UNIVERSITY OF THE PELOPONNESE & NCSR "DEMOCRITOS"
MSC PROGRAMME IN DATA SCIENCE

Multimodal Workplace Monitoring for Work-Related Fatigue Recognition

by

Alexandros Mitsou

A thesis submitted in partial fulfillment
of the requirements for the MSc
in Data Science

Supervisor: Theodoros Giannakopoulos
Principal Researcher of Multimodal Machine Learning

Athens, July 2021

Multimodal Workplace Monitoring for Work-Related Fatigue Recognition

Alexandros Mitsou

MSc. Thesis, MSc. Programme in Data Science

University of the Peloponnese & NCSR "Democritos", July 2021

UNIVERSITY OF THE PELOPONNESE & NCSR "DEMOCRITOS"
MSC PROGRAMME IN DATA SCIENCE

Multimodal Workplace Monitoring for Work-Related Fatigue Recognition

by

Alexandros Mitsou

A thesis submitted in partial fulfillment
of the requirements for the MSc
in Data Science

Supervisor: Theodoros Giannakopoulos
Principal Researcher of Multimodal Machine Learning

Approved by the examination committee on July, 2021.

(Signature)	(Signature)	(Signature)
.....	.....	.....
Theodoros Giannakopoulos Principal Researcher	Spiros Skiadopoulos Professor	Evaggelos Spyrou Assistant Professor

Athens, July 2021

UNIVERSITY OF THE PELOPONNESE & NCSR "DEMOCRITOS"
MSC PROGRAMME IN DATA SCIENCE

Declaration of Authorship

(1) I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

(2) I confirm that this thesis presented for the degree of Master of Science in Informatics and Telecommunications, has

(i) been composed entirely by myself

(ii) been solely the result of my own work

(iii) not been submitted for any other degree or professional qualification

(3) I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or processional qualification except as specified.

(Signature)

. . . . . .

Alexandros Mitsou

Athens, July 2021

Acknowledgments

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor, Dr.Giannakopoulos Theodoros, for his invaluable advice, continuous support, and patience during my dissertation within the postgraduate program of Data Science of the University of the Peloponnese in collaboration with the National Centre for Scientific Research "Demokritos". I would also like to thank Dr.Spyrou Evaggelos whose expertise was invaluable in formulating the research questions and methodology. I would also like to thank Computational Intelligence Laboratory for providing knowledge and support. And my biggest thanks to my family for all the support you have shown me through these years.

To my family and friends.

Περίληψη

Σκοπός αυτής της εργασίας είναι ο σχεδιασμός και η υλοποίηση ενός συστήματος το οποίο ασχολείται με την αναγνώριση της ανθρώπινης συμπεριφοράς πάνω σε θέματα ψυχικής φύσεως. Δεδομένου ότι, η είσοδος της τεχνολογίας στην ανθρώπινη καθημερινότητα αυξάνεται ολοένα και περισσότερο μέρα με τη μέρα και έτσι, νέες εφαρμογές και μέθοδοι της επιστήμης της Πληροφορικής εισέρχονται σε αυτή. Επομένως, προκύπτει ένα αυξανόμενο ενδιαφέρον για την δημιουργία εφαρμογών οι οποίες, θα λειτουργούν ως βοήθημα για τα άτομα που αντιμετωπίζουν προβλήματα εργασιακής κόπωσης, στρες και άγχους. Επί της ουσίας, η εργασία αυτή, αποτελεί έναν σύγχρονο τομέα εφαρμογών της επιστήμης των δεδομένων, η οποία συνδυάζει τις επιστημονικές περιοχές της πληροφορικής και της ψυχολογίας. Σε αυτή την εργασία, παρουσιάζεται μία προσέγγιση με στόχο την αυτόματη αναγνώριση δραστηριοτήτων που λαμβάνουν χώρα, στα πλαίσια ενός εργασιακού περιβάλλοντος, μέσω μιας σειράς παρατηρήσεων οι οποίες συμπεριλαμβάνουν ενέργειες, συμπεριφορές και καταστάσεις που αποτελούν πηγή κόπωσης, άγχους και στρες. Για να επιτευχθεί ο στόχος της εργασίας, χρησιμοποιείται ένας συνδυασμός πηγών πληροφορίας, οι οποίες προέρχονται από τα περιφερειακά μέρη ενός ηλεκτρονικού υπολογιστή, καθώς και πληροφορία η οποία προέρχεται από αισθητήρες αντίστασης-δύναμης. Προκειμένου να πραγματοποιηθεί η εξαγωγή γνώσης διερευνώνται διάφορες τεχνικές μηχανικής μάθησης και πιο συγκεκριμένα μάθησης υπό-επίβλεψη. Αρχικά, ορίζεται ένα σύνολο δραστηριοτήτων, οι οποίες λαμβάνουν χώρα, στα πλαίσια ενός εργασιακού περιβάλλοντος. Βάση αυτών, συγκεντρώνονται μετρήσεις, οι οποίες χρησιμοποιούνται για την κατασκευή ενός συνόλου δεδομένων το οποίο αντικατοπτρίζει τις δραστηριότητες που έγιναν εντός του χρόνου εργασίας. Στη συνέχεια εκπαιδεύονται διάφοροι ταξινομητές με σκοπό την

