Individual Diploma Thesis

Exploring Different Methods For Annotating Real-Time E4-Based Psychophysiological Data Using Semi-Supervised Machine-Learning

Ioulianos Christou

University Of Cyprus

Computer Science

May 2024

University of Cyprus

Department of Computer Science

EXPLORING DIFFERENT METHODS FOR ANNOTATING REAL-TIME E4-BASED PSYCHOPHYSIOLOGICAL DATA USING SEMI-SUPERVISED MACHINE-LEARNING

Ioulianos Christou

Supervisor

Chryssis Georgiou

Thesis project has been submitted for partial fulfillment of the requierements of Computer Science Degree acquisition from the University of Cyprus

May 2024

Acknowledgement

I would like to express my gratitude to my supervisor, Dr. Chryssis Georgiou, who holds the position of Professor at the Department of Computer Science at the University of Cyprus, for providing me with the opportunity to work on this topic and for offering me invaluable assistance throughout my thesis. His guidance, encouragement, and support have been instrumental in helping me accomplish my research goals, and we are deeply thankful for his unwavering dedication to my success.

I would also like to extend my sincere appreciation to the Department of Psychology at the University of Cyprus for providing us with the experimental data samples. Without their generous contribution, this research project would not have been possible, and we RE incredibly grateful for their support.

Lastly, we want to express my appreciation to Andria Trigeorgi for providing me with pertinent information related to the topic and for her continued support and feedback. Her insights and advice have been tremendously helpful throughout the research process, and we are deeply grateful for her assistance.

Abstract

The integration of wearable technologies and Artificial Intelligence (AI) offers a promising avenue for understanding and addressing problematic pain coping behaviors. This thesis examines the potential collaboration between wearable devices and AI algorithms to distinguish between effective and ineffective pain coping strategies in real-world situations.

Participants were equipped with Empatica E4 wearable devices and instructed to interact with a customized smartphone application, providing valuable insights into social context, stress or pain experiences, and coping strategies. By using a variety of psychophysiological signals such as Photoplethysmography (PPG), Electrodermal Activity (EDA), Accelerometer (ACC), and Temperature (TEMP), alongside self-reported responses, features were extracted to capture subtle physiological responses linked to different coping mechanisms.

Since there is a shortage of labeled data and in more large scale experiment may be expensive to label the data, a semi-supervised learning approach was employed, in which a limited amount of labeled instances were combined with abundant unlabeled data to improve the model's generalization. The study examined the effectiveness of semi-supervised algorithms in accurately classifying individuals into acceptance and avoidance groups, using techniques such as self-training, co-training, and label propagation.

Content

Chapter 1

Introduction

1.1 Motivation

Over the past decade, there has been a significant increase in the use of wearable technology, including smartwatches and smart bands. These devices have the ability to track various psychophysiological signals, such as heart rate and sweat gland activity. These signals, also known as psychophysiological indicators, have been demonstrated to be indicative of an individual's emotional response. Some examples of psychophysiological signals include *Electrocardiogram (ECG)* [16], *Electrodermal Activity (EDA)* [17], and facial *electromyography (fEMG)* [18].

Several previous studies [1][2][3][4][5] have analyzed such data. However, most previous works focused on signals recorded from stationary devices, with the exception of one study [5] that included measures from wearable devices. Furthermore, the only features that were examined and used to train the models were HRV time-domain features. These measures are used to quantify the amount of variation in the intervals between heartbeats over a specific period of time and are derived from ECG data.

The purpose of this study is to delve deeper into the subject matter in order to gain a better understanding of the potential uses and benefits of psychophysiological signals captured by wearable devices. The findings of this study could have significant implications for the healthcare industry and could help to improve the care and treatment of patients.

1.2 Goal of Study

This research utilizes data that was gathered during an experiment about pain management techniques, conducted by the Department of Psychology at the University of Cyprus. The primary objective of this thesis is to aid in integrating Acceptance and Commitment Therapy (ACT) [19] into daily life. ACT is a type of psychotherapy that encourages individuals to confront their thoughts and feelings rather than blaming themselves or ignoring them. This therapy is particularly useful for those struggling with anxiety, depression, and similar conditions. ACT divides individuals into two categories based on their reactions at a particular moment: the first group is referred to as "acceptance" or "functional," consisting of individuals who acknowledge their problems and attempt to deal with them head-on. The second group is the "avoidance" or "dysfunctional" group, consisting of individuals who refuse to interact with their thoughts and sensations and try to avoid them. An individual's classification changes depending on the environment and circumstances, and they may not always fall into the same category. The aim of this thesis is to effectively categorize individuals as functional or dysfunctional in terms of coping with pain.

1.3 Methodology

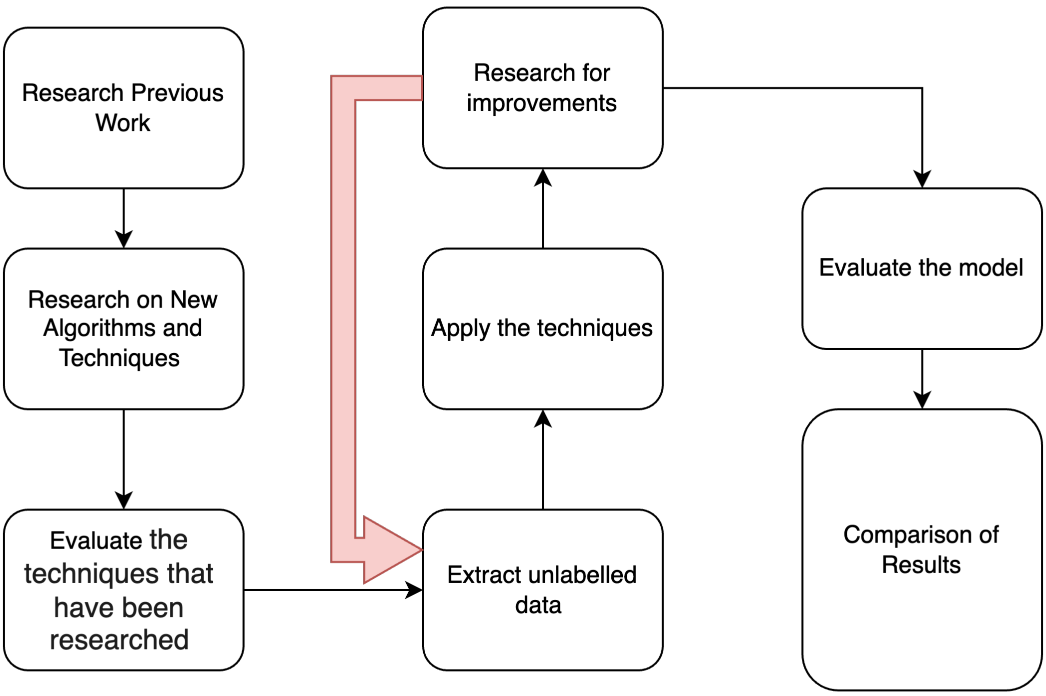

Figure 1.1 : Methodology

Firstly, the Department of Psychology of the University of Cyprus conducted an experiment in the lab, where psychophysiological signals were recorded from the Empatica E4 wearable device. These signals are *Photoplethysmography* (PPG), *Electrodermal Activity* (EDA).

As shows in the Figure 1.1 before anything, a lot of attention was given to examining the efforts of the past years. For this reason, from all the models that was used in the last thesis [5] we choose to use the *GradientBoostingClassifier* [11] as we tested it came with the most promising results.

Then with the use of previous code and techniques data was extracted and cleared. With the data ready we started researching about ways that we can get all this unlabeled data that was capture by the wearables in use. Semi supervised learning techniques [6][7][8] that will be explain later on the thesis is the way that we choose to proceed.

Finally we started a cycle of extracting unlabeled data, applying the techniques [6][7][8] and researching about room for improvements. At the end we evaluated all the models and compared the results.

1.4 Document Organization

The document is structured into four main sections, as shown in Table 1.1.

Chapters	Description
Chpter 2 : Background Knowledge	A comprehensive explanation of the algorithms, methods of evaluations, and an analysis of the device used for data collection is provided
Chapter 3 : Previous Work	Detailed review of all the relevant prior research in the field
Chapter 4 : Results	Displays various algorithm combinations and their corresponding outcomes
Chapter 5 : Summary and Future Improvements	Overview of the findings and potential avenues for further research and enhancements.

Table 1.1 : Document Organization

Chapter 2

Background knowledge

2.1 Psychophysiological Signal

Four experiments were conducted by the Department of Psychology of the University of Cyprus, which are detailed in Section 3. In this section, the psychophysiological signals that were collected are explained.

2.1.1 Electrocardiogram

The electrical signals generated by the heart during each beat can be detected noninvasively from the surface of the body using an electrocardiogram (ECG) [16]. There are three waves in the basic pattern of this electrical activity, which are referred to as P, QRS, and T, as shown in Figure 2.1. The ECG signal allows for the extraction of three groups of features, which include frequency-domain, spectral, and time-domain. However, for the purposes of this study, we are primarily focusing on time-domain features, as they are more relevant to our research, according to a previous study. Time-domain measures are primarily concerned with Heart Rate Variability (HRV), which refers to the variations in the time intervals between successive heartbeats, specifically the RR intervals. The RR intervals are the duration between two sequential R peaks in the ECG signal, with the R wave being part of the QRS complex.

