. PANAS questionnaire input was mapped into a Circumplex model of affect corresponding to the ground truth of our new data set. This is explained in more detail in section 3.1.1. Following the steps depicted in figure 3.4, first, the data was cleaned, and features were extracted from the time series data. Finally, the machine learning engine was pre-trained with the used IAPS images.

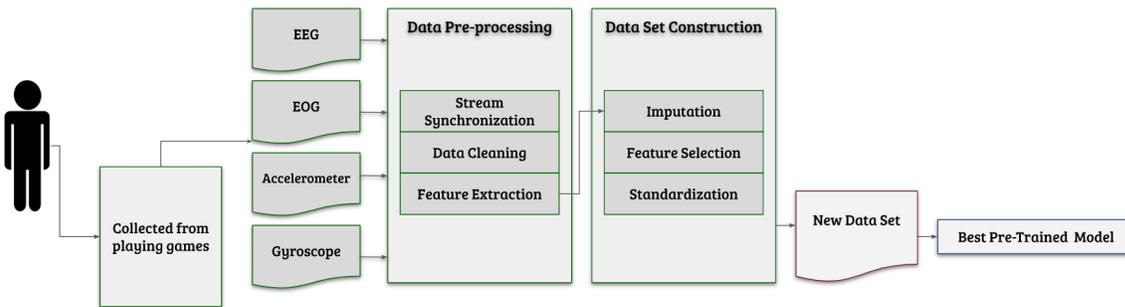


Figure 3.5: Flowchart of the proposed approach to predict the emotional state of the participants while playing the game.

3.4.1 Data-Stream Synchronization

Collected data from OpenBCI (OBCI) and JINS MEME (JM) were emitted to a collecting computer in the Service Systems Research Lab at the University of Alberta.

We used the timestamp of the initial appearance of each IAPS image on the subject's display as sentinels for synchronizing the data collected from the two wearable devices. Since the OBCI

and JM devices had different sample rates (OBCI - 125 Hz, JM 100 Hz), one signal had to be oversampled, and the other had to be undersampled. The oversampling approach can amplify the noise, and therefore, the topological characteristics of the signal can be deformed. On the other hand, the undersampling can remove essential features from the original signal. To avoid these issues, we designed an emotion profile per participant, which contained a synchronization based on the stages ("start of image display", "fixation cross appearance", and "PANAS questionnaire response").

Python was used to slice and organize the data using an associative array. In this case, four main key-values (IDs, stages, data types, and event counters) were used. IDs are unique identifiers to protect the identity of the participants. The stages were defined as "start of image display", "fixation cross appearance", and "PANAS questionnaire response", with a predefined time of 6s, 4s, and 6s, respectively. The data types were based on wearable devices such as OBCI and JM, and the event counter was defined as the number of displayed images. We used the timestamp to slice the data with the events from the displaced images without the need for over- or undersampling the signals.

In the third session, the timestamps collected by playing the game Whack-a-Mole were used to synchronize the data obtained from wearable devices, using the same approach. These timestamps corresponded to player's start and end time of each the three levels (Too Easy, Optimal, Too Difficult) in random order.

3.4.2 Feature Extraction

Feature extraction started from an initial set of measured data and derived values from the time series of the raw signals were built. The feature extraction intends to be informative and non-redundant, to facilitate the subsequent learning, and to generalize the processing steps. For the evaluation of the participant's reactions to the shown images, Python was used to calculate 281 attributes using EOG, EEG, H-ACC, G-ACC and G-GYRO data, which is shown in table 3.3. These attributes contain information about head movements, neurofeedback, and eye movements, as well as attributes calculated from the raw signals based on a set of attributes used for emotion recognition with EEG known from previous researches [72, 79, 80].

EEG: Based on the literature [66, 72, 70, 69], a 50 Hz Notch filter and a band-pass filter with 0.3-60Hz on the 16 channels were applied. The extraction of the EEG Power Band was carried out *via* Fourier transformation. Therefore, the high-frequency signals (Theta, Alpha, Beta, Gamma) were calculated for each of the 16 EEG channels. Also, the asymmetry of the average power spectral density of the band relations Delta/Theta, Delta/Alpha, Delta/Beta, Theta/Alpha, Theta/Beta, and Alpha/Beta were extracted.

Table 3.3: Feature extraction per signals.

Extracted Features	
Electroencephalography (EEG) (112 features)	Mean, Standard Deviation (Std) of the Average 5 Bands (Theta, Alpha, Beta, Gamma, Delta) and Delta to Theta, Delta to Alpha, Delta to Beta, Theta to Alpha, Theta to Beta, Alpha to Beta and Power Spectrum Density (PSD) in 16 channels (112 features from 16 ch x 7 features per channel)
Electrooculography (EOG) (63 features)	Horizontal and Vertical Saccade Movement, Peak of FFT Horizontal and Vertical, Horizontal and Vertical Maximum, Maximum, Blink Mean, Std, Max, Min, Sqr, Kurtosis, Skew, PSD, Mean Jerk, Std Jerk , Max FFT, Total Saccades Movements (Up, Down, Left and Right), Total Number of Blinks (63 features from 3 ch (Horizontal, Vertical and Result Vector) x 21)
Head Acceleration (H-ACC) (66 features)	Mean, Standard Deviation, Max, Min, Sqr, Kurtosis, Skew, Power Band, (PSD), Energy, Mean Jerk, Std Jerk , Max FFT, Periodogram (66 features from 3 ch (X, Y and Z) x 22)
Glasses Acceleration (G-ACC) (66 features)	
Glasses Gyroscope (G-GYRO) (40 features)	Mean, Standard Deviation, Max, Min, Sqr, Kurtosis, Power Band, Periodogram (40 features from 4 ch (Pitch, Roll, Yaw, and Normal Vector) x 10)

EOG: Based on the literature [11, 12, 71], a 20 Hz low-pass filter and 0.5 Hz high-pass filter were applied to remove the DC noise of the signal. A 0.05 seconds moving average filter was used to smooth the EOG signal and to remove signal noise and to preserve the overall structure of the eye movements. The most characteristic feature was the difference between steady periods without eye movements and periods with eye movements. Hence, to detect the eye movements, the eye acceleration was extracted by using the first derivation of the EOG signal from JM. The first derivative of the EOG signal characterizes all eye movements by a local maximum and a local minimum. The maximum and minimum lag-lead characterization determines the direction of an eye movement in either the vertical or horizontal dimension. This leads to the extraction of features such as horizontal and vertical saccade movement, the peak of FFT horizontal and vertical, horizontal and vertical maximum, minimum, and blink mean.

Accelerometer & Gyroscope: Following the literature [68, 70, 12, 73], a 0.5 Hz high-pass

filter was used to eliminate the dynamic human kinematic motion while preserving the static human kinematic motion or posture information. A 5-20 Hz band-pass filter was applied to extract the components of the static signals and to remove the DC components of the signal. Then, standard features were calculated, such as mean, standard deviation, jerk, and skewness in the X, Y, and Z directions of the accelerometer.

Imputation

Imputation is the process of estimating and replacing missing data with synthetic values [81]. In our study, missing values occurred due to Bluetooth communication issues. Therefore, we used the median value imputation to determine the missing values from the EOG features mean jerk and peak of FFT horizontal and vertical. According to Setz *et al.* [81], in cases where the applied classification model employs a feature whose value is missing in the test sample, imputation is required.

Feature Selection

Feature selection is required, in order to eliminate redundant features and reduce the model complexity of machine learning algorithms [72]. Previous research proved that using feature selection for emotion recognition based on Random Forest (RF) can enhance the sadness recognition (low arousal and valence) and improve the classifier performance. RF is constructed out of many unpruned decision trees on several bootstrap samples and uses the OUT-OF-BAG (OOB) error to measure the prediction error. As shown in table 3.4, we selected the 30 most important features based on the feature importance score using the Random Forest method as known in the literature [82].

$$FI(X^j) = \frac{1}{\text{number of trees}} \sum_t (|err(X_t^j) - err(X_{t-oob}^j)|) \quad (3.3)$$

Feature importance (FI) is defined as the increment of the mean error while changing the feature value X_t^j randomly by one of its values from the OOB set (see equation (3.3)) [82].

Standardization

After the feature selection, feature scaling was applied. The two common ways for scaling are min-max scaling and standardization. In this study, we chose the latter due to less affection by outliers. In the case of min-max scaling, the min value was subtracted and divided by the difference of max and min. The shifted and rescaled values ended up ranging from 0 to 1. For standardization,

Table 3.4: 30 Features selected for emotion recognition based on Random Forest. (The common features between Valence and Arousal are underlining and bold).

	Valence	Arousal
EEG	Mean of Theta P4, Alpha to Beta C4, Mean of Alpha F7, Standard Deviation of Delta O1, Delta to Beta C4, Delta to Alpha C4, Theta to Beta C4, Average Band Power of Beta T6, Standard Deviation of Alpha O1, Average Band Power of Beta Fp1, Mean of Gamma Fp2, Delta to Theta C3, Delta to Beta T6, Delta to Alpha T6, Mean of Beta T5 (15 features)	Standard Deviation of Gamma Fp1, Standard Deviation of Beta T3, Alpha to Beta Fp1, Theta to Alpha P3, Theta to Alpha F4, Standard Deviation of Delta T5, Standard Deviation of Delta C3, Delta to Theta P3, Average Band Power of Alpha Fp1, Standard Deviation of Delta Fp2, Mean of Alpha C3, Theta to Alpha T3, Average Band Power of Delta Fp2, Theta to Alpha T5, Theta to Alpha C4, Theta to Alpha O2, Standard Deviation of Delta T6 (17 features)
EOG	Horizontal Saccade Movement, Peak of FFT Horizontal, Horizontal Maximum, Peak of FFT Vertical , Maximum of Result Vector (5 features)	Vertical Maximum, Peak of FFT Vertical (2 features)
H-ACC	Peak of FFT of Acceleration in Z-Direction, Median of Result Vector, Peak of FFT of Acceleration in Y-Direction, Peak of FFT of Result Vector (4 features)	Median of Acceleration in Z-Direction, Maximum of Acceleration in X-Direction, PSD of Acceleration in Y-Direction (3 features)
G-ACC	PSD of the Acceleration in Z-Direction, Peak of FFT of Acceleration in Z-Direction, Peak of FFT of Result Vector (3 features)	Energy of Result Vector, Peak of FFT of Result Vector , Standard Deviation of Jerk, Energy of Acceleration in Y-Direction, Energy of Acceleration in Z-Direction (5 features)
G-GYRO	Minimum of Pitch Axis, Maximum of Pitch Axis, Standard Deviation of Pitch Axis (3 features)	Maximum of Acceleration in Y-Direction, Standard Deviation of Roll Axis, IQR of Result Vector (3 features)

the mean value was subtracted and then divided by the variance. Hence, the final distribution had a unit variance [83].