αναγώριση των διάφορων δραστηριοτήτων.

Abstract

The aim of this thesis is to design and implement a system that deals with the recognition of human behavior on issues of a mental nature. Since the input of technology in human daily life is growing more and more every day and so, new applications and methods of computer science to enter it. Therefore, there is a growing interest in creating applications that will serve as an aid for people experiencing problems with work fatigue, stress and anxiety. In essence, this work is a modern application field of data science, which combines the disciplines of information technology and psychology. In this work, an approach is presented with the aim of automatically recognizing activities that take place, within a work environment, through a series of observations which include actions, behaviors and situations that are a source of fatigue, anxiety and stress. In order to achieve the goal of this work, a combination of information sources is used, which come from the peripheral parts of a computer, as well as information coming from force sensitive resistor sensors. In addition, in order to extract knowledge various machine learning techniques, and more specifically supervised learning are explored. Initially, a set of activities is defined, which take place within a working environment. Based on these activities, measurements are collected, which are used to construct a data set that reflects the activities performed during the working time. Various classifiers are then trained to identify the various activities.

Contents

List of Tables

List of Figures

List of Abbreviations

e.g.	exempli gratia
etc.	et cetera
i.e.	id est
ss	segment size
ANS	Autonomous Nervous System
CNS	Central Nervous System
HAR	Human Activity Recognition
FSR	Force Sensitive Resistor Sensor
SVM	Support Vector Machine
DT	Decision Tree
NB	Naive Bayes
KNN	K-Neareset Neighbor
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
TPR	True Positive Rate
FPR	False Positive Rate
IDE	Integrated Development Environment

ECG	Electrocardiography
EMG	Electromyogram
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability
BVP	Blood Volume Pressure
EDA	Electrodermal Activity
RSP	Respiration
GPS	Global Positioning System

Chapter 1

Introduction

1.1 Problem Statement

The task that we will attempt to solve in this thesis is the classification of user activities that take place within work environments. First and foremost, we are called to solve a problem that arises from the research areas of human behavior recognition in combination with psychology. In other words, the activities that a user does during a working day affect his/her psychology, which translates into anxiety, stress and, fatigue. Thus, by combining tools and techniques from the world of information technology, we can achieve more organization in the collection, processing and synthesis of psychological data [1]. To that end, we created a multimodal data collection pipeline, using a combination of modalities of three types. More specifically and respecting the ethical and legal protocols [2] we collected information coming from the user's keyboard and mouse log, in combination with a set of FSR sensors that were placed in the user's chair. In addition we collected the video and audio stream from the webcam, but this information was used only for verification. Finally, we extracted a selection of hand-crafted features in order to evaluate our statistical models.