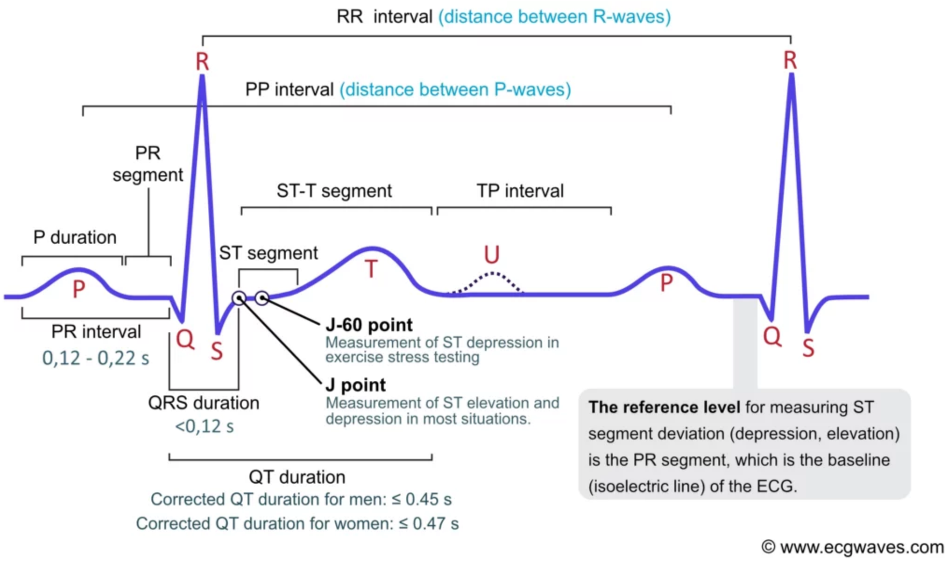

Figure 2.1 Electrocardiogram

2.1.2 Photoplethysmography

Optical sensors can detect fluctuations in blood volume throughout the body during a cardiac cycle, as blood volume changes from the start of one heartbeat to the beginning of the next. This technology, known as photoplethysmography (PPG) [17], involves the use of an LED light source and a photodetector. The LED emits light into the microvascular bed of the tissue, and the photodetector records the amount of light absorbed or reflected. Based on the amount of light absorbed or reflected, changes in blood volume can be detected as shown in the Figure 2.2 . Heart Rate Variability (HRV) can also be estimated from the photoplethysmography signal, which is equivalent to the distance between consecutive R-peaks of the ECG signal.

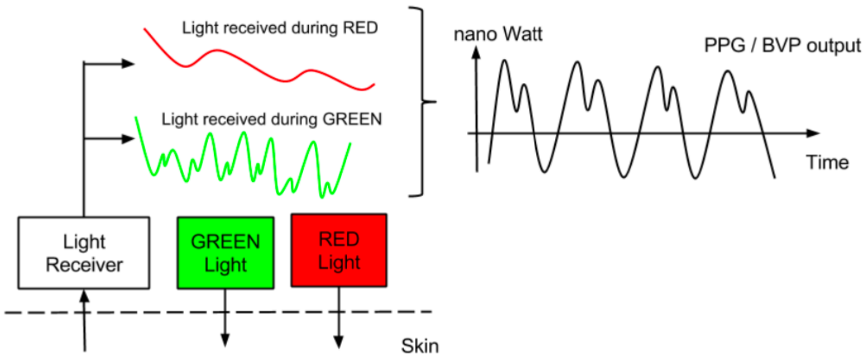

Figure 2.2 PPG

2.1.3 Electrodermal Activity

Electrodermal activity (EDA), also known as Galvanic Skin Response (GSR) [20] or Skin Conductance (SC), refers to changes in the electrical properties of the skin caused by sweating. It is an indicator of a person's emotional state or arousal. Measuring skin conductance variations is possible by applying an electrical potential between two points on the skin and measuring the current flow between them. The EDA signal is useful in assessing pain, and its applications in clinical settings are diverse. For a more detailed explanation of the formation of the EDA signal and the features that can be extracted from it, please refer to Section 4.2.2.

2.1.4 Inter-Beat Interval

IBI, which stands for Inter-Beat Interval [21], as shown in the Figure 2.3 refers to the duration between successive heartbeats. This measure is utilized to determine the current heart rate.

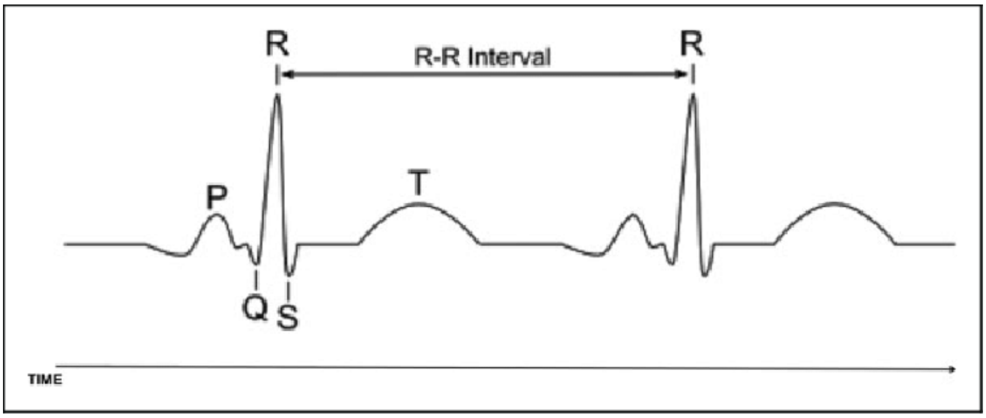

Figure 2.3 : Inter.Beat Interval

2.1.5 Blood Volume Pulse

The main output of the PPG sensor is the Blood Volume Pulse (BVP) [22]. This signal is obtained through a special algorithm that combines the light signals observed during green and red exposure, as shown in Figure 2.2. The BVP has a fixed sampling rate of 64 Hz, meaning it is sampled 64 times every second.

2.2 Machine Learning Algorithms

The field of Machine Learning [13] has seen rapid growth in recent times, with algorithms being developed to cater to various problem domains. Amongst these algorithms, *supervised*, *unsupervised*, and *reinforcement learning* techniques have been identified as the three primary categories. Supervised learning involves data points that come with an associated label, which allows the algorithm to identify patterns and generalize to situations not found in the dataset. The process is often referred to as "learning with a teacher". Unsupervised learning, on the other hand, deals with input data that does not have predefined outputs for each input. Therefore, such algorithms focus on uncovering shared characteristics within the input data. Reinforcement learning operates like "training under the guidance of a judge", as the algorithm receives rewards for correct results and penalties for incorrect ones.

In academic research, semi-supervised learning has gained increasing attention due to its ability to combine supervised and unsupervised learning techniques. This category of learning is particularly useful when the collection of data is easy, but labeling them is considered difficult due to the time it takes to label them. The main techniques of the semi supervised learning, that were used in this thesis are Self Learning [8], co-training [6] and Label propagation [7].

2.2.1 Self-Learning

Self-learning [8] is a popular approach in semi-supervised learning that relies on the assumption that a model's (figure 2.4) predictions on unlabeled data are trustworthy enough to be treated as true labels. However, the reliability of this assumption may not always be guaranteed, especially when the model encounters ambiguous or out-of-distribution samples. As a result, it is imperative to exercise caution when selecting the confidence threshold and closely monitor the model's performance to ensure the effectiveness of self-learning. Despite its potential shortcomings, self-learning remains a simple and effective way to leverage unlabeled data in various domains, including natural language processing, computer vision, and healthcare, as demonstrated by its successful application in recent research studies.