3.4.3 Learning a Classifier Model

In the literature, machine learning algorithms such as Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF) were successfully applied to classify biosignal data associated with affective/emotional states [84, 85, 86]. On that basis, in this approach, we used machine learning approaches mentioned previously, and created our hard voting classifier.

Support Vector Machine is a machine learning model suitable for linear or nonlinear classification, regression, and outlier detection. According to Sohaib *et al.* [84], SVM is predicted to classify EEG data accurately. The kernel function will be applied for data, which cannot be separated linearly. The most common kernel functions are linear, poly, radial basis, sigmoid, etc. Following Chen *et al.* [87], in this work, we used a linear kernel function for the SVM method.

Logistic Regression is commonly used for the estimation of the probability that an instance belongs to a specific class. With a binary labeling system of 1 (positive class) and 0 (negative class), the instance will belong to that class where the estimated probability is higher than 50%. Therefore, a weighted sum of the input features is computed. The final result is their logistic output, where logistic is defined as a sigmoid function giving a number between 0 and 1 as output [83]. LR can also be used for brainwave classification, according to the literature [87].

Random Forest is a method generally used in the classification of EEG signals and based on the decision-tree method. Multiple trees are integrated as an algorithm following the idea of machine-learning-integrated learning methods [87]. The randomness of the Random Forest algorithm increases with growing trees. The algorithm finds the best feature among a random subset of features. This yields in general in a better model, due to greater tree diversity, which lowers the variance for a higher bias [83].

Ensemble Classifier Given a set of classifiers (SVM, LR, and RF) trained on the same data, a potentially better classifier can be constructed by combining their predictions to identify the class with the most votes. The resulting hard voting classifier can obtain higher accuracy, in comparison to the best classifier in the ensemble [83]. According to Schuller *et al.* [88], emotion samples are hard to access, especially at a higher number of emotion samples. Hence, even if each classifier is a weak learner by itself, the ensemble can be a strong learner [83]. This technique is called ensemble learning, and its algorithm is called ensemble method. Nevertheless, it has to be noted, that this is only true if all classifiers are perfectly independent. This will provide an improvement of the ensemble's accuracy [83].

3.5 Findings

3.5.1 Recognizing the Self-Reported Emotional States Through Biosignals

In this section, we present the classification results for the detection of the degrees low, medium, and high of valence and arousal, achieved by biosignal features while participants were observing IAPS images to response our first research question **Q1**. First of all, the feature set and the classification system were evaluated. The comparison of the classifiers SVM, LR, RF, and EC was carried out using 10-cross-validation for each subject.

To investigate our first research question, **Q1: How accurately can the player's self-reported valence and arousal be predicted based on EEG, EOG, and kinematic motion data?**, we experimented with different machine-learning methods for learning classifier on different feature vectors. Each different feature-vector+algorithm combination was evaluated with 10-cross-validation for each subject.

For the evaluation of our machine learning approach, four experiments were conducted.

1. For the first experiment, the defined pipeline shown in figure 3.4 was used. Therefore, no feature selection was applied to train the machine learning algorithm with all features.
2. For the second experiment, 281 features and SMOTE (Synthetic Minority Over-sampling Technique) were used in order to balance the learning, and no feature selection was applied. SMOTE was applied for the counteracting of the class-imbalance problem. Therefore, an algorithm to construct new data based on the existing data was used. This procedure increased the amount of data in underrepresented minority classes while avoiding overfitting.
3. In the third experiment, the 30 most important features were used for the classifier construction.
4. For the last experiment, we used the 30 most important features and applied SMOTE.

The tables 3.5 and 3.6 show the average values for each participant. As can be seen in table 3.5, all classifiers learned from the original pipeline exhibited fairly low performance. One possible explanation for this is the high number of feature vectors, which results in the need for more complex decision boundaries. Furthermore, the complexity of the learning algorithm increases with additional dimensions. Another explanation can be the unequal representation of the classes.

Reviewing the results of experiments 2 and 4, an improvement by using SMOTE of the classification performance for arousal and valence can be observed. As described before, classification

build on class-imbalanced data favors the majority class. With an increasing dimension of the data, this tendency is even more significant, due to a higher number of variables than several samples. Nevertheless, it has to be noted that SMOTE is less efficient for high-dimensional data [89].

Of the four tested machine learning algorithms, the better overall performance was obtained using the 30 most important features and SMOTE. Using the 10-cross-validation across all subjects, the arousal and valence accuracy ranges from 70% to 94% for the binary classifier, and from 62% to 89% for the multi-classifier (see table 3.5).

The Ensemble Classifier and Random Forest showed the best results. While their performances are similar, the Ensemble Classifier obtained a more efficient recall, indicating that it returned most of the relevant results. Moreover, further evaluation based on the SMOTE data with the 30 most important features showed a two times better performance in comparison to the other feature set as it can be seen in table 3.5. Overall, it can be said that the emotional state can be recognized based on EEG, EOG, and kinematic motion data.

Table 3.5: Comparison of mean performance across 10-fold user’s average of the classifiers Logistic Regression, Random Forest, polynomial liner SVC, and Ensemble Learning.

		Multi-Classifier						Binary Classifier					
		F1-Score		Recall		Accuracy		F1-Score		Recall		Accuracy	
	Features	V	A	V	A	V	A	V	A	V	A	V	A
SVM	All (281)	0.40	0.41	0.40	0.41	0.41	0.43	0.49	0.41	0.51	0.46	0.58	0.58
	All (281) + SMOTE	0.68	0.70	0.70	0.71	0.69	0.71	0.70	0.76	0.74	0.87	0.70	0.79
	Important (30)	0.66	0.73	0.67	0.74	0.67	0.74	0.51	0.48	0.51	0.50	0.58	0.66
	Important (30) + SMOTE	0.63	0.72	0.64	0.73	0.64	0.74	0.66	0.81	0.69	0.87	0.66	0.81
Logistic Regression	All (281)	0.37	0.41	0.37	0.41	0.37	0.43	0.47	0.42	0.49	0.43	0.55	0.58
	All (281) + SMOTE	0.65	0.68	0.66	0.69	0.67	0.69	0.67	0.81	0.71	0.88	0.65	0.79
	Important (30)	0.64	0.67	0.65	0.69	0.61	0.70	0.51	0.53	0.52	0.53	0.59	0.73
	Important (30) + SMOTE	0.61	0.71	0.62	0.72	0.61	0.73	0.65	0.82	0.66	0.86	0.64	0.82
Random Forest	All (281)	0.40	0.45	0.43	0.49	0.40	0.43	0.44	0.31	0.40	0.30	0.61	0.68
	All (281) + SMOTE	0.60	0.68	0.43	0.68	0.61	0.69	0.65	0.78	0.65	0.81	0.66	0.79
	Important (30)	0.61	0.68	0.43	0.68	0.61	0.67	0.57	0.48	0.52	0.47	0.69	0.76
	Important (30) + SMOTE	0.73	0.78	0.75	0.77	0.75	0.79	0.75	0.85	0.74	0.86	0.75	0.84
Ensemble Learning	All (281)	0.39	0.43	0.45	0.42	0.39	0.49	0.38	0.31	0.36	0.32	0.61	0.69
	All (281) + SMOTE	0.66	0.70	0.66	0.71	0.67	0.71	0.69	0.82	0.79	0.88	0.65	0.80
	Important (30)	0.65	0.71	0.67	0.71	0.66	0.71	0.47	0.44	0.44	0.43	0.64	0.76
	Important (30) + SMOTE	0.67	0.77	0.72	0.78	0.73	0.79	0.75	0.85	0.83	0.88	0.72	0.84

Table 3.6 shows the comparison of f1-score and accuracy using the three signals, EEG, EOG, and kinematic motion between subjects. Following four scenarios were tested: using only EOG, EEG, or kinematic motion, and using all three signals. As can be seen in table 3.6, high accuracies

across all users for the classification of arousal and valence were achieved. This provides further evidence for the hypothesis that movement sensor data can be used for emotion recognition. For the creation of a personal model as future work, we can avoid EEG signal features due to their complexity for measuring and tracking emotions, and focus on EOG and kinematic motion data. Overall, the best accuracies for valence (75%) and arousal (84%) were achieved using the binary classifier.

Table 3.6: Comparison of Valence and Arousal for the signals EEG, EOG, kinematic motion and their sum using the 30 most important features and SMOTE.

		Binary Classifier					
		F1-Score		Recall		Accuracy	
	Data Set	V	A	V	A	V	A
Random Forest	EOG	0.68	0.80	0.69	0.82	0.69	0.79
	EEG	0.71	0.80	0.70	0.69	0.72	0.81
	Motion	0.70	0.82	0.70	0.81	0.70	0.84
	All	0.75	0.85	0.74	0.86	0.75	0.84
Ensemble Learning	EOG	0.67	0.80	0.66	0.82	0.67	0.79
	EEG	0.67	0.80	0.69	0.83	0.68	0.80
	Motion	0.69	0.83	0.75	0.85	0.68	0.83
	All	0.75	0.85	0.83	0.89	0.72	0.84

Table 3.7 summarizes the overall final results of the presented analysis and compares it to similar research found in the literature. Due to a high overall performance, it is possible to state that our method is able to predict emotional states accurately. These results suggest that the JINS MEME glasses, EEG, kinematic motion, and Ensemble Classifier can be used to classify emotional states reasonably. It has to be noted that this table is primarily meant as an overview and not as a valid comparative evaluation of the reported methods.