1.2 Motivation

One of the biggest scourges of our time is stress and anxiety. Stress is to blame for causing serious health problems. Briefly, stress is directly related to diseases of the cardiovascular system, psychiatric disorders, digestive problems, decreased libido as it directly affects the hormones associated with the reproductive process and many more. In addition, it has been shown that people living and working in large urban centers increase the chances of premature aging as well as the chances of suffering from chronic stress. Eventually, people get sick because the stress-response becomes damaging due to the reason that stress will disrupt a wide variety of immune functions. In essence, the modern way of life, combined with the socio-economic changes contribute to the increase of risks that arise from problematic work planning, organization, and management, as well as from an unhealthy social work environment that may lead to negative psychological, physical, and social outcomes, such as work stress, burnout, or depression. Some examples of working conditions that may lead to psychosocial risks are: excessive workload, conflicting demands and ambiguities regarding the role of the employee, lack of participation in decision-making that affects the employee, lack of influence on the way work is done and low productivity.

When considering job requirements, it is important not to confuse psychosocial risks such as overwork with situations that, despite being demanding and sometimes difficult, they are characterized by a supportive work environment that provides motivation and proper training to employees in order to maximize their skills. Employees experiencing anxiety attacks when the demands of the job are excessive and go beyond their ability to address them. A healthy psychosocial environment increases productivity and personal development, as well as the mental and physical well-being of employees. In terms of organization or enterprise, negative consequences may include poor overall business performance, increased rate of absence from work, formal presence at work and increased rates of accidents and injuries. The costs for business and society is estimated to be significant, amounting to billions of euros at national level.

Thus, for the reasons above, achievements in computer hardware and software, shifting social needs and the continuous research in the field of Computer Science and Psychology, have enabled the evolution and progress has been made in order to better understand the human mind. These concepts seem to constantly cause the philosophical and theoretical concepts of exploring the subconscious mind, thus giving practical solutions in terms of stress, such as one to understand the real reasons for experiencing stress. Although various attempts have been made, none of them got very far from the respective areas of psychology and computer science, in terms of multimodal information analysis and fusion. In this work, we examine an alternate path, introducing the problem of analyzing and classifying human behaviour using multimodal information coming from multiple sources.

1.3 Related Work

In this section we try to systematically document the relevant research work that has been proposed in the last decade. The works we present are divided into categories depending on the nature of the data collection. More specifically, first research is presented which mentions the possibilities of combining computer science and psychology. Then, a bibliography is presented, which aims to highlight how data is collected and processed. In essence, the data collected fall into either the **behavioral** data category or the **physiological** data category in terms of stress detection. In this work, our main focus includes the data collection process and the signal nature that were collected.

1.3.1 Approaches for Stress Detection

During the last few years, many research efforts have focused on the problem of stress and anxiety recognition. In this section, we present works that attempt to recognize stress and anxiety using non-invasive approaches. Non-invasive methods are similar to those that are recognized in this work. Stress is an important risk factor for a variety of diseases. This is a reaction mode of today's man, who feels that he can not cope with a modern lifestyle, a key feature of which are the fast pace of everyday life, and excessive demands. As it is obvious, stress has a wide range

of effects on the psychic sphere, for this reason if we observe more carefully we will see that stress manifests itself based on some factors. In the relevant literature it has been recorded that there are many physical and emotional reactions such as the following: Stuttering, neck pain, lower back pain, muscle spasms, dizziness, lethargy, red cheeks, sweating, dry mouth, indigestion, shortness of breath, deep sighing, difficulties in concentration, attention, constant fatigue, weakness, exhaustion. Based on this, scientists were inspired to model and design biodata monitoring systems in order to make measurements around stress.

1.3.1.1 An overview

Linsey Barker, et al.[3], brought to the fore the issue of workers fatigue. They studied the causes of fatigue in the workplaces, informing that the sources of fatigue can originate from the organization of the work, the shift type and patterns, the total working time and recovery and the number and duration of breaks. In order to associate workplace fatigue with stress, anxiety and human activity detection, we had to research more.

Based on this, Christian Montag, et al.[4], introduced a new research field, under the name 'Psychoinformatics' that could be very useful for psychologists. This field, can enrich their methods, due to the reason that they can observe human daily routines using large datasets. Furthermore psychoinformatics brings in new tools and techniques, originated from the field of computer science in combination with the field of psychology, in order to improve data acquisition, organization and synthesis of psychological data [1]. That means that, for the first time, a specialist does not have to rely on a patient's self assessment.