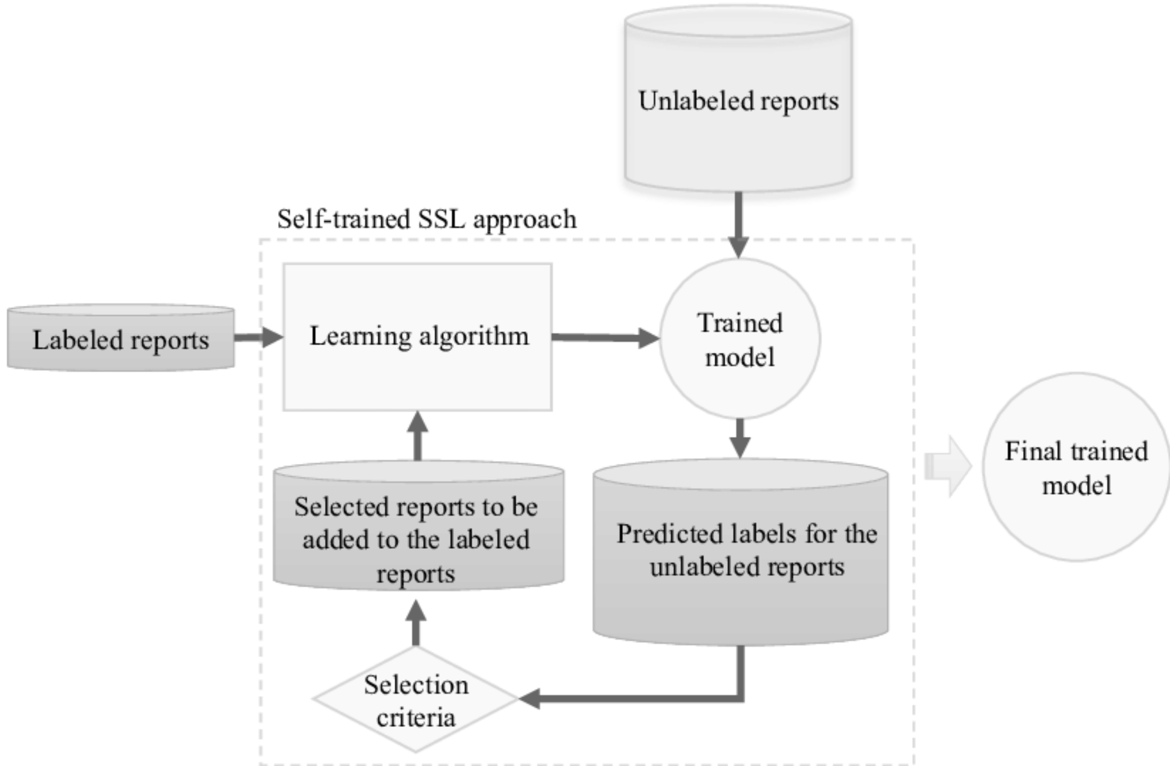

Figure 2.4 Self-Learning

2.2.2 Co-Training

The co-training [6] technique is a popular approach in semi-supervised learning that leverages the assumption that different views or representations of the data can provide complementary information, and instances, where the models agree, are likely to be correctly labeled. By iteratively expanding the labeled dataset with confident predictions from multiple models, co-training can effectively utilize unlabeled data to improve model performance. Co-training is especially useful when data can be naturally partitioned into distinct views or when multiple sources of information are available. Its successful application in various domains, including text classification, image recognition, and bioinformatics, attests to its effectiveness. However, the selection of views or representations and the monitoring of model agreement require careful consideration to ensure the success of co-training.

2.2.3 Label Propagation

The final technique that was tested in the study was label propagation [7], which is grounded in the assumption that similar instances should possess similar labels. By harnessing the associations between data points that are encoded in the graph or similarity matrix, label propagation effectively capitalizes on the information contained in both labeled and unlabeled data in order to deduce labels for the entire dataset. Label propagation is frequently employed in cases where the data can be depicted as a graph or where similarity information between data points is accessible, such as in social network analysis, citation networks, and image segmentation. Nevertheless, the efficacy of label propagation is contingent upon the quality of the graph or similarity matrix and the choice of propagation algorithm. Moreover, meticulous consideration of the initial labeling and stopping criterion is crucial to attain accurate label inference.

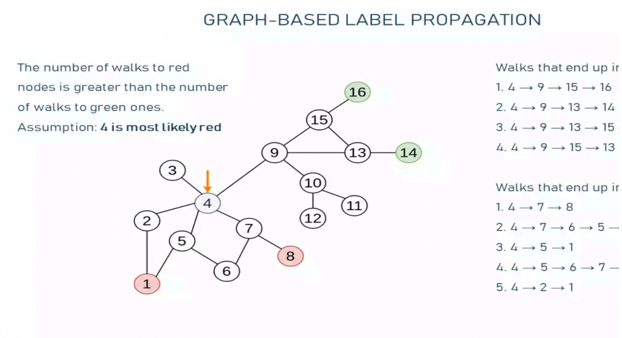

Figure 2.5 Label Propagation

2.2.4 GradientBoostingClassifier

The GradientBoostingClassifier [11] was chosen as the model for training the semi-supervised learning algorithm due to its demonstrated superior performance in the previous thesis [5]. This classifier is a powerful ensemble learning algorithm commonly employed for classification tasks in the field of machine learning. It constructs a model in a stepwise fashion using multiple weak learners, typically decision trees, to create a robust predictive model. The algorithm operates by iteratively adding new models that correct the errors made by the previous ones, focusing on the most challenging instances to classify. This is achieved through a technique known as boosting, where each subsequent model is trained to minimize the remaining errors of the collective ensemble of models constructed thus far. By leveraging the strengths of various weak models, the GradientBoostingClassifier enhances the overall accuracy and robustness of the predictive performance. Furthermore, it encompasses parameters to regulate overfitting, such as learning rate and tree depth, rendering it a versatile and widely utilized approach in diverse applications, including finance, healthcare, and marketing.

2.3 Model Evaluation

In this thesis, it is important to identify the best classification algorithm for the subject matter. To achieve this, it is crucial to choose the most appropriate evaluation methodology and performance metrics to compare the potential algorithms. This will help ensure that the chosen algorithm is the most effective and accurate for the task at hand.

The Machine Learning algorithms' performance is evaluated using the Stratified k-fold cross-validation methodology [9], which has been utilized and in previous studies in the past [1][2][3][4][5]. This method ensures that each fold contains a proportional representation of each class. The dataset is divided into k equal-sized folds while maintaining the original class label proportions. During k iterations of training and validation, one fold is reserved for validation while the remaining (k-1) folds serve as the training set. The model is trained on the training set and evaluated on the validation set, producing performance metrics (such as accuracy, precision, and recall) for each iteration. Finally, the average of these metrics across all k iterations provides a comprehensive evaluation of the model's performance.

To compute the performance metrics, four measures are crucial as shows in the figure 2.5: true positives, false positives, true negatives, and false negatives. True positives (TP) represent the number of correctly classified positive samples, while false positives (FP) represent the number of incorrectly classified positive samples. True negatives (TN) refer to the number of correctly classified negative samples, and false negatives (FN) indicate the number of incorrectly classified negative samples.

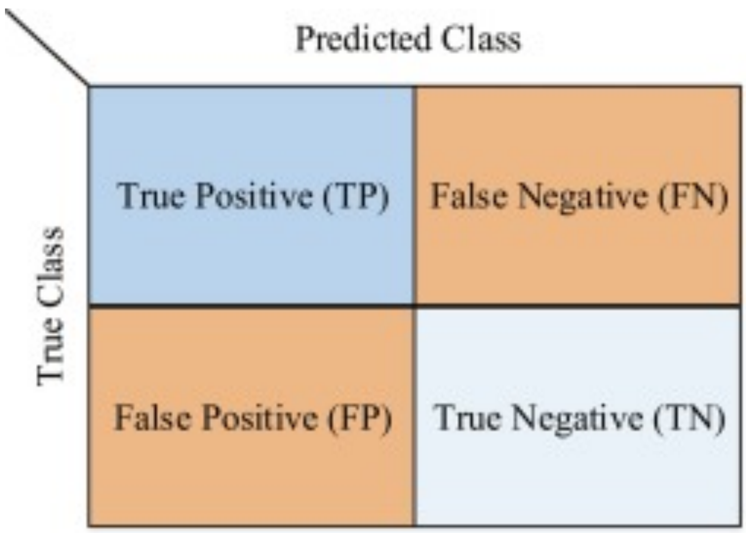

Figure 2.6 Confusion Matric

2.4 Monitor Device Empatica E4

The Empatica E4 wristband [23] is a wearable device that is designed to collect and analyze real-time physiological data. It is a small, lightweight, and comfortable device that can be worn on the wrist, making it an invaluable tool for researchers and healthcare professionals. The device is equipped with a photoplethysmography (PPG) sensor, which measures blood volume pulse to derive heart rate, heart rate variability, and other cardiovascular parameters. This sensor is used to monitor the heart rate and heart rate variability, which are important indicators of cardiovascular health.

In addition to the PPG sensor, the Empatica E4 wristband [10] also has an electrodermal activity (EDA) sensor. This sensor measures the electrical conductance of the skin to provide insights into emotional arousal, stress, and various psychological states. The EDA sensor is used to measure skin conductance, which is an important measure of emotional arousal and stress. The device also features a 3-axis accelerometer, which captures information on physical activity, movement, and gestures. This sensor is used to monitor physical activity and movement, which are important indicators of overall health and well-being.

Furthermore, the Empatica E4 wristband (figure 2.6) has an infrared thermopile that measures skin temperature for studying thermal regulation and other physiological parameters. This sensor is used to monitor skin temperature, which is an important indicator of various physiological functions, such as blood flow and skin hydration.

The Empatica E4 wristband is widely utilized in diverse research settings, including stress monitoring, sleep studies, emotion recognition, and mental health research. The device's versatility and effectiveness make it a powerful tool for both clinical trials and academic research projects. The Empatica E4 wristband was chosen for the ongoing study because of its wide range of sensors and demonstrated effectiveness in various research environments. This ensures precise, dependable data gathering and the capability to examine multiple physiological parameters simultaneously.

Figure 2.7 Empatica E4 wristband

Chapter 3

Data Collection and Previous Work

3.1 Psysiological Experiments

The Department of Psychology at the University of Cyprus conducted four experiments, focusing on diagnosing experiential avoidance in smokers, eating disorders, experimental avoidance for anxiety, and functional versus dysfunctional coping with acute pain. These experiments were conducted in the ACTHealth lab and involved volunteer participants. This section provides an overview of the methodology used in each experiment, as well as related research methodologies.

All of these experiments were connected to Acceptance and Commitment Therapy, which involves dividing individuals into two groups based on their reactions: acceptance or avoidance. As described in Section 1.2, acceptance-based strategies involve accepting one's thoughts and sensations as functional, while avoidance-based strategies involve attempting to avoid uncomfortable thoughts and sensations or control and alter them, which is considered dysfunctional. Furthermore, a person's classification can change based on the environment and circumstances. During the data collection for the first three experiments, it was assumed that each participant belonged to a single group throughout the entire procedure. However, this assumption did not apply to the fourth experiment.