3.5.2 Correlation between Gameplay Difficulty and Self-Reported Emotional State

To identify the dependency between the three Whack-a-Mole levels (*Too Easy*, *Optimal*, and *Too Difficult*), and the self-reported emotional states (arousal and valence) of the players during these sessions, chi-squared tests were performed. Chi-squared tests were inputted using emotional states mapped from PANAS to a 2-D (arousal and valence) Circumplex Model (see section 3.1.1), and the game levels. The result of emotional states mapped from PANAS is shown in figure 3.6.

Following, the next research question we set to examine was **Q2: How does the player's**

Table 3.7: Accuracy between proposed method and similar works using Databases (DB).

Year: Author(s)	Wearable Sensor: Sensor Type(s)	# of Subjects	Accuracy	
			Binary Classifier	
			V	A
2019: Our method	Electrooculography, Accelerometer, Gyroscope, Electroencephalography	21	72.02%	84.30%
2018: Wang <i>et al.</i> [72]	Galvanic Skin Response, Electrocardiography, Electroencephalography	33 (DB)	80.01%	68.40%
2017: Wiem <i>et al.</i> [69]	Electrocardiogram, Galvanic Skin Response, Skin Temperature, Respiration Volume	24 (DB)	68.75%	64.23%

emotional state changes as the game difficulty changes? To answer to this question, we tested three different Null Hypotheses (NH). The first hypothesis assumed that there is no association between the game levels (*Too easy*, *Optimal*, *Too hard*) of the game and the player's arousal (high and low). The second hypothesis assumed that there is no association between the levels and the valence (high and low). The third hypothesis assumed that there is no association between the emotional state (arousal and valence) and the game levels.

Correspondingly, the alternative hypotheses posited that there is an association between the levels of the game and the player's arousal and valence.

The chi-square analysis revealed a correlation between the game levels and arousal with $\chi^{2(2)} = 7.843$, rejection of the first NH with $\rho = 0.020$, no correlation between the game levels, valence with $\chi^{2(2)} = 2.471$, and a failure to reject the second NH with $\rho = 0.291$. The result of the chi-square analysis was determined by a 0.05 level of significance and two degrees of freedom. This result confirms that the players' arousal emotional state changes while playing the game Whack-a-Mole and their change of difficulty as can be seen in figure 3.6. The arousal decreases with increasing difficulty of the game levels. On the other hand, the players' valence emotional state increases with increasing difficulty of the game levels, but there is a lack of association between the valence and levels difficulty.

To examine the third hypothesis, a chi-squared test between emotional state (arousal and valence) and the game levels was performed. The chi-square analysis revealed a correlation between emotional state and the game levels *Too easy*, and *Too difficult* with $\chi^{2(2)} = 10.255$, rejection of the third NH with $\rho = 0.017$, but not for the *Optimal* level. The disassociation between the *Optimal* and (*Too easy*, and *Too difficult*) levels reveals a significant difference between the difficulties *Too Easy*



Figure 3.6: Mapping of PANAS as 2-D Russell diagrams for all participants. The y-axis presents the overall participants emotional states. For each participant, scores for valence and arousal for each level based on their self-reported emotional response (high = 1 and Low = -1) were assigned.

and *Optimal*.

3.5.3 Predicting Player's Arousal & Valence Through The Pre-trained Machine Learning Models

Finally, we used the pre-trained Random Forest and Ensemble Classifier models which have been used for the classification of arousal and valence based on IAPS images in order to predict the emotional states after the participants were playing the game Whack-a-Mole. This section answers **Q3: How accurately can the players' emotional states be predicted using pre-trained machine learning models with the use of IAPS images during gameplay when the difficulty of the game changes?**. The results of arousal and valence predicted by playing the game are presented in table 3.8. We used the transformation of the PANAS to the Circumplex Model as ground truth. First of all, we can observe that the performance of the Ensemble Classifier is better than the Random Forest. Hence, our hard voting system can be more robust to predict emotions. On the other hand, our results show that the predictions for valence are 7% higher than the predictions for arousal. Table 3.9 summarizes results obtained in this work compared to the literature.

Table 3.8: Obtained results of arousal and valence predicted by playing the game Whack-a-Mole.

		Binary Classifier	
		Accuracy	
	Level	V	A
Random Forest	<i>Too easy</i>	0.65	0.65
	<i>Optimal</i>	0.71	0.70
	<i>Too difficult</i>	0.90	0.70
	Overall	0.75	0.68
Ensemble Learning	<i>Too easy</i>	0.70	0.80
	<i>Optimal</i>	0.75	0.65
	<i>Too difficult</i>	0.95	0.75
	Overall	0.80	0.73

It has to be noted that this table is primarily meant as an overview and not as a valid comparative evaluation of the reported methods since these methods have not been applied to a common data set and experimental scenario yet. Nevertheless, our approach seems to show high performance in accuracy concerning the literature.

Table 3.9: Accuracy between proposed method and similar works using different games.

				Accuracy	
				Recognized Emotions or Affects	
Author(s)	Physiological Signal(s) or Sensor Type(s)	Game	# of Subjects	V	A
2019: Our method	Electrooculography, Electroencephalography, Accelerometer, Gyroscope	Whack-a-Mole	21	80.32%	73.02%
2018: Yang <i>et al.</i> [75]	ECG, EDA, EMG, respiration and body movement with a 3-axis accelerometer, facial recording, game screen recording	FIFA 2016	58	50.70%	50.40%
2017: Alhargan <i>et al.</i> [76]	Pupil responses with HT (eyes tracking)	Speedboat	14	61.40%	76.00%
				Excitement	Happiness
2016: Huynh <i>et al.</i> [77]	Galvanic Skin Conductance (GSR), Photoplethysmography (PPG), Electroencephalography (EEG)	Tank 1990 HD	22	77.38%	73.21%

Our results are also consistent with previous physiological signals based on affect recognition studies (see table 3.9). Our results are relatively lower on arousal. This could be a result of the similarity of the difficulty of the game levels *Too easy* and *Optimal*. Hence, a strong enough stimuli can not be achieved. Thus, this game design will investigate the maximizing of the recognition performance on the valence dimension.

3.6 Summary

In this work, we proved that emotion could be recognized based on EEG, EOG, and kinematic motion data. We verified the effectiveness of 4 classifiers (SVM, LR, RF, and EVC) in subject emotion recognition, including 3 kind of data sets (all-features, all-features and SMOTE, the 30 most important features and SMOTE) and emotion prediction for 3 levels (*Too easy*, *Optimal*, *Too hard*) of the game Whack-a-Mole. We adopted a 10-cross-validation method for each subject to verify the performance of the proposed features. Feature extraction methods were enhanced by using the random forest to evaluate the complexity of physiological signals. The proposed method reaches 83% and 73% accuracy for valence and arousal, respectively. These results are comparable to those achieved in past research using biosignals for emotion recognition, with the advantage that JINS MEME may have a higher likelihood of user acceptance.

We evaluated the players' emotional state changes while playing the game Whack-a-Mole and their change of difficulty. We found a strong correlation between arousal and the levels, but no clear relationship between valence and level difficulty was observed. This can be a result of missing stimuli of the activity performed. Furthermore, our study predicted that the players' emotional state increases during gameplay when the difficulty of the game changes.

On the other hand, the study suggests that it is possible to detect emotions based on eye movements and kinematic motion data. Using wearable devices such as JM and OBCI, is an excellent complement for collecting data, to provide biofeedback, and for intervention. On the other hand, playing the game Whack-a-Mole can raise emotions based on the difficulty of the levels.

Conclusion and Future Work

4.1 Conclusion

This thesis presents two studies designed to examine the potential of an off-the-shelf eyewear device, JINS MEME, and OpenBCI, as a means of sensing and accurately recognizing activities of daily living (first study) and emotion recognition (second study). Additionally, this thesis contributes to the state of the art of the use of non-traditional EOG signals for activity recognition, and the extraction of indications for arousal and valence which can imply the increase of cognitive load.

In the first study, twenty-six participants spent a one to two-hour session in a one-bedroom living environment, either alone or in pairs, and performed a scripted protocol of activities of daily-living. Twelve of these participants wore the commercial smart eyewear device JINS MEME, which collected electrooculography, accelerometer and gyroscope data throughout their sessions. In this study, we demonstrated that *both motion-based and visual-based activities can be classified accurately*, but *we cannot differentiate specific activities* such as taking medication and eating food. This study shows that our *method yields equal or better results* with a variety of activities compared to approaches that involve more restrictive wearable device setups. Furthermore, we develop *a method for calculating information about eye and head movements from JINS MEME*, in addition to standard characteristics of EOG and kinematic motion signals. The results demonstrate the *suitability of JINS MEME for recognition of activities* of daily-living.

In the second study, 21 participants viewed a sequence of images from the International Affective Picture System (IAPS) database. Wearable sensor devices collected electroencephalography (EEG), electrooculography (EOG), and kinematic motion data as participants viewed the images; the participants also characterized their own emotional responses to the images. Participants then played

the serious game “Whack-a-Mole”, wearing the sensor devices. They played three levels of the game that required varying amounts of cognitive effort. This study demonstrated that the *emotional state during a task can be accurately determined using collected data from wearable sensor devices*, with and without self-reported measures, and we found *a strong correlation between arousal and the levels, but no clear relation between valence and level difficulty was observed*. Further, we created promising *methods to predict emotion states during gameplay*.

The usefulness of these devices can be seen in many aspects. Their application can reduce costs, enhance accessibility, improve the quality of life, provide suitable adaptive care for the caregiver and clinicians, and provide an alternative system of evaluation and enable automated assisted living systems. Prior research into subject and clinician preferences for wearable sensor systems suggested that they should be compact, simple to operate and maintain, and should not affect daily behaviour. JINS MEME’s convenient packaging of multiple sensors – three EOG electrodes, an accelerometer and a gyroscope – within the frame of traditional eyeglasses, which the majority of older adults already wear, makes it a promising and practicable device in this field of research. On the other hand, the OpenBCI is not a cosmetic and compact device, but it can be useful to record brain signals for the evaluation of different factors which change as the patient progresses (such as the emotional state), and grow from day to day. Our approach could help to test the usefulness of other daily activities, to determine their validity and potential as tools for assessment in rehabilitation for people with cognitive impairment.