Alexander Markowetz, et al. [5], highlighted the need to study human behaviour to a greater scale. They claimed that, big data is an opportunity for tracking and analyzing behaviours. Thus specifically devices are nowadays available for self-tracking [6], in order to draw conclusions about a person's mental state and physiological behaviour. These devices are designed to be worn upon the body to automatically collect data on bodily functions that include physiological and behavioural signals.

1.3.1.2 Physiological Signal Approaches

Thus, in order to continue our research, we had to identify what signals can be used as a means to detect stress. In the literature, the plurality of signals which are displayed comprise signals derived from physiological functions. Dvijesh Shastri, et al.[7], applied physiological signals through thermal imaging. They took advantage of respiration, a physiological phenomenon that contains a clear thermal print. They argued that stress can be detected using neurophysiological responses on the perinasal area. That is clearly explained for the reason that respiratory responses during stress are not the same as respiratory responses in normal conditions. In addition they used EDA probes on the index finger.

Takuto Hayashi, et al. [8], chose to study whether stress states affect brain responses in the regions related to emotional and cognitive processing. In order to achieve their purpose, they applied functional magnetic response imaging (fMRI).

Rafal Kocielnik, et al. [9], created a framework, allowing people to discover their stress reaction patterns using obtrusive monitoring technologies. More specifically, data were collected through a wristband device for non-invasive stress detection and data logging. Their framework, collected a variety of data to provide the user feedback about the stress experienced in relation to the context in which it was experienced. They collected and processed physiological signals such as skin conductivity, temperature, ambient temperature and 3D accelerations.

Kais Riani, et al.[10] introduced a multimodal dataset in order to detect levels of alertness in drivers, proposing a machine learning framework aiming to investigate that multimodal features increase the detection of that activity. Their set of modalities, include an RGB-Camera and physiological indicators such as skin conductance, blood pressure, respiration rate, and skin temperature. Moreover regarding such datasets, Burcu Cinaz et al.[11] collected data such as ECG, HRV with the addition of salivary cortisol in order to measure stress levels. In order to recognize stress emotion.

Changzhi Wei [12], applied a union of multiple physiological signals such EMG and RSP signals, claiming that they can improve the correctness of the stress detection

rate. What we observed is that most research associated with data mainly from ECG, BVP, EMG, HRV, EDA as a standalone solution or in combination with other sources such as questionnaires.

1.3.1.3 Behavioural and Psychological Signal Approaches

Jonathan Aigrain et al. [13], developed a multi assessment methodology to analyze stress detection results. Various test were conducted for the evaluation process, including a socially evaluated mental arithmetic test, self-assessment from external observers and an assessment from a physiology expert. What set them apart from most studies was that they studied stress from biological perspective i.e how the body responds to a stressful stimulus, phenomenological perspective i.e stress can occur with no body changes present and from behavioural perspective i.e behaviours made by humans. Therefore, the sources of information they used came from video, a depth camera, physiological and behavioural signals such as BVP, EGC, EMG, GSR, HR, HRV, speech signals, body movements, the position of the head, etc. Armando Barreto et al. [14], proposed a methodology in order to determine the affective state of a computer user from the measurement of physiological signals and more specifically signals originated from the pupil diameter.

Some approaches include the analysis of behavioural patterns when research participants interacting with technological devices. In [15], they argued that the level of the users stress, could be estimated using sensory data derived from a cellphone e.g. touch patterns, accuracy, intensity, duration, amount of movement, etc. Clayton Epp et al. [16], proposed a solution in order to determine user emotion by analyzing the rhythm of their typing patterns on a standard keyboard. A formal definition of keyboard analysis states that 'Keystroke Dynamics' is the study of the unique timing patterns in an individual's typing and typically includes extracting keystroke timing features such as the keystroke duration, latency, etc.

As an extension of Clayton's et al. work, Belinda Eijckelhof et al. [17], studied associations between workplace stressors and office employees computer use patterns. Based on this assumption, they collected the following data including, the time a participant spent working at a computer, the number and duration of computer

breaks, the pace and frequency of keystroke and mouse dynamics per minutes, as well as the mean duration of the individual key strikes in miliseconds. As for the mouse events, they monitored the cursor velocity and the frequency of mouse button clicks.