3.1.1 Diagnosis of Experiential Avoidance in Smokers

The experiment was divided into five timeframes, each lasting 8 minutes. The first timeframe was used as a starting point, while the next two timeframes showed an emotionally neutral video. The final two timeframes showed a video meant to evoke negative emotions. The participant's signals were recorded using ECG, COR (fEMG), and GSR with a sampling rate of 1000Hz. Throughout the latter four timeframes, the participants took a series of cognitive tests to assess brain function. An expert from the Department of Psychology analyzed the collected data and categorized the participants accordingly.

3.1.2 Diagnosis of Eating Disorders

The goal of this study was to compare the emotional regulation abilities of individuals with low and high risk of developing an Eating Disorder. The experiment was divided into five timeframes, each lasting 2.5 minutes. The first timeframe was used as a baseline to ensure the participants' calm state. In the second and fourth timeframes, an emotionally neutral video was presented, while in the third timeframe, a general-content distressing video was shown. In the fifth timeframe, a video related to eating disorders was presented to the participants.

ECG, COR (fEMG), and GSR signals were recorded from the participants throughout the procedure with a sampling rate of 1000Hz. An expert from the Department of Psychology classified the participants into one of the categories using the collected data. Additionally, the Body Image Acceptance and Action Questionnaire (BI-AAQ) was administered to the participants to measure body image flexibility. The participants responded on a scale from never true to always true, with higher scores indicating greater body image flexibility.

3.1.3 Diagnosis of Experiential Avoidance for Anxiety

The objective of this experiment was to compare the emotional regulation abilities of individuals in the acceptance category versus those in the avoidance category regarding anxiety. The experiment consisted of 72 consecutive timeframes, each lasting approximately 1.8 minutes. In each timeframe, the participant viewed a single image that was meant to elicit a different reaction depending on whether they exhibited signs of anxiety or not. Using the collected data, an expert from the Department of Psychology classified the participants into one of the categories.

ECG, GSR, and fEMG (COR, ORB, and ZYG muscles) signals were recorded from the participants throughout the procedure with a sampling rate of 1000Hz.

3.1.4 Functional Versus Dysfunctional Coping with Acute Pain

In this study, the researchers aimed to investigate the effectiveness of acceptance and avoidance coping strategies in managing pain. The study involved 80 participants who were randomly assigned to four different conditions, each receiving a different set of instructions on how to deal with pain. The participants underwent three timeframes, which included a baseline period to ensure they were in a state of calm, followed by the Cold Pressor Task (CPT), where they were required to immerse their hand in cold water for as long as they could endure, and a final period to assess their pain management strategies. During these timeframes, multiple measures were recorded, including behavioral measures such as pain threshold and tolerance, psychophysiological measures such as ECG and EDA signals, and self-reported measures examining various aspects such as participants' psychological condition and their use of pain-coping strategies.

In the most recent experiment, the researchers used an Ecological Momentary Assessment (EMA) approach, which involved providing participants with smartphones and wearable psychophysiological monitors to wear for three days. During this phase, participants were prompted to respond to questions on an app pre-installed on the provided smartphones at fixed intervals throughout the day, asking about their social context, experiences of stress or pain (both physical and emotional), and their use of coping strategies. The use of smartphones and wearable monitors allowed for real-time data collection, providing a more accurate and comprehensive understanding of participants' experiences and coping strategies. If participants did not respond to the prompts, reminder messages were automatically sent every 30 minutes to ensure maximum participation.

Overall, these experiments provide valuable insights into the effectiveness of pain management strategies, and the use of different methodologies and technologies allows for a more comprehensive understanding of pain and coping strategies in different contexts.

3.1.5 Functional Versus Dysfunctional Coping in Real Time

The experiment that this thesis focuses on is the most recent one. During the Ecological Momentary Assessment (EMA) phase, participants were provided with smartphones and wearable psychophysiological monitors which they wore for three consecutive days. The participants were given instructions on how to wear and charge the Empatica E4 wristband to ensure proper data collection. They were then prompted to answer questions three times a day at fixed intervals every three hours from 10am to 10pm using an app installed by the researchers on the smartphones provided. The questions pertained to social context, experiences of stress or pain (both physical and emotional), and the use of coping strategies. Participants wore the monitors until bedtime and charged them at that time. Reminder messages were sent automatically every 30 minutes if a participant did not respond. The devices were returned by the participants after the three-day period.

3.2 Prior Work

In the current section, the methodologies of four previous works analyzing data from experiments on smoking, eating disorders, anxiety, and pain and emotions management are presented. This thesis focuses on an experiment mentioned in Section 3.1.5 that studies real-time pain and emotion management.

3.2.1 Diploma Project of Ch. Galazis in 2017

The aim of the project of Galazis [1] was to find the best combination of features to classify smoking and eating disorders experiments, based on previous research. For the anxiety experiment, additional work was done. The Random Forest classifier was trained and tested using all unique feature combinations, with the candidate features being the mean values of each recorded signal in each timeframe. The combination that had the highest accuracy and the fewest features was chosen. Different machine learning algorithms were studied, including Logistic Regression, Naive Bayes, K-Nearest Neighbours, Classification Tree, Neural Network, SVM, Bagging (using Decision Tree as the Base Learner), AdaBoosting (using Decision Tree as the Base Learner), Gradient Tree Boosting, and Random Forest. The data was divided into ten different training and test sets, and each algorithm was executed ten times. The results from the best-performing distribution were used for algorithm comparison.

3.2.2 Master Thesis of A. Trigeorgi in 2018

A different approach was employed in the study of the Trigeorgi [4], which focused more on feature extraction. The ECG signal was used to extract time-domain features, generating candidate features that included not only the mean values of each signal but also ECG-derived time-domain features (explained in detail in Section 4.1.1). To identify the optimal feature combination, a Random Forest Classifier [14] was used along with Stratified k-fold cross-validation, with the performance averaged across k iterations.

The same algorithms from the previous study were examined, but the execution method differed, using Stratified 5-fold cross-validation. The data were divided into training and test sets in five different ways, and each algorithm was executed five times, with the average performance of the five runs measured.

3.2.3 Master Thesis of G. Demosthenous in 2019

Breiman and Friedman's method was utilized in the study of Demosthenous [2] to extract even more features from the ECG signal. The method was employed to calculate feature importance using Gradient Boosting Decision Tree, ranking candidate features based on node impurity, to identify the most effective feature combination.

The algorithms studied in this research were different from the previous two studies, as they focused on tree-based algorithms. The five algorithms analyzed were Gradient Boosting Decision Tree (GBDT), Ada Boosting Decision Tree, Bagging Decision Tree (BDT), Random Forest (RF), and Extra Trees (ET). To increase the sample size and counter the assumption that each participant belonged to the same group throughout the experiment, an additional step was performed, which was training data multiplication.

Two methodologies, Moving Window Methodology (MWM) and Rectangular Window Methodology (RWM), were used and compared. The algorithm execution method combined the methods used in the previous two studies, using 10-fold cross-validation and executing each algorithm 10 times for each split, totaling 100 executions per algorithm. The prediction was also made for samples from previous experiments in this work.

3.2.4 Diploma Project of E. Georgiou in 2022

In the study of Georgiou [3] various physiological signals were monitored using BIOPAC, Microsoft Band 2, and the Moodmetric Smart Ring. These signals were recorded at different frequencies. Due to a limited dataset, the Rectangular Window Methodology was used to generate artificial samples. Four datasets were created using different window sizes (10, 20, 30, and 40 seconds). Time-domain measures were the main focus for ECG and HRV signals, and statistical metrics were extracted from the SCL and SCR components of the EDA signal. The thesis aimed to identify the most relevant features using three feature selection methods: Wrapper, Embedded, and Filter Methods. In addition, the thesis compared the common signals from different monitoring devices. The results varied depending on the methods used for comparison. Finally, the thesis concluded that data multiplication using Rectangular Window Methodology improved classifier performance and that data from the Microsoft Band 2 could match the performance of stationary devices.

3.2.5 Diploma Project of S. Zeniou in 2023

The work of Zeniou [5] goes beyond previous studies by analyzing the convergence of wearable technologies and Artificial Intelligence to tackle ineffective pain coping mechanisms. The study participants from the University of Cyprus utilized the Empatica E4 wearable device and a specialized app to gather real-time data on social context, stress, pain, and coping strategies. The study recorded and analyzed various physiological signals, with a particular focus on identifying essential characteristics. The study leveraged machine learning algorithms like Adaptive Boosting, Gradient Boosting Decision Tree, Random Forest, and Extra Trees to classify the coping strategies of the participants. The analysis emphasized the significance of heart rate variability features in the classification process. The Gradient Boosting Decision Tree model turned out to be particularly effective, with a 70% accuracy rate in distinguishing between functional and ineffective coping mechanisms. Interestingly, participants who did not match either coping category displayed characteristics of avoidance coping. Furthermore, the study found that the data collected from wearable devices yielded comparable results to those gathered from stationary devices, indicating the encouraging potential of wearable technology in this field.

Chapter 4

Classification

4.1 Signal Analysis

4.1.1 Data selection

A thorough data collection process was conducted for machine learning purposes, as outlined in Section 3.1.5. The study involved 88 participants, each with five associated files containing various physiological metrics, including Interbeat Interval (IBI), Heart Rate Variability (HRV), Temperature (TEMP), and (EDA). The data was collected over a three-day period for each patient.

In addition to the metrics, participants completed a questionnaire three times a day, with the time of completion recorded. The primary question asked was “What are you doing right now to manage your thoughts, feelings and emotions?”. Participants were given three answer options: avoidance for “I distract myself by doing or thinking about something else so that we avoid thinking about them,” acceptance for “I let the unpleasant thoughts and experiences be there without doing anything to drive them away,” and mindfulness for “I focus on what I’m doing now.”. In this thesis we take in consideration only the acceptance and the avoidance.