4.2 Future Work

Future research shall focus on the creation of an on-line framework that is able to predict activity recognition and emotion recognition at the same time, for individuals with mild cognitive impairment using data collected from wearable sensor in the Smart Condo™. In order to create an on-line system it is necessary to conduct a follow-up experiment that includes a period of training activities, in which participants will perform the protocol activities for short but equal amounts of time. This would mimic the user-specific training that may be necessary for the setup of the proposed method in living environments, remove the problem of imbalanced classes, and allow the current method, including the application of SMOTE, to be validated in classifying the protocol activities based on the training data only. In the case of emotion recognition it is necessary to conduct a follow-up experiment that includes the use of more biosignals such as electromyography (EMG), electrocardiogram (ECG), galvanic skin response (GSR), and the performance of the machine learning classifiers. Furthermore, a deep learning approach and transfer learning shall be used to pre-train a more complex emotional state classifier. Additionally, the system shall be evaluated in different activities such as watching videos, playing video games, reading books, and listening

to music. To study the classification of more contextually realistic activities and emotions, and to further investigate the degree of user acceptance of the device, future research may be conducted with older adults. The long-term contribution of this work will be to integrate multiple wearable sensors to enhance health monitoring systems in the Smart Condo™.

Bibliography

- [1] N. M. Boers, D. Chodos, P. Gburzynski, L. Guirguis, J. Huang, R. Lederer, L. Liu, I. Nikolaidis, C. Sadowski, and E. Stroulia, “The Smart Condo Project: Services for Independent Living,” in *E-Health, assistive technologies and applications for assisted living: Challenges and solutions*. IGI Global, 2011, pp. 289–314.
- [2] JINS. (2017) JINS MEME Academic. [Online]. Available: <https://jins.meme.com/en/academic>
- [3] OpenBCI. (2019) OpenBCI Cython Board. [Online]. Available: <https://https://openbci.com/>
- [4] D. Watson, L. A. Clark, and A. Tellegen, “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales.” *Journal of personality and social psychology*, vol. 54, no. 6, p. 1063, 1988.
- [5] U. DESA, “United Nations Department of Economic and Social Affairs/Population Division (2009b): World Population Prospects: The 2008 Revision,” *Internet: http://esa.un.org/unpp (gelesen am 16, 2010)*.
- [6] V. Ganev, D. Chodos, I. Nikolaidis, and E. Stroulia, “The smart condo: Integrating sensor networks and virtual worlds,” in *Proceedings of the 2nd Workshop on Software Engineering for Sensor Network Applications*, ser. SESENA '11. New York, NY, USA: ACM, 2011, pp. 49–54. [Online]. Available: <http://doi.acm.org/10.1145/1988051.1988061>
- [7] E. Stroulia, D. Chodos, N. M. Boers, Jianzhao Huang, P. Gburzynski, and I. Nikolaidis, “Software Engineering for Health Education and Care Delivery Systems: The Smart Condo Project,” in *2009 ICSE Workshop on Software Engineering in Health Care*, May 2009, pp. 20–28.
- [8] P. Mohebbi, E. Stroulia, and I. Nikolaidis, “Indoor Localization: A Cost-Effectiveness vs. Accuracy Study,” in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, June 2017, pp. 552–557.

- [9] D. Hickam, W. J.-M. Guise, J. B. M. G. N. Wasson, and S. , “Outpatient Case Management for Adults with Medical Illness and Complex Care Needs: A Comparative Effectiveness Review,” 01 2013.
- [10] S. J. Preece, J. Y. Goulermas, L. P. Kenney, D. Howard, K. Meijer, and R. Crompton, “Activity Identification Using Body-Mounted Sensors—A Review of Classification Techniques,” *Physiological measurement*, vol. 30, no. 4, p. R1, 2009.
- [11] D. Díaz, N. Yee, C. Daum, E. Stroulia, and L. Liu, “Activity Classification in Independent Living Environment with JINS MEME eyewear,” in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2018, pp. 1–9.
- [12] A. Cruz, D. Garcia, G. Pires, and U. Nunes, “Facial Expression Recognition Based on EOG Toward Emotion Detection for Human-Robot Interaction,” in *Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4*, ser. BIOSTEC 2015. Portugal: SCITEPRESS - Science and Technology Publications, Lda, 2015, pp. 31–37. [Online]. Available: <https://doi.org/10.5220/0005187200310037>
- [13] B. Wallace, F. Knoefel, R. Goubran, P. Masson, A. Baker, B. Allard, V. Guana, and E. Stroulia, “Detecting Cognitive Ability Changes in Patients With Moderate Dementia Using a Modified Whack-a-Mole Game,” *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 7, pp. 1521–1534, July 2018.
- [14] T. Tong, M. Chignell, M. C. Tierney, and J. Lee, “A Serious Game for Clinical Assessment of Cognitive Status: Validation Study,” *JMIR Serious Games*, vol. 4, p. e7, 05 2016.
- [15] T. J. Mattimore, N. S. Wenger, N. A. Desbiens, J. M. Teno, M. B. Hamel, H. Liu, R. Califf, A. F. Connors Jr, J. Lynn, and R. K. Oye, “Surrogate and Physician Understanding of Patients’ Preferences for Living Permanently in a Nursing Home,” *Journal of the American Geriatrics Society*, vol. 45, no. 7, pp. 818–824, 1997.
- [16] G. Demiris and B. K. Hensel, “Technologies for an Aging Society: a Systematic Review of “Smart Home” Applications,” *Yearbook of medical informatics*, vol. 17, no. 01, pp. 33–40, 2008.
- [17] A. Bulling and D. Roggen, “Recognition of Visual Memory Recall Processes Using Eye Movement Analysis,” in *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 2011, pp. 455–464.
- [18] A. Fleury, M. Vacher, and N. Noury, “Svm-Based Multimodal Classification of Activities of Daily Living in Health Smart Homes: Sensors, Algorithms, and First Experimental Results,”

- IEEE transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 274–283, 2009.
- [19] Y.-J. Hong, I.-J. Kim, S. C. Ahn, and H.-G. Kim, “Mobile Health Monitoring System Based on Activity Recognition Using Accelerometer,” *Simulation Modelling Practice and Theory*, vol. 18, no. 4, pp. 446–455, 2010.
- [20] N. Roy, A. Misra, and D. Cook, “Infrastructure-Assisted Smartphone-Based Activity Recognition in Multi-Inhabitant Smart Environments,” in *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2013, pp. 38–46.
- [21] T. G. Pavey, N. D. Gilson, S. R. Gomersall, B. Clark, and S. G. Trost, “Field Evaluation of a Random Forest Activity Classifier for Wrist-Worn Accelerometer Data,” *Journal of science and medicine in sport*, vol. 20, no. 1, pp. 75–80, 2017.
- [22] U. Maurer, A. Rowe, A. Smailagic, and D. Siewiorek, “Location and Activity Recognition Using eWatch: A Wearable Sensor Platform,” in *Ambient Intelligence in Everyday Life*. Springer, 2006, pp. 86–102.
- [23] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, “Elderly Activities Recognition and Classification for Applications in Assisted Living,” *Expert Systems with Applications*, vol. 40, no. 5, pp. 1662–1674, 2013.
- [24] D. Maguire and R. Frisby, “Comparison of Feature Classification Algorithm for Activity Recognition Based on Accelerometer and Heart Rate Data,” 2009.
- [25] K. Zhan, S. Faux, and F. Ramos, “Multi-Scale Conditional Random Fields for First-Person Activity Recognition,” in *2014 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2014, pp. 51–59.
- [26] T. Szttyler and H. Stuckenschmidt, “Online Personalization of Cross-Subjects Based Activity Recognition Models on Wearable Devices,” in *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2017, pp. 180–189.
- [27] S. Kanoh, S. Ichi-nohe, S. Shioya, K. Inoue, and R. Kawashima, “Development of an Eyewear to Measure Eye and Body Movements,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 2267–2270.
- [28] M. Desai, L. Pratt, H. Lentzner, and K. Robinson, “Trends in Vision and Hearing Among Older Americans (Aging Trends No. 2),” *Hyattsville, MD: National Center for Health Statistics*, 2001.

- [29] S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling, "In the Blink of an Eye: Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass," in *Proceedings of the 5th augmented human international conference*. ACM, 2014, p. 15.
- [30] J. Bergmann and A. McGregor, "Body-Worn Sensor Design: What Do Patients and Clinicians Want?" *Annals of biomedical engineering*, vol. 39, no. 9, pp. 2299–2312, 2011.
- [31] M. Kheirkhahan, S. Mehta, M. Nath, A. A. Wanigatunga, D. B. Corbett, T. M. Manini, and S. Ranka, "A Bag-of-Words Approach for Assessing Activities of Daily Living Using Wrist Accelerometer data," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2017, pp. 678–685.
- [32] M. A. Ayu, S. A. Ismail, A. F. A. Matin, and T. Mantoro, "A Comparison Study of Classifier Algorithms For Mobile-Phone's Accelerometer Based Activity Recognition," *Procedia Engineering*, vol. 41, pp. 224–229, 2012.
- [33] S. L. Lau and K. David, "Movement Recognition Using the Accelerometer in Smartphones," in *2010 Future Network & Mobile Summit*. IEEE, 2010, pp. 1–9.
- [34] Z. Prekopcsák, S. Soha, T. Henk, and C. Gáspár-Papanek, "Activity Recognition for Personal Time Management," in *European Conference on Ambient Intelligence*. Springer, 2009, pp. 55–59.
- [35] S. Das, L. Green, B. Perez, M. Murphy, and A. Perring, "Detecting User Activities Using the Accelerometer on (a)ndroid Smartphones," *TRUST REU The Team for Research in Ubiquitous Secure Technology*, vol. 29, 2010.
- [36] S. Ishimaru, K. Kunze, K. Tanaka, Y. Uema, K. Kise, and M. Inami, "Smart Eyewear for Interaction and Activity Recognition," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2015, pp. 307–310.
- [37] S. Niwa, M. Yuki, T. Noro, S. Shioya, and K. Inoue, "A Wearable Device for Traffic Safety-A Study on Estimating Drowsiness with Eyewear, JINS MEME," SAE Technical Paper, Tech. Rep., 2016.
- [38] T. Ogawa, M. Takahashi, and R. Kawashima, "Human Cognitive Control Mode Estimation Using JINS MEME," *IFAC-PapersOnLine*, vol. 49, no. 19, pp. 331–336, 2016.
- [39] F. I. Mahoney and D. W. Barthel, "Functional Evaluation: the Barthel Index: A Simple Index of Independence Useful in Scoring Improvement in the Rehabilitation of the Chronically Ill." *Maryland state medical journal*, 1965.