A further investigation based on such signals, has been proposed by Klemen Peternal et al. [18]. They addressed that in order to measure stress levels using only the traditional way of questionnaires is the major fact that the results are usually poor. The reasoning behind this thought, is that the patients are not always able to recall the entire history of encounters with different stressors. On the other hand, the proposed contextual information as a means of improving stress detection as a whole process. User context is mostly described by a set of parameters, which depict information about the user's state at a certain point in time. In order to achieve their goal, they implemented a multimodal dataset collection pipeline including: A simple questionnaire incorporated into a mobile application, the user's GPS location and physical activity. In addition, they took advantage of audio signals, keystroke and mouse dynamics as well as the user's phone call status.

A similar work has been proposed by Daniel McDuff et al. [19]. They developed a multimodal sensor setup system named 'AffectAura' for continuous logging of audio, visual, physiological and contextual data streams. They exploited these types of signals, in order to predict a user's affective state when working in front of a computer. Moreover, participants were able to reconstruct their workday activities, based on the logged data. That means that, users could explore the reasoning over the data artifacts presented in this system. For the purpose of data logging, they used a simple web camera, Microsoft Kinect, an EDA sensor, the GPS location as well as the file activity and the calendar's information. At this point it is important to mention another source of stress indication. The vocal projection of a human could be a very useful source of emotional leakage. Psychologist Paul Ekman has shown that we are not always aware of the emotions we transmit in our speech [20]. Vijay Patil et al. [21], proved that voice stress analysis is an alternative for obtaining a non-invasive way, to extract information about a user's stress state. This can be achieved through the vocal pitch (frequency) changes from low to high pitch during

vocalization.

1.3.2 Smart Furniture for Measuring Stress and Anxiety Levels

In the most recent publications, work has been presented concerning the use of sensors in smart furniture. The goal is to support people with an emphasis on the elderly population. As it has been observed that the elderly spend a lot of time sitting or lying on furniture, the need arose to create smart furniture. The central idea is that by integrating networks of sensors and mechanical parts into objects such as furniture, it is possible to build a system that will provide supervision similar to a professional who provides care services to the elderly. A very interesting size that [22] took into account is the installation of humidity sensors in order to measure liquid leaks in the bed or sofa. Moreover, they have observed that many older people are overweight without knowing their weight, which can be dangerous for their deteriorating health. For this reason, they placed weight sensors on furniture such as the bed, so that the user knows his/her current weight without further inconvenience. Wang [23] modeled and processed voice data in order to make specially designed furniture. In particular, they integrated an encoder and a voice decoder on a sofa, so that the user's stored voice data, when recognized, would perform the functions designed for the sofa. Regarding the development of software and the use of measurements through sensors, [24] presented in their work the use of pressure sensors which were placed at the feet of a chair. In essence, for this task the question that had to be answered was ultimately, how much information can be extracted about the activities performed by a user if we take into account data coming from his chair. So they were able to model changes in the weight distribution of a chair, based on a combination of posture-related features, in order to identify high-level activities such as whether a user is working on their computer or whether a user watches a movie, etc. Panasonic [25] has developed a smart mirror, which is one level ahead of the common makeup sets and mirrors used by beauticians. This mirror, also known as 'Snow Beauty Mirror' captures the image of a user's face and then creates beauty tips using processed data stored in its memory. As a step further,

this mirror can provide the user with a printable make-up that can be placed on his face, thus replacing the conventional make-up.

In conditions where employees perform 'office work' it is observed that musculoskeletal health problems often occur such as back pain (back pain), back pain and neck pain, due to poor posture. In addition, problems occur in muscles, nerves, ligaments, tendons and articular cartilage, which can aggravate pre-existing conditions such as tendonitis, muscle fractures and chondropathies, turning them from acute to chronic conditions, which are now difficult to treat. The term 'posture' refers to the position of the body: how a person holds himself when he is standing, sitting or lying down. In orthopedics, posture also refers to how the muscles, joints, and spine work together to keep a person upright. Poor posture is considered an asymmetrical body position. Bert Arnrich et al.[26], stated that posture affective states information related to stress can be found in the posture channel during office work. Thus, they took measurements that derived from FSR sensors mounted on the surface of a chair, in order to create features from the pressure distribution.