The first step of the analysis process involved identifying the time at which the pain-coping question was answered by each participant in their respective questionnaires. From there, the corresponding metric files were accessed and the data points were recorded five minutes before and after the question was answered. This approach allowed for a focused examination of the relationship between physiological metrics and patients’ pain-coping strategies.

4.1.2 Feature Extraction

Various features were extracted from the raw data to train algorithms efficiently. This was accomplished by previous researchers, and all details regarding the feature extraction and calculation can be found in theses from the past, specifically in Section 4.2.

4.2 Classification Process

As already mentioned, semi-supervised learning is a valuable technique for leveraging both labeled and unlabeled data in training machine learning models [8]. This approach is especially useful when acquiring labeled data is difficult or expensive. Notably, semi-supervised learning has been successfully applied in various fields, such as healthcare, finance, language processing, and computer vision, to produce more accurate models by incorporating large amounts of unlabeled data.

In this study, the outcomes of three different semi-supervised learning methods for classification were investigated. These algorithms, namely self-learning, co-training, and label propagation, were chosen to showcase a variety of techniques and compared for their effectiveness. In this thesis Section 2.2 provides detailed descriptions of each algorithm, as well as the feature selection process used to train them. Additionally, hyperparameters were optimized to maximize performance on the training data.

To evaluate the trained models, various metrics were used, as outlined in Section 2.3.1, including accuracy, recall, specificity, and F1-score. This assessment allows for a comparison of classifier performance, both across different datasets and different algorithms. The most efficient classifier(s) are determined based on these metrics, with a particular focus on their suitability for detecting and preventing dysfunctional pain coping behaviors.

The classification process employed in this study provides valuable insights into the efficacy of the selected machine learning algorithms and the impact of dataset optimization on their performance. By effectively cleaning and optimizing the dataset, the study aims to offer valuable guidance in improving the detection and prevention of dysfunctional pain coping behaviors.

4.2.1 Data Management

The semi-supervised algorithm techniques we used in this study required both labeled and unlabeled data for training and validation. To achieve this, we separated the labeled data into training and testing datasets, as explained in Section 2.3 of the report. However, we still needed to find a way to extract unlabeled records. After evaluating the situation and the available tools, we decided to modify the data extraction program. We extracted records before and after the label (5/10/20/30/60/120/180/240 minutes before and after) and considered them as unlabeled data.

To find the best combination of unlabeled data that would help train our model, we experimented with different grouping strategies. The two groups of unlabeled data that we used was all of them and in the second group we used only the *distance*s one (60/120/180/240 minutes before and after). Our methodical approach to tweaking the data and experimenting with various strategies allowed us to fine-tune our semi-supervised learning technique, improving the model's adaptability and effectiveness.

4.3 Result Comparison

This section describes the results of the three algorithms/techniques. In each combination of technique and data we used the k-cross validation method and we calculated the k-mean value and standard deviation of 5 times.

In the tables that are showing in each case there are the below metrics and there is one variable that is changing(*threshold* and *alpha*). The metrics formula is shown in the Figure 5.1

F1 Score :

The F1 score is the harmonic mean of precision and recall, providing a single measure of a model’s accuracy that balances both false positives and false negatives. It is particularly useful for imbalanced datasets.

Recall (Sensitivity):

Recall, also known as sensitivity, measures the proportion of actual positives that are correctly identified by the model. It is crucial in contexts where it is important to capture as many positives as possible. The positives in this scenario is the avoidance

Precision:

Precision quantifies the proportion of positive identifications that are actually correct. This metric is important when the cost of false positives is high.

AUC (Area Under the ROC Curve):

The AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. A higher AUC value indicates a better-performing model.

Specificity:

Specificity measures the proportion of actual negatives that are correctly identified. It is crucial in scenarios where it is important to identify as many true negatives as possible. The negatives in this scenario in the acceptance

Accuracy:

Accuracy is the ratio of correctly predicted instances (both true positives and true negatives) to the total instances. It provides an overall effectiveness of the model but can be misleading for imbalanced datasets.

$$
\begin{aligned}
precision &= \frac{TP}{TP + FP} \\
recall &= \frac{TP}{TP + FN} \\
F1 &= \frac{2 \times precision \times recall}{precision + recall} \\
accuracy &= \frac{TP + TN}{TP + FN + TN + FP} \\
specificity &= \frac{TN}{TN + FP}
\end{aligned}
$$

Figure 5.1

4.3.1 Self-Learning Results

For this technique, we used various combinations of the parameters and unlabeled data. As mentioned in the previous Section 4.2.1 Data Management we extracted a multiple window of unlabeled data. So to find the best combination of unlabeled data to feed the model we had to try a few. Also, another parameter that was important to test was the threshold.

Firstly we tried various combinations of unlabeled data and we started modifying the threshold bit by bit so we could find the optimal.

4.3.1.1 Full Labeled and Unlabeled Data

Threshold	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.5	0.829 ± 0.015	0.894 ± 0.037	0.775 ± 0.023	0.684 ± 0.029	0.474 ± 0.074	0.754 ± 0.020
0.55	0.814 ± 0.015	0.875 ± 0.068	0.767 ± 0.039	0.663 ± 0.037	0.452 ± 0.138	0.734 ± 0.014
0.6	0.830 ± 0.020	0.917 ± 0.029	0.759 ± 0.024	0.665 ± 0.036	0.413 ± 0.068	0.749 ± 0.029
0.65	0.819 ± 0.038	0.879 ± 0.057	0.768 ± 0.029	0.672 ± 0.047	0.466 ± 0.056	0.742 ± 0.050
0.7	0.824 ± 0.021	0.901 ± 0.031	0.759 ± 0.034	0.661 ± 0.046	0.421 ± 0.102	0.742 ± 0.033
0.75	0.829 ± 0.029	0.905 ± 0.032	0.764 ± 0.030	0.670 ± 0.054	0.434 ± 0.086	0.749 ± 0.044
0.8	0.829 ± 0.021	0.928 ± 0.025	0.750 ± 0.025	0.650 ± 0.048	0.373 ± 0.085	0.744 ± 0.035
0.85	0.825 ± 0.017	0.890 ± 0.061	0.773 ± 0.023	0.678 ± 0.024	0.470 ± 0.099	0.749 ± 0.014
0.9	0.826 ± 0.023	0.894 ± 0.054	0.769 ± 0.008	0.676 ± 0.019	0.458 ± 0.043	0.749 ± 0.026
0.95	0.812 ± 0.037	0.849 ± 0.047	0.779 ± 0.040	0.680 ± 0.066	0.511 ± 0.011	0.737 ± 0.055

Table 4.1 Self Learning Results Full Labeled and Unlabeled Data

The evaluation of the model’s recall indicates (Table 4.1) a proficient ability to detect instances of avoidance, yet a notable deficiency in identifying acceptances. This unbalanced performance, while consistent with the objective of emphasizing avoidance detection, introduces bias by elevating the rate of avoidance identification. Moreover, altering the threshold to 0.5, despite producing the most favorable outcomes, does not substantially address the issue. Ultimately, the existing model setup appears inadequate in effectively discerning between the two categories.

4.3.1.2 Full Labeled and Distance Unlabeled Data

Threshold	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.5	0.825 ± 0.032	0.874 ± 0.025	0.782 ± 0.041	0.690 ± 0.061	0.504 ± 0.102	0.752 ± 0.048
0.55	0.837 ± 0.024	0.890 ± 0.036	0.791 ± 0.024	0.708 ± 0.041	0.526 ± 0.073	0.769 ± 0.035
0.6	0.848 ± 0.020	0.931 ± 0.030	0.779 ± 0.016	0.698 ± 0.034	0.465 ± 0.048	0.777 ± 0.030
0.65	0.843 ± 0.030	0.913 ± 0.050	0.785 ± 0.020	0.704 ± 0.033	0.496 ± 0.042	0.774 ± 0.037
0.7	0.833 ± 0.037	0.890 ± 0.061	0.782 ± 0.027	0.697 ± 0.046	0.504 ± 0.060	0.762 ± 0.048
0.75	0.845 ± 0.037	0.920 ± 0.025	0.783 ± 0.049	0.701 ± 0.075	0.481 ± 0.121	0.774 ± 0.057
0.8	0.830 ± 0.030	0.879 ± 0.059	0.788 ± 0.033	0.698 ± 0.049	0.518 ± 0.106	0.759 ± 0.040
0.85	0.837 ± 0.024	0.898 ± 0.019	0.785 ± 0.030	0.700 ± 0.052	0.503 ± 0.092	0.767 ± 0.038
0.9	0.825 ± 0.046	0.863 ± 0.066	0.792 ± 0.041	0.703 ± 0.061	0.543 ± 0.092	0.757 ± 0.059
0.95	0.826 ± 0.015	0.856 ± 0.042	0.801 ± 0.033	0.710 ± 0.037	0.565 ± 0.103	0.759 ± 0.021

Table 4.1 Self Learning Results Full Labeled and Distance Unlabeled Data

The assessment of the model's recall indicates (Table 4.2) a proficient detection of avoidances, but an inadequate classification of acceptances. This unbalanced performance, though consistent with the objective of prioritizing avoidance detection, introduces bias by artificially inflating the avoidance identification rate. Despite a novel combination of factors demonstrating a marginal enhancement in accurately classifying both categories based on specificity, even with a threshold of 0.75, the overall performance remains unsatisfactory. Ultimately, the current model configuration appears inept at effectively discerning between the two categories.