-
- [40] E. Munguia Tapia, "Using Machine Learning for Real-Time Activity Recognition and Estimation of Energy Expenditure," Ph.D. dissertation, Massachusetts Institute of Technology, 2008.
- [41] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition Using Wearable Sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.
- [42] X. Long, B. Yin, and R. M. Aarts, "Single-Accelerometer-Based Daily Physical Activity Classification," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 6107–6110.
- [43] S. Liu, R. Gao, and P. Freedson, "Computational Methods for Estimating Energy Expenditure in Human Physical Activities," *Medicine and science in sports and exercise*, vol. 44, no. 11, p. 2138, 2012.
- [44] M. Toivanen, K. Pettersson, and K. Lukander, "A Probabilistic Real-Time Algorithm for Detecting Blinks, Saccades, and Fixations from EOG Data," *Journal of Eye Movement Research*, vol. 8, no. 2, pp. 1–14, 2015.
- [45] G. M. Weiss, "Mining with Rarity: A Unifying Framework," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 7–19, 2004.
- [46] G. M. Weiss and F. Provost, "Learning when Training Data are Costly: The Effect of Class Distribution on Tree Induction," *Journal of artificial intelligence research*, vol. 19, pp. 315–354, 2003.
- [47] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [48] H. Kuehne, A. Arslan, and T. Serre, "The Language of Actions: Recovering the Syntax and Semantics of Goal-Directed Human Activities," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 780–787.
- [49] B. N. Klein, S. L. Lau, and K. David, "Evaluation of the Influence of Time Synchronisation on Classification Learning Based Movement Detection with Accelerometers," in *2011 IEEE/IPSJ International Symposium on Applications and the Internet*. IEEE, 2011, pp. 56–64.
- [50] Y.-Y. Song and L. Ying, "Decision Tree Methods: Applications for Classification and Prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015.

- [51] E. Frank and I. H. Witten, "Generating Accurate Rule Sets Without Global Optimization," in *Proceedings of the Fifteenth International Conference on Machine Learning*, ser. ICML '98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 144–151. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645527.657305>
- [52] C. P. John, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," *MSRTR: Microsoft Research*, vol. 3, no. 1, pp. 88–95, 1998.
- [53] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [54] A. A. Motsinger and M. D. Ritchie, "The Effect of Reduction in Cross-Validation Intervals on the Performance of Multifactor Dimensionality Reduction," *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, vol. 30, no. 6, pp. 546–555, 2006.
- [55] G. Andersson, J. Hagman, R. Talianzadeh, A. Svedberg, and H. C. Larsen, "Effect of Cognitive Load on Postural Control," *Brain Research Bulletin*, vol. 58, no. 1, pp. 135 – 139, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0361923002007700>
- [56] G. L. Pellecchia, "Postural Sway Increases with Attentional Demands of Concurrent Cognitive Task," *Gait & Posture*, vol. 18, no. 1, pp. 29 – 34, 2003. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0966636202001388>
- [57] S. Willis, S. L. Tennstedt, M. Marsiske, K. Ball, J. Elias, K. Mann Koepke, J. Morris, G. Rebok, F. W. Unverzagt, A. M. Stoddard, and E. Wright, "Long-Term Effects of Cognitive Training on Everyday Functional Outcomes in Older Adults," *JAMA : the journal of the American Medical Association*, vol. 296, pp. 2805–14, 01 2007.
- [58] J. Anguera, J. Boccanfuso, J. Rintoul, O. Claffin, F. Faraji, J. Janowich, E. Kong, Y. Larraburo, C. Rolle, E. Johnston, and A. Gazzaley, "Video Game Training Enhances Cognitive Control in Older Adults," *Nature*, vol. 501, pp. 97–101, 09 2013.
- [59] R. Nouchi, Y. Taki, H. Takeuchi, H. Hashizume, T. Nozawa, T. Kambara, A. Sekiguchi, C. Miyauchi, Y. Kotozaki, H. Nouchi, and R. Kawashima, "Brain Training Game Boosts Executive Functions, Working Memory and Processing Speed in the Young Adults: A Randomized Controlled Trial," *PloS one*, vol. 8, p. e55518, 02 2013.
- [60] P. Toril, J. Reales, and S. Ballesteros, "Video Game Training Enhances Cognition of Older Adults: A Meta-Analytic Study," *Psychology and Aging*, vol. 29, 09 2014.

- [61] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 12 1980.
- [62] K. E. Hutchison, R. P. Trombley, F. L. Collins, D. W. McNeil, C. L. Turk, L. E. Carter, B. J. Ries, and M. J. Leftwich, "A Comparison of Two Models of Emotion: Can Measurement of Emotion Based on One Model be Used to make Inferences about the other?" *Personality and Individual Differences*, vol. 21, no. 5, pp. 785 – 789, 1996. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0191886996001079>
- [63] D. Watson and A. Tellegen, "Toward a Consensual Structure of Mood," *Psychological bulletin*, vol. 98, pp. 219–35, 10 1985.
- [64] S. von Humboldt, A. Monteiro, and I. Leal, "Validation of the PANAS: A Measure of Positive and Negative Affect for Use with Cross-National Older Adults," *Review of European Studies*, vol. 9, p. 10, 03 2017.
- [65] A. Mackinnon, A. Jorm, H. Christensen, A. Korten, P. Jacomb, and B. Rodgers, "A Short Form of the Positive and Negative Affect Scale: Evaluation of Factorial Validity and Invariance Across Demographic Variables in a Community Sample," *Personality and Individual Differences*, pp. 405 – 416, 1999.
- [66] Y. Wu, Y. Wei, and J. Tudor, "A Real-Time Wearable Emotion Detection Headband Based on EEG Measurement," *Sensors and Actuators A: Physical*, vol. 263, pp. 614 – 621, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S092442471631192X>
- [67] W. Liu, W.-L. Zheng, and B.-L. Lu, "Emotion Recognition Using Multimodal Deep Learning," in *Neural Information Processing*, A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, and D. Liu, Eds. Cham: Springer International Publishing, 2016, pp. 521–529.
- [68] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams, "A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures," *Frontiers in Psychology*, vol. 8, p. 2116, 2017. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpsyg.2017.02116>
- [69] M. B. H. Wiem and Z. Lachiri, "Emotion Classification in Arousal Valence Model using MAHNOB-HCI Database," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 3, 2017. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2017.080344>

- [70] X. Liu, Q. Wang, D. Liu, Y. Wang, Y. Zhang, O. Bai, and J. Sun, "Human Emotion Classification Based on Multiple Physiological Signals by Wearable System," in *Technology and health care : official journal of the European Society for Engineering and Medicine*, 2018.
- [71] Y. Wang, Z. Lv, and Y. Zheng, "Automatic Emotion Perception Using Eye Movement Information for E-Healthcare Systems," *Sensors*, vol. 18, no. 9, 2018. [Online]. Available: <http://www.mdpi.com/1424-8220/18/9/2826>
- [72] S.-H. Wang, H.-T. Li, E.-J. Chang, and A.-Y. A. Wu, "Entropy-Assisted Emotion Recognition of Valence and Arousal Using XGBoost Classifier," in *Artificial Intelligence Applications and Innovations*, L. Iliadis, I. Maglogiannis, and V. Plagianakos, Eds. Cham: Springer International Publishing, 2018, pp. 249–260.
- [73] J. C. Quiroz, M. H. Yong, and E. Geangu, "Emotion-Recognition Using Smart Watch Accelerometer Data: Preliminary Findings," *CoRR*, vol. abs/1709.09148, 2017. [Online]. Available: <http://arxiv.org/abs/1709.09148>
- [74] M. Roozeboom, G. Visschedijk, and E. Oprins, "The Effectiveness of Three Serious Games Measuring Generic Learning Features," *British Journal of Educational Technology*, vol. 48, 10 2015.
- [75] W. Yang, M. Rifqi, C. Marsala, and A. Pinna, "Physiological-Based Emotion Detection and Recognition in a Video Game Context," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, Rio, Brazil, Jul. 2018, pp. 194–201. [Online]. Available: <https://hal.sorbonne-universite.fr/hal-01784795>
- [76] A. Alhargan, N. Cooke, and T. Binjammaz, "Affect Recognition in an Interactive Gaming Environment Using Eye Tracking," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, Oct 2017, pp. 285–291.
- [77] S. Huynh, Y. Lee, T. Park, and R. K. Balan, "Jasper: Sensing Gamers' Emotions Using Physiological Sensors," in *Proceedings of the 3rd Workshop on Mobile Gaming*, ser. MobiGames '16. New York, NY, USA: ACM, 2016, pp. 1–6. [Online]. Available: <http://doi.acm.org/10.1145/2934646.2934648>
- [78] Y. Liu, Z. Teng, H. Lan, X. Zhang, and D. Yao, "Short-Term Effects of Prosocial Video Games on Aggression: An Event-Related Potential Study," *Frontiers in Behavioral Neuroscience*, vol. 9, p. 193, 2015. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnbeh.2015.00193>

- [79] S. Chen, Z. Gao, and S. Wang, "Emotion Recognition from Peripheral Physiological Signals Enhanced by EEG," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 2827–2831.
- [80] N. Jadhav, R. Manthalkar, and Y. Joshi, "Electroencephalography-Based Emotion Recognition Using Gray-Level Co-Occurrence Matrix Features," in *Proceedings of International Conference on Computer Vision and Image Processing*, B. Raman, S. Kumar, P. P. Roy, and D. Sen, Eds. Singapore: Springer Singapore, 2017, pp. 335–343.
- [81] C. Setz, J. Schumm, C. Lorenz, B. Arnrich, and G. Tröster, "Using Ensemble Classifier Systems for Handling Missing Data in Emotion Recognition from Physiology: One Step Towards a Practical System," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, 2009, pp. 1–8.
- [82] S. Gharsalli, B. Emile, H. Laurent, X. Desquesnes, and D. Vivet, "Random Forest-Based Feature Selection for Emotion Recognition," in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Nov 2015, pp. 268–272.
- [83] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. " O'Reilly Media, Inc.", 2017.
- [84] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. Jerčić, "Evaluating Classifiers for Emotion Recognition Using EEG," in *International conference on augmented cognition*. Springer, 2013, pp. 492–501.
- [85] X. Liu, Q. Wang, D. Liu, Y. Wang, Y. Zhang, O. Bai, and J. Sun, "Human Emotion Classification Based on Multiple Physiological Signals by Wearable System," *Technology and Health Care*, vol. 1, no. Preprint, pp. 1–11, 2018.
- [86] J.-M. López-Gil, J. Virgili-Gomá, R. Gil, T. Guilera, I. Batalla, J. Soler-González, and R. García, "Method for Improving EEG Based Emotion Recognition by Combining it with Synchronized Biometric and Eye Tracking Technologies in a Non-Invasive and Low Cost Way," *Frontiers in Computational Neuroscience*, vol. 10, p. 85, 2016. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fncom.2016.00085>
- [87] D.-W. Chen, R. Miao, W.-Q. Yang, Y. Liang, H.-H. Chen, L. Huang, C.-J. Deng, and N. Han, "A Feature Extraction Method Based on Differential Entropy and Linear Discriminant Analysis for Emotion Recognition," *Sensors*, vol. 19, no. 7, 2019. [Online]. Available: <http://www.mdpi.com/1424-8220/19/7/1631>

- [88] B. Schuller, M. Lang, and G. Rigoll, "Robust Acoustic Speech Emotion Recognition by Ensembles of Classifiers," in *Tagungsband Fortschritte der Akustik-DAGA# 05, München*, 2005.
- [89] R. Blagus and L. Lusa, "SMOTE for High-Dimensional Class-Imbalanced Data," *BMC Bioinformatics*, vol. 14, no. 1, p. 106, Mar 2013. [Online]. Available: <https://doi.org/10.1186/1471-2105-14-106>

Appendix Activity Detection

A.1 Activity Classification

Activity Mode 1 – Literal Activities

This activity mode includes labels for every activity explicitly performed by the participant, including all activities outlined in the protocol. See Table A.1 for a list of all literal activities.

Activity Mode 2 – General Activities

This activity mode includes labels for every activity explicitly performed by the participant, translating some of them into the corresponding Basic Activity of Daily Living (BADL) on the Barthel Index [39] while grouping some of the Instrumental Activities of Daily Living (IADLs) into more broad categories. This activity mode was included in order to define a level of detail which might be more reasonably expected to yield accurate activity classification results than the literal activities in activity mode 1, but which would still contain the same amount of useful information for a caretaker, clinician or assisted living system. See Table A.1 for a list of all general activities.

Activity Mode 3 – Posture and Motion

This activity mode identifies the posture of the participant as either lying down, sitting, standing or walking and indicates whether the participant is either in motion or relatively stationary. This concept is similar to the body state recognition that has been performed in past work involving activity recognition [19]. See Table A.1 for a list of posture and motion labels.

Each participant performs a set of tasks outlined in a protocol over a period of approximately 1.5 - 2.5 hours. The participant performs these tasks while wearing the JINS MEME glasses, which continuously transmit data via Bluetooth to a computer located within the Smart Condo™. No participant-specific calibration or adjustments were made to the JINS MEME glasses.

A.1.1 Manual Label-Activity Mode

Based on video recordings collected from cameras embedded in the ceiling of the Smart Condo™, the actions of each participant are manually time-stamped. Three different activity modes are defined for use in analysis: See table A.2

Literal	General
Washing hands	Grooming (Barthel Index)
Toilet use	Toilet use (Barthel Index)
Eating meal	Feeding (Barthel Index)
Taking medication	
Walking	Mobility (Barthel Index)
Dressing	Dressing (Barthel Index)
Bathing	Bathing (Barthel Index)
Virtual Gym	Stretching/Exercise
Questionnaire	Typing/writing
Preparing coffee	Cooking
Making eggs and toast	
Setting table	Housework - Low intensity
Cleaning dishes	
Sweeping	Housework - High intensity
Laundry	
Ironing shirt	
Tablet games	
Watching TV	Still sitting

Figure A.1: List of Activities

Posture / Motion
Lying down - Low Motion
Lying down - High Motion
Sitting - Low Motion
Sitting - High Motion
Standing - Low Motion
Standing - High Motion
Walking - <u>Low</u> Motion
Walking - High Motion

Figure A.2: List of Posture and Motion Characteristics

A.2 Classifies models – Parameters

IBK's Scheme: Weka.classifiers.lazy.IBk -K 3 -W 0 -X -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last" **Ramdon Forest's Scheme:** Weka.classifiers.trees.RandomForest -P 100 -O -attribute-importance -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 -B

SMO's Scheme: Weka.classifiers.functions.SMO -C 2.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"

C45's Scheme: Weka.classifiers.trees.C45 -R -N 3 -Q 1 -B -M 2 -A

A.3 Data Analysis Framework

Figures A1, A2 and A3 show the activity classification rates by activity under each activity modes 1, 2 and 3, respectively. Although the overall classification rate varies between activity modes, the Random Forest classifier has the highest PRC area and recall rate in each activity mode. This indicates that the Random Forest classifier may be the most For further results, only the Random Forest classifier will be used.

A.4 Data-Stream Synchronization - MATLAB

The data JINS MEME glasses includes a timestamp for each data point. Although the 100 Hz sample rate of the device is known, the timestamp was used to synchronize the data with the video collected from the Smart Condo™ since it was found that certain activities and certain locations

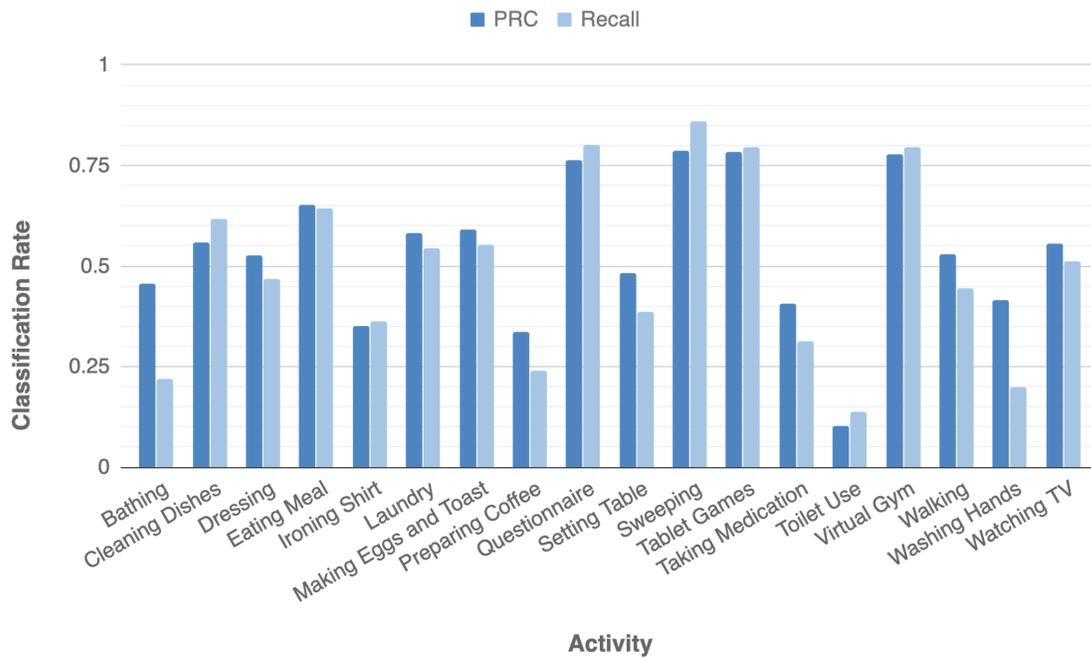


Figure A.3: Mean activity classification rates by activity (literal activities). This Figure shows the activity classification accuracies by activity under activity mode 1 (literal activities). The mean recall rate and area under the PRC curve are shown across all classifiers for each activity. This graph gives a general overview of which activities within the protocol performed by the participants are reasonably able to be classified using data from the JINS MEME glasses and machine learning algorithms.

caused the device to lose connection with the computer receiving the data, resulting in a skipping or loss of data.

For times when participants were performing activities, MATLAB was used to calculate 214 attributes using EOG, Accelerometer, and Gyroscope data. These attributes include information about eye movements and head movements, as well as attributes calculated from the raw signals based on a set of attributes used for activity recognition with accelerometers in past research [40]. This set of raw signal-derived attributes includes those calculated from the time domain as well as the frequency domain. For accelerometer data, previous research has suggested a relationship between the mean of the signal in the time domain and the intensity of movement [42]. It has also been shown that activities with a similar energy intensity can be identified in the frequency domain by the period of the signal [43]. While past research only used this set of raw signal-derived attributes on accelerometer data, the same attributes were calculated in the present study using EOG data, accelerometer data, and angular velocity of the head along Euler axes as determined

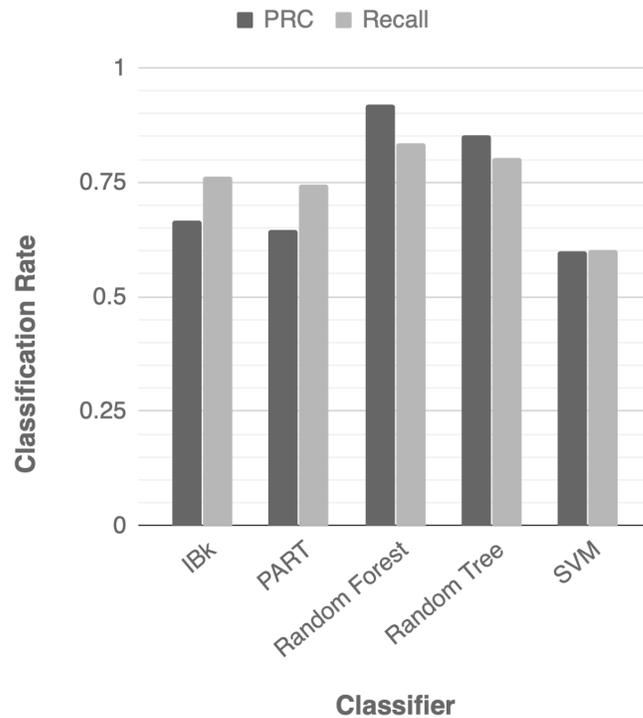


Figure A.4: Mean activity classification rates by classifier (literal activities), all classifiers + activities.

from accelerometer and gyroscope data. Angular position along Euler axes could not be accurately determined without drift due to the exclusion of a magnetometer in the JINS MEME glasses. A complete list of attributes used in the present study are shown in **List A.5** (Appendix A).