4.3.1.3 Balance Labeled and Full Unlabeled

Threshold	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.5	0.681 ± 0.0.47	0.682 ± 0.077	0.684 ± 0.043	0.681 ± 0.045	0.679 ± 0.062	0.681 ± 0.045
0.55	0.676 ± 0.032	0.704 ± 0.045	0.652 ± 0.038	0.661 ± 0.036	0.617 ± 0.066	0.661 ± 0.036
0.6	0.686 ± 0.040	0.667 ± 0.062	0.712 ± 0.042	0.696 ± 0.033	0.725 ± 0.055	0.696 ± 0.033
0.65	0.634 ± 0.036	0.622 ± 0.047	0.650 ± 0.049	0.639 ± 0.043	0.656 ± 0.080	0.639 ± 0.044
0.7	0.653 ± 0.060	0.628 ± 0.093	0.687 ± 0.054	0.669 ± 0.045	0.710 ± 0.068	0.669 ± 0.044
0.75	0.646 ± 0.081	0.613 ± 0.089	0.685 ± 0.075	0.665 ± 0.072	0.717 ± 0.065	0.665 ± 0.072
0.8	0.693 ± 0.066	0.727 ± 0.112	0.670 ± 0.041	0.684 ± 0.048	0.641 ± 0.063	0.684 ± 0.049
0.85	0.620 ± 0.750	0.638 ± 0.118	0.612 ± 0.036	0.617 ± 0.043	0.596 ± 0.067	0.616 ± 0.045
0.9	0.677 ± 0.072	0.712 ± 0.059	0.651 ± 0.101	0.653 ± 0.094	0.594 ± 0.157	0.654 ± 0.094
0.95	0.649 ± 0.088	0.644 ± 0.124	0.660 ± 0.056	0.657 ± 0.074	0.671 ± 0.055	0.657 ± 0.073

Table 4.2 Self Learning Results Balance Labeled and Full Unlabeled

The analysis reveals on Table 4.2 a trade-off between accurate class separation and overall model performance. While improvements are seen in correctly classifying both acceptance and avoidance, this comes at a cost to overall accuracy. Despite achieving the best accuracy at a threshold of 0.6, the model's performance remains inadequate, indicating the need for further exploration and potential model reconfiguration.

4.3.1.4 Balance Labeled and Distance Unlabeled

Threshold	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.5	0.689 ± 0.059	0.666 ± 0.098	0.723 ± 0.035	0.703 ± 0.038	0.741 ± 0.060	0.703 ± 0.038
0.55	0.696 ± 0.074	0.689 ± 0.103	0.711 ± 0.061	0.703 ± 0.057	0.717 ± 0.072	0.703 ± 0.056
0.6	0.693 ± 0.036	0.675 ± 0.080	0.723 ± 0.036	0.704 ± 0.016	0.733 ± 0.066	0.704 ± 0.017
0.65	0.701 ± 0.084	0.691 ± 0.129	0.720 ± 0.051	0.712 ± 0.072	0.732 ± 0.056	0.711 ± 0.072
0.7	0.710 ± 0.026	0.698 ± 0.059	0.731 ± 0.058	0.716 ± 0.028	0.734 ± 0.084	0.715 ± 0.028
0.75	0.706 ± 0.068	0.690 ± 0.98	0.730 ± 0.049	0.715 ± 0.058	0.740 ± 0.063	0.715 ± 0.058
0.8	0.702 ± 0.055	0.682 ± 0.070	0.726 ± 0.050	0.712 ± 0.057	0.741 ± 0.047	0.711 ± 0.050
0.85	0.695 ± 0.031	0.674 ± 0.048	0.720 ± 0.037	0.703 ± 0.056	0.703 ± 0.056	0.703 ± 0.301
0.9	0.713 ± 0.019	0.727 ± 0.044	0.702 ± 0.022	0.707 ± 0.015	0.686 ± 0.047	0.707 ± 0.014
0.95	0.696 ± 0.076	0.683 ± 0.080	0.712 ± 0.086	0.7 ± 0.789	0.717 ± 0.094	0.7 ± 0.079

Table 4.4 Self Learning Results Balance Labeled and Distance Unlabeled

While previous attempts struggled to achieve a balance between accurate classification and overall performance, this new combination shows promising improvement. Here, we see on Table 4.4 increases in accuracy, specificity, and recall, indicating the model is learning to distinguish between the two classes. However, it's important to acknowledge that accuracy is still lower than ideal. This suggests a trade-off: the model can achieve better class separation at the expense of some overall accuracy. The best accuracy is achieved at a threshold of 0.7, but further refinement might be necessary to optimize both metrics.

4.3.1.5 Results

Overall as it seems from the results the best combination of data in Self learning technique was balance labeled data with distance unlabeled and threshold equals 0.7.

4.3.2 Co-Training

This technique we had trouble using a ready library so we decided to implement it myself.

But in the beginning, we had to find a way to split the features into 2 views. There are so many different ways and so many combinations. So we took the features from the last theses and tried on with many difference combinations(learning rate , balance/unbalance data and lastly we tried with all the features with no results.)
A sample of results of this technique results is shown below.

	Predicted negative	**Predicted positive**
Actual negative	3	25
Actual positive	3	48

Table 4.3 : Co-Training Confusion Matrix

F1	0.7741935483870968
Recall-Sensitivity	0.9411764705882353
Precision (PPV)	0.6575342465753424
AUC	0.5241596638655461
Specificity	0.10714285714285714
Accuracy	0.6455696202531646

Figure 4.6 : Co-Training Metrics

The analysis of the two tables above (Table 5.5 and Table 5.6) revealed that the model was unable to accurately identify negatives (acceptance). Consequently, it was determined that this technique was not suitable for the task at hand. As a result, the decision was made to discontinue its use and shift the focus to the remaining two techniques.

4.3.3 Label Propagation

In this technique as explained in previous section the algorithm works with distance. In the sklearn library in python there are 2 option for that(using rbf or using knn). Like previously we tested the algorithm both ways and also we made the same combinations as self-learning (filtering the unlabeled data and/or balancing the labeled). Each time we modified the alpha parameter (as the sklearn library write “ Clamping factor. A value in (0, 1) that specifies the relative amount that an instance should adopt the information from its neighbors as opposed to its initial label. alpha=0 means keeping the initial label information, alpha=1 means replacing all initial information.”) to find the optimal.

4.3.3.1 RBF

The RBF kernel is a powerful tool for analyzing data points in a high-dimensional space [15]. Its ability to identify complex relationships among data points makes it an ideal choice for a wide range of machine learning tasks, including label propagation. By leveraging the RBF kernel's similarity measure, label propagation can effectively propagate labels from labeled data points to their unlabeled counterparts.

4.3.3.1.1 Full Labeled and Unlabeled Data

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.784 ± 0.038	0.776 ± 0.063	0.796 ± 0.035	0.685 ± 0.048	0.595 ± 0.101	0.716 ± 0.043
0.2	0.787 ± 0.037	0.772 ± 0.043	0.803 ± 0.033	0.696 ± 0.046	0.619 ± 0.052	0.722 ± 0.045
0.3	0.767 ± 0.054	0.754 ± 0.096	0.787 ± 0.022	0.671 ± 0.036	0.588 ± 0.062	0.699 ± 0.054
0.4	0.780 ± 0.020	0.758 ± 0.021	0.806 ± 0.042	0.692 ± 0.046	0.627 ± 0.099	0.714 ± 0.031
0.5	0.774 ± 0.042	0.766 ± 0.069	0.786 ± 0.024	0.673 ± 0.045	0.580 ± 0.057	0.704 ± 0.049
0.6	0.805 ± 0.011	0.792 ± 0.029	0.820 ± 0.010	0.720 ± 0.008	0.649 ± 0.039	0.744 ± 0.009
0.7	0.776 ± 0.032	0.765 ± 0.062	0.791 ± 0.035	0.676 ± 0.039	0.588 ± 0.098	0.706 ± 0.034
0.8	0.795 ± 0.041	0.796 ± 0.080	0.801 ± 0.038	0.696 ± 0.046	0.596 ± 0.113	0.729 ± 0.043
0.9	0.783 ± 0.026	0.784 ± 0.022	0.785 ± 0.055	0.671 ± 0.068	0.558 ± 0.141	0.709 ± 0.045

Table 4.7 Label propagation RBF Full Labeled and Unlabeled Data

The combination we tested fails to adequately classify either avoidances or acceptances, and adjusting the alpha parameter doesn't seem on Table 4.7 to significantly improve the outcome.

4.3.3.1.2 Full Labeled and Distance Unlabeled Data

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.808 ± 0.027	0.811 ± 0.061	0.809 ± 0.024	0.711 ± 0.032	0.611 ± 0.081	0.744 ± 0.030
0.2	0.754 ± 0.033	0.746 ± 0.061	0.764 ± 0.026	0.640 ± 0.030	0.534 ± 0.073	0.676 ± 0.033
0.3	0.768 ± 0.042	0.754 ± 0.081	0.787 ± 0.016	0.671 ± 0.033	0.588 ± 0.054	0.699 ± 0.045
0.4	0.793 ± 0.040	0.784 ± 0.074	0.806 ± 0.029	0.701 ± 00.040	0.618 ± 0.077	0.729 ± 0.045
0.5	0.772 ± 0.044	0.758 ± 0.055	0.788 ± 0.042	0.673 ± 0.063	0.588 ± 0.092	0.701 ± 0.056
0.6	0.779 ± 0.036	0.776 ± 0.047	0.783 ± 0.042	0.670 ± 0.059	0.564 ± 0.103	0.706 ± 0.049
0.7	0.754 ± 0.043	0.743 ± 0.058	0.770 ± 0.060	0.642 ± 0.076	0.542 ± 0.146	0.676 ± 0.056
0.8	0.777 ± 0.026	0.762 0.038	0.796 ± 0.056	0.679 ± 0.063	0.596 ± 0.137	0.706 ± 0.041
0.9	0.765 ± 0.018	0.757 ± 0.029	0.774 ± 0.031	0.653 ± 0.040	0.549 ± 0.090	0.689 ± 0.026

Table 4.8 Label propagation RBF Full Labeled and Distance Unlabeled Data

The results of this combination overall on Table 4.8 shows that with alpha value 0.1 we get increase on the accuracy, recall and Specificity. That indicates that the algorithm priorities the existing labels on the graph and doesn’t give much weight to the similar nodes.