Past research determined that a window length of 5.6 seconds was most appropriate for activity recognition in living environments, as it is either optimal or near optimal for many attributes and also allows for posture recognition [40]. For real-time applications, a window length of 5.6 seconds is also short enough to allow fast interventions at the point of decision that could be triggered almost as soon as the activity of interest is recognized. Tapia *et al.* [40] also concede that a disadvantage of a short window duration is expected lower performance on activities with high motion variability since these activities can be performed in different ways depending on the situation.

The 214 attributes were calculated for each activity over a sliding 5.6 second-long window that moved incrementally by 1 second in order to capture information within the duration of the overlap. Due to the missing or skipped data in certain activities, windows containing at least 60% of the expected data (according to the sample rate) were interpolated to contain 560 (5.6 s x 100 Hz) data points. Low-pass filters and bandpass filters were used in the calculation of many of the attributes. Low-pass filters eliminate most of the signal generated by dynamic human motion while

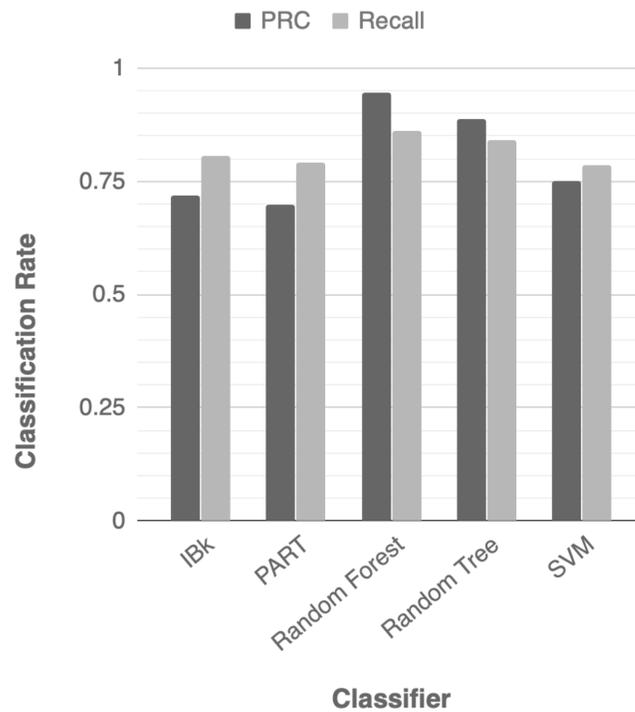


Figure A.5: Mean activity classification rates by activity (general activities), all classifiers + activities.

preserving the information generated by static human motion or posture information [40]. Bandpass filters eliminate the static signal components containing reference information. For accelerometers, bandpass filters remove DC components with posture information about the orientation of the sensor [40]. Applying these filters allows for clarity in the attributes calculated for activity classification

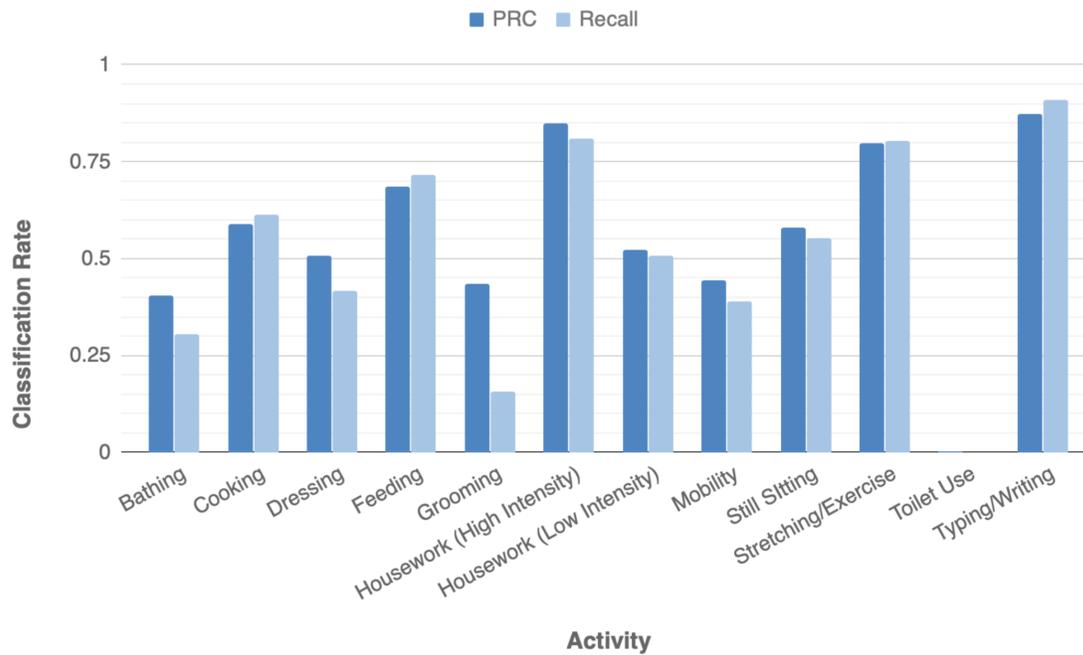


Figure A.6: Mean Activity Classification Rates by Activity (Posture and Motion), All Classifiers + Activities, this figure shows the activity classification accuracies by activity under activity mode 3 (posture and motion). The labels of Sitting (High Motion) and Walking (Low Motion) had the lowest classification accuracies. For both of these posture and motion labels, this could be attributed to a lack of data due to the fact that participants did not spend much time doing activities with motion while in a sitting position. For Walking (Low Motion), the low classification accuracy could also be due to inconsistent attributes caused by inconsistent walking styles at this low motion state, as well as the short amounts of time that participants walked at a time

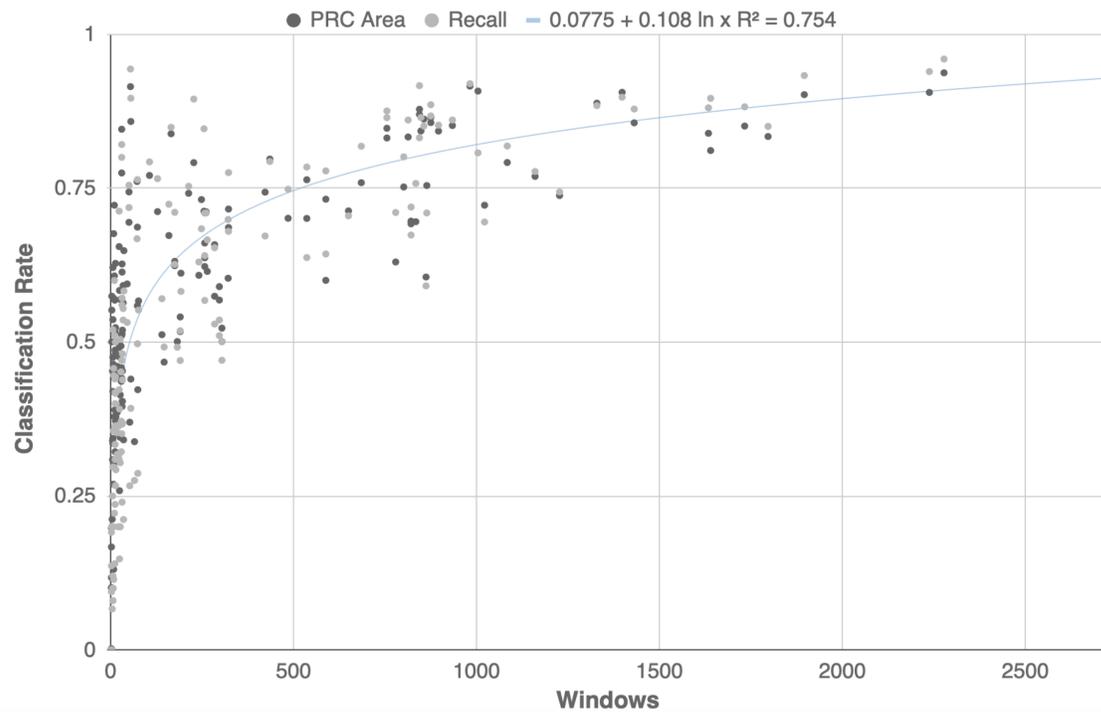


Figure A.7: activity classification rate and available data. All activities are plotted for each participant, including the three modes of activity definition. Trendline is determined from Recall values only. PRC Area data is included to emphasize the trend.

A.5 Extracted Attributes

Uses raw EOG data

- 1 Ratio of “good” data to all data
- 2 Number of blinks
- 3 Mean peak amplitude of blinks
- 4 Mean length of blinks
- 5 Number of saccades
- 6 Mean peak amplitude of saccades
- 7 Mean length of saccades

A.5.1 EOG Attributes

Uses raw EOG data

- 8 Mean EOG (Reference)
- 9 Mean EOG (Horizontal)
- 10 Mean EOG (Vertical)
- 11 Mean EOG (All)
- 12 Area under EOG (Reference)
- 13 Area under EOG (Horizontal)
- 14 Area under EOG (Vertical)
- 15 Mean distance (Reference - Horizontal)
- 16 Mean distance (Reference - Vertical)
- 17 Mean distance (Horizontal - Vertical)
- 18 Mean EOG absolute (Reference)
- 19 Mean EOG absolute (Horizontal)
- 20 Mean EOG absolute (Vertical)
- 21 Cumulative sum absolute (Reference)
- 22 Cumulative sum absolute (Horizontal)
- 23 Cumulative sum absolute (Vertical)
- 24 Cumulative sum absolute (All)
- 25 Signal vector magnitude (All)
- 26 Entropy (All)
- 27 Skewness (Reference)
- 28 Skewness (Horizontal)
- 29 Skewness (Vertical)
- 30 Kurtosis (Reference)
- 31 Kurtosis (Horizontal)
- 32 Kurtosis (Vertical)
- 33 Quartile 1 (Reference)
- 34 Quartile 2 (Reference)
- 35 Quartile 3 (Reference)
- 36 Quartile 1 (Horizontal)
- 37 Quartile 2 (Horizontal)
- 38 Quartile 3 (Horizontal)
- 39 Quartile 1 (Vertical)
- 40 Quartile 2 (Vertical)
- 41 Quartile 3 (Vertical)
- 42 Variance (Reference)
- 43 Variance (Horizontal)
- 44 Variance (Vertical)
- 45 Coefficient of variation absolute (Reference)
- 46 Coefficient of variation absolute (Horizontal)
- 47 Coefficient of variation absolute (Vertical)
- 48 Interquartile Range (Reference)
- 49 Interquartile Range (Horizontal)
- 50 Interquartile Range (Vertical)
- 51 Interquartile Range of maximum signal amplitude (Reference)
- 52 Range of maximum signal amplitude (Horizontal)
- 53 Range of maximum signal amplitude (Vertical)
- 54 Power Spectral Density Estimate (Reference)
- 55 Power Spectral Density Estimate (Horizontal)
- 56 Power Spectral Density Estimate (Vertical)
- 57 Power of frequency peak (Reference)
- 58 Power of frequency peak (Horizontal)

59	Power of frequency peak (Vertical)	87	Cumulative sum absolute (X)
60	Power of frequency peak (All)	88	Cumulative sum absolute (Y)
61	Frequency peak (All)	89	Cumulative sum absolute (Z)
62	Total energy (Reference)	90	Cumulative sum absolute (All)
63	Total energy (Horizontal)	91	Signal vector magnitude (All)
64	Total energy (Vertical)	92	Entropy (All)
65	Activity band energy (Reference)	93	Skewness (X)
66	Activity band energy (Horizontal)	94	Skewness (Y)
67	Activity band energy (Vertical)	95	Skewness (Z)
68	Energy of low intensity activity (Reference)	96	Kurtosis (X)
69	Energy of low intensity activity (Horizontal)	97	Kurtosis (Y)
70	Energy of low intensity activity (Vertical)	98	Kurtosis (Z)
71	Energy of moderate intensity activity (Reference)	99	Quartile 1 (X)
72	Energy of moderate intensity activity (Horizontal)	100	Quartile 2 (X)
73	Energy of moderate intensity activity (Vertical)	101	Quartile 3 (X)
		102	Quartile 1 (Y)
		103	Quartile 2 (Y)
		104	Quartile 3 (Y)
		105	Quartile 1 (Z)
		106	Quartile 2 (Z)
		107	Quartile 3 (Z)
		108	Variance (X)
		109	Variance (Y)
		110	Variance (Z)
		111	Coefficient of variation absolute (X)
		112	Coefficient of variation absolute (Y)
		113	Coefficient of variation absolute (Z)
		114	Interquartile Range (X)
		115	Interquartile Range (Y)
		116	Interquartile Range (Z)
		117	Range of maximum signal amplitude (X)
		118	Range of maximum signal amplitude (Y)
		120	Power Spectral Density Estimate (X)
		121	Power Spectral Density Estimate (Y)
		122	Power Spectral Density Estimate (Z)
		123	Power of frequency peak (X)

A.5.2 Accelerometer Attributes

Uses raw Accelerometer data

74	Mean accelerometer (X)
75	Mean accelerometer (Y)
76	Mean accelerometer (Z)
77	Mean accelerometer (All)
78	Area under accelerometer (X)
79	Area under accelerometer (Y)
80	Area under accelerometer (Z)
81	Mean distance (X - Y)
82	Mean distance (X - Z)
83	Mean distance (Y - Z)
84	Mean accelerometer absolute (X)
85	Mean accelerometer absolute (Y)
86	Mean accelerometer absolute (Z)

- 124 Power of frequency peak (Y)
- 125 Power of frequency peak (Z)
- 126 Power of frequency peak (All)
- 127 Frequency peak (All)
- 128 Total energy (X)
- 129 Total energy (Y)
- 130 Total energy (Z)
- 131 Activity band energy (X)
- 132 Activity band energy (Y)
- 133 Activity band energy (Z)
- 134 Energy of low intensity activity (X)
- 135 Energy of low intensity activity (Y)
- 136 Energy of low intensity activity (Z)
- 137 Energy of moderate intensity activity (X)
- 138 Energy of moderate intensity activity (Y)
- 139 Energy of moderate intensity activity (Z)

A.5.3 Head Movement Features

Uses the derivative of Euler Angle data (Accelerometer and Gyroscope)

- 140 Number of head movements (Phi)
- 141 Number of head movements (Theta)
- 142 Number of head movements (Psi)
- 143 Mean peak amplitude of head movements (Phi)
- 144 Mean peak amplitude of head movements (Theta)
- 145 Mean peak amplitude of head movements (Psi)
- 146 Mean length of head movements (Phi)
- 147 Mean length of head movements (Theta)
- 148 Mean length of head movements (Psi)

A.6 Euler Angle Features

Uses the derivative of Euler Angle data (Accelerometer and Gyroscope)

- 149 Mean Euler angle (Phi)
- 150 Mean Euler angle (Theta)
- 151 Mean Euler angle (Psi)
- 152 Mean Euler angle (All)
- 153 Area under Euler angle (Phi)
- 154 Area under Euler angle (Theta)
- 155 Area under Euler angle (Psi)
- 156 Mean distance (Phi - Theta)
- 157 Mean distance (Phi - Psi)
- 158 Mean distance (Theta - Psi)
- 159 Mean Euler angle absolute (Phi)
- 160 Mean Euler angle absolute (Theta)
- 161 Mean Euler angle absolute (Psi)
- 162 Cumulative sum absolute (Phi)
- 163 Cumulative sum absolute (Theta)
- 164 Cumulative sum absolute (Psi)
- 165 Cumulative sum absolute (All)
- 166 Signal vector magnitude (All)
- 167 Entropy (All)
- 168 Skewness (Phi)
- 169 Skewness (Theta)
- 170 Skewness (Psi)
- 171 Kurtosis (Phi)
- 172 Kurtosis (Theta)
- 173 Kurtosis (Psi)
- 174 Quartile 1 (Phi)
- 175 Quartile 2 (Phi)
- 176 Quartile 3 (Phi)
- 177 Quartile 1 (Theta)
- 178 Quartile 2 (Theta)
- 179 Quartile 3 (Theta)

- 180 Quartile 1 (Psi)
- 181 Quartile 2 (Psi)
- 182 Quartile 3 (Psi)
- 183 Variance (Phi)
- 184 Variance (Theta)
- 185 Variance (Psi)
- 186 Coefficient of variation absolute (Phi)
- 187 Coefficient of variation absolute (Theta)
- 188 Coefficient of variation absolute (Psi)
- 189 Interquartile Range (Phi)
- 190 Interquartile Range (Theta)
- 191 Interquartile Range (Psi)
- 192 Range of maximum signal amplitude (Phi)
- 193 Range of maximum signal amplitude (Theta)
- 194 Range of maximum signal amplitude (Psi)
- 195 Power Spectral Density Estimate (Phi)
- 196 Power Spectral Density Estimate (Theta)
- 197 Power Spectral Density Estimate (Psi)
- 198 Power of frequency peak (Phi)
- 199 Power of frequency peak (Theta)
- 200 Power of frequency peak (Psi)
- 201 Power of frequency peak (All)
- 202 Frequency peak (All)
- 203 Total energy (Phi)
- 204 Total energy (Theta)
- 205 Total energy (Psi)
- 206 Activity band energy (Phi)
- 207 Activity band energy (Theta)
- 208 Activity band energy (Psi)
- 209 Energy of low intensity activity (Phi)
- 210 Energy of low intensity activity (Theta)
- 211 Energy of low intensity activity (Psi)
- 212 Energy of moderate intensity activity (Phi)
- 213 Energy of moderate intensity activity (Theta)
- 214 Energy of moderate intensity activity (Psi)

Appendix Questionnaires

B.1 Questionnaires

Participants wore the sensor devices as they view a random sequence of images. The images were selected from the International Affective Picture System (IAPS) and were represent different emotional states. After each image, participants were indicate their emotional response through a questionnaire (see figure B.1). The participant have then be given a version of the Whack-a-Mole game configured based on their performance in the first session.

This version of the game display only the Too Easy, Optimal and Too Difficult levels in a random sequence. Each level has a consistent length. Note that this differs from gameplay succession in the first session, where each level displayed a consistent number of targets. After the end of each level, the participant will be asked to fill out the PANAS questionnaire [4] (see in figure B.2).

PANAS

This scale consists of a number of words that describe different feelings and emotions. Read each item and then marks the appropriate answer in the space next to the word. Indicate to what extent you feel this way right now, that is, at the present moment.

1	2	3	4	5
Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
___	Interested		___	Irritable
___	Distressed		___	Alert
___	Excited		___	Ashamed
___	Upset		___	inspired
___	Strong		___	Nervous
___	Guilty		___	Determined
___	Scared		___	Attentive
___	Hostile		___	Jittery
___	Enthusiastic		___	Active
___	Proud		___	Afraid

Figure B.2: Self-assessment of PANAS gameplay experience [4]