4.3.3.1.3 Balance Labeled and Full Unlabeled

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.652 ± 0.059	0.637 ± 0.067	0.675 ± 0.083	0.658 ± 0.062	0.680 ± 0.112	0.658 ± 0.062
0.2	0.638 ± 0.067	0.615 ± 0.097	0.672 ± 0.062	0.655 ± 0.057	0.694 ± 0.078	0.654 ± 0.057
0.3	0.694 ± 0.066	0.689 ± 0.057	0.702 ± 0.092	0.692 ± 0.074	0.695 ± 0.116	0.692 ± 0.074
0.4	0.651 ± 0.074	0.673 ± 0.118	0.637 ± 0.036	0.646 ± 0.046	0.619 ± 0.050	0.646 ± 0.047
0.5	0.696 ± 0.060	0.675 ± 0.070	0.722 ± 0.069	0.704 ± 0.086	0.732 ± 0.086	0.704 ± 0.061
0.6	0.646 ± 0.055	0.636 ± 0.093	0.663 ± 0.049	0.654 ± 0.037	0.672 ± 0.079	0.654 ± 0.036
0.7	0.650 ± 0.118	0.651 ± 0.147	0.654 ± 0.088	0.657 ± 0.096	0.664 ± 0.066	0.658 ± 0.095
0.8	0.649 ± 0.061	0.667 ± 0.092	0.637 ± 0.048	0.643 ± 0.052	0.619 ± 0.063	0.642 ± 0.053
0.9	0.656 ± 0.066	0.660 ± 0.133	0.671 ± 0.043	0.663 ± 0.036	0.666 ± 0.111	0.662 ± 0.037

Table 4.9 Label propagation RBF Balance Labeled and Full Unlabeled

Analysis of the results shows on Table 4.9 improved separation of the two classes compared to previous attempts. This improvement is achieved with an alpha parameter of 0.5, indicating that assigning equal weight to existing and propagated labels during the labeling process is most effective here.

4.3.3.1.4 Balance Labeled and Distance Unlabeled

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.695 ± 0.066	0.696 ± 0.084	0.697 ± 0.062	0.695 ± 0.058	0.694 ± 0.070	0.695 ± 0.058
0.2	0.707 ± 0.039	0.705 ± 0.046	0.710 ± 0.037	0.708 ± 0.037	0.710 ± 0.036	0.707 ± 0.037
0.3	0.684 ± 0.095	0.674 ± 0.104	0.697 ± 0.094	0.688 ± 0.089	0.702 ± 0.098	0.688 ± 0.089
0.4	0.700 ± 0.065	0.719 ± 0.110	0.691 ± 0.053	0.696 ± 0.050	0.673 ± 0.086	0.696 ± 0.051
0.5	0.665 ± 0.032	0.682 ± 0.095	0.665 ± 0.065	0.658 ± 0.038	0.634 ± 0.140	0.658 ± 0.037
0.6	0.696 ± 0.034	0.681 ± 0.056	0.716 ± 0.034	0.703 ± 0.027	0.725 ± 0.053	0.703 ± 0.028
0.7	0.703 ± 0.020	0.682 ± 0.038	0.726 ± 0.016	0.711 ± 0.012	0.740 ± 0.029	0.711 ± 0.012
0.8	0.670 ± 0.054	0.681 ± 0.073	0.661 ± 0.044	0.665 ± 0.046	0.648 ± 0.047	0.665 ± 0.046
0.9	0.690 ± 0.041	0.690 ± 0.080	0.698 ± 0.034	0.692 ± 0.033	0.694 ± 0.067	0.692 ± 0.033

Table 4.10 Label propagation RBF Balance Labeled and Distance Unlabeled

As Table 4.10 shows, with the best alpha value now at 0.7. This suggests that while both existing labels and propagated labels are important, the model benefits slightly more from prioritizing the propagated labels during the labeling process.

4.3.3.1.5 Results

Overall as it seems the best result of Label Propagation with rbf kernel is full label and distance unlabeled data with the alpha parameter equal with 0.1, that indicates that the algorithm gives doesn't give much weight to the around similar nodes.

4.3.3.2 KNN

In this setting, each data point is represented as a node in a graph, and the edges between nodes represent the similarity between them. The KNN kernel calculates the similarity between each pair of data points based on their distance in the feature space. Specifically, for each data point, it considers its k nearest neighbors and assigns higher similarity values to closer neighbors and lower values to farther ones.

4.3.3.2.1 Full Labeled and Unlabeled Data

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.777 ± 0.036	0.829 ± 0.060	0.732 ± 0.018	0.609 ± 0.032	0.389 ± 0.040	0.683 ± 0.040
0.2	0.783 ± 0.016	0.863 ± 0.031	0.717 ± 0.013	0.588 ± 0.024	0.312 ± 0.053	0.681 ± 0.020
0.3	0.770 ± 0.031	0.833 ± 0.023	0.715 ± 0.037	0.581 ± 0.060	0328 ± 0.100	0.666 ± 0.048
0.4	0.759 ± 0.027	0.830 ± 0.036	0.699 ± 0.022	0.556 ± 0.037	0.283 ± 0.040	0.648 ± 0.037
0.5	0.745 ± 0.029	0.795 ± 0.059	0.703 ± 0.023	0.558 ± 0.037	0.320 ± 0.084	0.638 ± 0.035
0.6	0.756 ± 0.042	0.841 ± 0.055	0.690 ± 0.042	0.538 ± 0.076	0.236 ± 0.013	0.641 ± 0.063
0.7	0.770 ± 0.045	0.852 ± 0.075	0.704 ± 0.025	0.567 ± 0.043	0.282 ± 0.060	0.663 ± 0.051
0.8	0.759 ± 0.033	0.852 ± 0.064	0.686 ± 0.022	0.533 ± 0.041	0.213 ± 0.087	0.641 ± 0.041
0.9	0.775 ± 0.011	0.883 ± 0.036	0.692 ± 0.009	0.545 ± 0.013	0.207 ± 0.048	0.658 ± 0.011

Table 4.11 Label propagation KNN Full Labeled and Unlabeled Data

Results show on Table 4.11 that this model is not able to separate the 2 classes independently of the alpha parameter. As it seems from the specificity it cannot classify the acceptance at all.

4.3.3.2.2 Full Labeled and Distance Unlabeled Data

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.742 ± 0.008	0.814 ± 0.013	0.681 ± 0.022	0.522 ± 0.034	0.229 ± 0.081	0.620 ± 0.020
0.2	0.744 ± 0.038	0.807 ± 0.064	0.691 ± 0.019	0.054 ± 0.032	0.275 ± 0.027	0.630 ± 0.043
0.3	0.736 ± 0.046	0.799 ± 0.089	0.684 ± 0.016	0.530 ± 0.029	0.260 ± 0.046	0.620 ± 0.047
0.4	0.727 ± 0.028	0.788 ± 0.043	0.675 ± 0.020	0.512 ± 0.045	0.236 ± 0.062	0.605 ± 0.040
0.5	0.755 ± 0.020	0.841 ± 0.033	0.685 ± 0.015	0.531 ± 0.026	0.221 ± 0.037	0.635 ± 0.027
0.6	0.731 ± 0.038	0.791 ± 0.067	0.680 ± 0.023	0.522 ± 0.036	0.252 ± 0.062	0.613 ± 0.043
0.7	0.743 ± 0.011	0.826 ± 0.031	0.675 ± 0.014	0.512 ± 0.022	0.199 ± 0.062	0.618 ± 0.015
0.8	0.748 ± 0.026	0.841 ± 0.044	0.674 ± 0.016	0.512 ± 0.034	0.183 ± 0.043	0.623 ± 0.034
0.9	0.769 ± 0.024	0.882 ± 0.039	0.682 ± 0.024	0.525 ± 0.051	0.167 ± 0.104	0.646 ± 0.038

Table 4.12 Label propagation KNN Full Labeled and Distance Unlabeled Data

Same in this case, again the model isn't able to identify the 2 classes. As it seems on Table 4.12 from the specificity it cannot classify the acceptance at all. The results are not satisfactory.

4.3.3.2.3 Balance Labeled and Full Unlabeled

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.595 ± 0.045	0.575 ± 0.071	0.628 ± 0.076	0.608 ± 0.054	0.640 ± 0.127	0.608 ± 0.054
0.2	0.572 ± 0.172	0.567 ± 0.193	0.588 ± 0.136	0.600 ± 0.111	0.634 ± 0.088	0.600 ± 0.111
0.3	o.629 ± 0.085	0.623 ± 0.113	0.644 ± 0.076	0.636 ± 0.081	0.648 ± 0.106	0.635 ± 0.081
0.4	0.576 ± 0.063	0.546 ± 0.057	0.615 ± 0.101	0.594 ± 0.074	0.641 ± 0.121	0.593 ± 0.073
0.5	0.582 ± 0.098	0.594 ± 0.116	0.575 ± 0.084	0.578 ± 0.092	0.564 ± 0.082	0.578 ± 0.092
0.6	0.573 ± 0.087	0.592 ± 0.130	0.562 ± 0.061	0.567 ± 0.069	0.542 ± 0.073	0.567 ± 0.069
0.7	0.571 ± 0.074	0.544 ± 0.079	0.606 ± 0.088	0.589 ± 0.074	0.635 ± 0.109	0.589 ± 0.073
0.8	0.561 ± 0.053	0.560 ± 0.069	0.565 ± 0.048	0.563 ± 0.042	0.566 ± 0.068	0.563 ± 0.042
0.9	0.590 ± 0.085	0.592 ± 0.115	0.591 ± 0.054	0.594 ± 0.065	0.595 ± 0.038	0.593 ± 0.065

Table 4.13Label propagation KNN Balance Labeled and Full Unlabeled

Like other combinations balancing the 2 classes gives better results. But again not satisfactory enough. The change on the alpha parameter does not seems to affect the model as shown on Table 4.13.

4.3.3.2.4 Balance Labeled and Distance Unlabeled

Alpha	F1	Recall-Sensitivity	Precision	AUC	Specificity	Accuracy
0.1	0.610 ± 0.078	0.634 ± 0.120	0.594 ± 0.059	0.600 ± 0.064	0.565 ± 0.089	0.601 ± 0.064
0.2	0.611 ± 0.040	0.636 ± 0.079	0.594 ± 0.032	0.596 ± 0.032	0.556 ± 0.096	0.597 ± 0.031
0.3	0.600 ± 0.084	0.607 ± 0.119	0.599 ± 0.060	0.601 ± 0.070	0.595 ± 0.068	0.601 ± 0.069
0.4	0.577 ± 0.040	0.576 ± 0.060	0.581 ± 0.030	0.578 ± 0.035	0.580 ± 0.057	0.580 ± 0.035
0.5	0.564 ± 0.130	0.576 ± 0.154	0.555 ± 0.056	0.563 ± 0.115	0.563 ± 0.080	0.563 ± 0.115
0.6	0.600 ± 0.106	0.631 ± 0.123	0.579 ± 0.108	0.579 ± 0.119	0.528 ± 0.116	0.579 ± 0.118
0.7	0.575 ± 0.082	0.592 ± 0.111	0.562 ± 0.058	0.567 ± 0.073	0.541 ± 0.059	0.567 ± 0.073
0.8	0.586 ± 0.055	0.592 ± 0.076	0.585 ± 0.063	0.582 ± 0.054	0.573 ± 0.090	0.582 ± 0.054
0.9	0.536 ± 0.096	0.547 ± 0.108	0.527 ± 0.527	0.525 ± 0.100	0.502 ± 0.115	0.525 ± 0.100

Table 4.14 Label propagation KNN Balance Labeled and Distance Unlabeled

In this combination, we got a bit worse results than the other cases as shown on Table 4.14. The alpha parameter is shown to be slightly better close to low values.

4.3.3.2.5 Results

Overall as it seems the KNN kernel doesn't suit in our case as in every combination fail to give satisfactory results

Chapter 5

Discussion

5.1 Summary

In general our results does not differ much from the last thesis results.

In conclusion the combination that gave the best results was Label Propagation with full labeled data with distance unlabeled and alpha value 0.1. Then the second best result is the self learning with balance labeled data and distance unlabeled with threshold 0.7.

In summary, it appears that the semi-supervised learning techniques utilized mostly had a positive impact on the data. However, the impact varies depending on the quality of the data used. The results show that selecting both labeled and unlabeled data has a noticeable effect. The model seems to perform best when given balanced data from both classes, although this does result in a drop in accuracy. This approach appears to increase the model's understanding of both classes, as indicated by acceptable levels of recall and specificity. When using all of the data, it becomes clear that the model struggles to recognize the acceptance class.

Furthermore, it seems that selecting distance records from the labeled data is more optimal for the unlabeled data than feeding in all of it. However, this also depends on the quality of the data and the experiment's design. It may be better to allow the user to clear their mind from the questionnaire and return to their normal activities. After testing various techniques, self-learning and label propagation with balanced labeled and distance unlabeled data proved to be the most suitable approach for this case.

Finally is worth noting that in the Label propagation technique, when the parameter alpha was low, it returned a slightly better result. That indicates that the labeled data is sparse or potentially *noisy* [24], as it allows the algorithm to better adapt to the underlying structure of the data.

5.2 Future Improvements

In the future, if the project continues with this approach, with the semi supervised learning, it will be a huge plus if data were recollected with more precautions so to have more clean data to work with, as explained in the summary (Section 5.1) the algorithm Label propagation shows that when the alpha parameter was low it gave better results, that indicates potential noisy labeled data. Also this data were collected on students and many students they were not taking the experiment on a serious level.

Another way that we found but we decided not to work with because it wasn't in the semi supervised learning techniques was try to generate records instead of trying to labeled the unlabeled with a form of GAN algorithms [12]. That might help because as it shows the unlabeled data they play a important role.

In the future, it would be beneficial to explore advanced machine learning algorithms, such as ensemble methods, deep learning models, and neural networks, in combination with semi-supervised learning. These advanced models have the potential to significantly enhance performance by capturing intricate data patterns and providing more accurate predictions. Furthermore, another technique that might be useful because of the noisy data it is some variation of *SOMs* (Self Organizing Maps) [25]. It's a type of unsupervised learning but with some modifications and after analyzing the neighborhoods it may give better results and the noisy data they will not affect the so much the model.

Βιβλιογραφία

[1]. C.Galazis. Non-Intrusive Physiological Wearable Devices for Identifying Individual Difference Parameters Using Supervised Classification Learning Algorithms. 2017.

[2]. G. Demosthenous. Machine Learning Approach to Predict Emotional Coping Using Psychophysiological Signals, MSc thesis, Department of Computer Science, University of Cyprus. 2019.

[3]. E.Georgiou. Feature Selection and Training of Machine Learning Algorithms to Classify Functional versus Dysfunctional Coping with Acute Pai, Department of Computer Science, University of Cyprus. 2022.

[4]. A. Trigiorgi. Μελέτη Μεθόδων Μηχανικής και Βαθιάς Μάθησης και Εφαρμογή σε Ψυχομετρικά Δεδομένα, Diploma Project, Department of Computer Science, University of Cyprus. 2016.

[5]. S. Zeniou analysis of real-time e4-based psychophysiological data for machine-learning based classification. Department of Computer Science, University of Cyprus 2023

[6]. Blum Avrim and Mitchell Tom. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*

[7]. *Xiajin Zhu,Zoubin Ghahramani Learning from labeled and unlabeled data with label propagation. Garniegie Mellon University Pittsburg, 2002*

[8]. Xiaojin Zhu, Andrew. B Goldberg. Introduction to Semi-Supervised Learning. Springer Nature, 31 May 2022

[9]. Jason Brownlee. A gentle introduction to k-fold cross-validation. 5 2018.

[10]. C. McCarthy, N. Pradhan, C. Redpath and A. Adler, "Validation of the Empatica E4 wristband," 2016 IEEE EMBS International Student Conference (ISC), Ottawa, ON, Canada, 2016

[11]. Prettenhofer, P., & Louppe, G. (23 February 2014). Gradient Boosted Regression Trees in Scikit-Learn [Paper presentation]. PyData 2014, London, United Kingdom.

[12]. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, "Generative Adversarial Networks: An Overview," in IEEE Signal Processing Magazine, Jan. 2018

[13]. Geron, Aurelien. Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd ed., O'Reilly Media, Inc., 2019.

[14]. Breiman, Leo. "Random Forests." Machine Learning 45.3 (2001): 5-32

[15]. Guang-Bin Huang, P. Saratchandran and N. Sundararajan, "A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation," in IEEE Transactions on Neural Networks, vol. 16, no. 1, pp. 57-67, Jan. 2005

[16]. Brandon Farnsworth. What is ecg and how does it work? 8 2021.

[17]. David Castaneda, Andres Esparza, Mohammad Ghamari, Cinthia Soltanpur, and Homer Nazeran. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron, 4(4):195–202, 2018.

[18]. John Wilson. What is facial emg and how does it work? iMotions, 2018.

[19]. C. E. Ackerman. How Does Acceptance And Commitment Therapy (ACT) Work?,"

[20]. iMotions. Galvanic skin response (gsr): The complete pocket guide. 2020.

[21]. Brüser, C., Stefan Winter, and Steffen Leonhardt. "Robust inter-beat interval estimation in cardiac vibration signals." Physiological measurement 34.2 (2013)

[22]. Nitzan, M., A. Babchenko, and B. Khanokh. "Verylow frequency variability in arterial blood pressure and blood volume pulse." Medical & biological engineering & computing 37 (1999): 54-58.

[23]. https://www.empatica.com/en-gb/research/e4

[24]. Zhu, Xiaojin, et al. "Semi-supervised Learning with Gaussian Fields and Expectation Propagation." (2005). https://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf.

[25]. Kohonen, T. (2001). The self-organizing map. [Paper presented at the International Conference on Neural Networks (ICNN'97), Perth, Western Australia, Australia, December 9-13, 1997]