1.4 Thesis Structure

The rest of this thesis is organized as follows: Initially, in chapter 2, we describe the necessary theoretical background, emphasizing the stress mechanisms and machine learning basics and methods we used. Then, in chapter 3 we discuss about the data collection pipeline and the proposed architecture in detail in order to recognize the activities of an employee during a workday. In chapter 4 we outline the experiments conducted and we present our findings and results. Finally in the last chapter we sum up our challenges and findings, and we discuss future work, focusing on how our own contributions could build up to a large scale data set that could better handle the task of multimodal activity recognition, in relation with the conclusions that emerged.

Chapter 2

Theoretical Background

2.1 Stress Mechanisms

Stress is a physical and mental reaction to the various experiences of our lives, related to our simple daily responsibilities, but also to more serious issues, such as health and death. For short-term situations that require an immediate response, stress can benefit us. However, if stress is a daily occurrence, then it can negatively affect many aspects of our health. In particular, chronic stress can cause a variety of symptoms that affect a person's overall well-being. The central unit of the stress system is located in the brain and consists of the hypothalamus and pituitary gland, while the peripheral part consists of the sympathetic nervous system and the adrenal glands. For this reason, below we present the systems of the human body and emphasis is given to the nervous system which is the stress system. In order to achieve the objectives of this work, it is important to understand where the signals and behaviors that are recorded come from as our purpose is to collect and model them. So it is important to take a few steps back and study the mechanisms of stress on a very basic but useful level. For this reason below we refer to issues of physiology of the human race.

2.1.1 Systems of the Human Body

The term system refers to a set of organs that work together to perform the same function in the human body. The human body includes the following systems:

1. Skeletal / excitatory system
2. Muscular system
3. Respiratory system
4. Circulatory system
5. Lymphatic system
6. Digestive / gastrointestinal system
7. **Nervous system**
8. Endocrine system
9. Immune system
10. Urinary system
11. Reproductive / genital system
12. Leather / cover system
13. The system of sensory organs

The Autonomous Nervous System: Of the systems found in the human body, the one that is most directly related to stress is the nervous system. The nervous system is responsible for controlling the harmonious cooperation of all organs of the body, it contributes to thought, perception, communication and emotion.

The organs that make up the nervous system are the brain, the nerves and the senses (nervous and sensory system). The functions of the nervous system are as follows:

1. To receive and transmit aesthetic information both from the external environment and from the rest of the body through the centripetal part of the Peripheral Nervous System to the Central Nervous System (CNS).

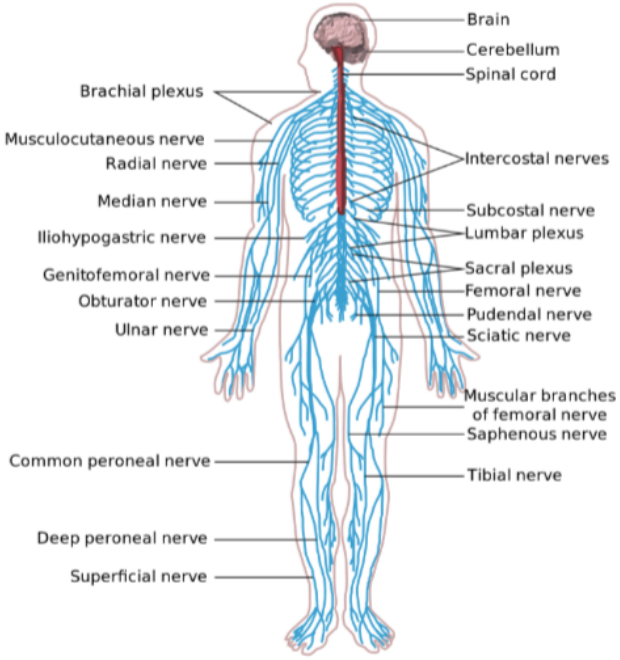


Figure 2.1: The human nervous system

2. To process that the CNS recruits (spinal cord for reflexes, brain for higher and more complex behaviors).

3. To respond to the **stimuli** it receives. That is, to regulate and control a response to the stimuli it receives through the centripetal fate of the Peripheral Nervous System. This answer can be either voluntary such as e.g. Moving away from a danger either unintentionally such as e.g. Sweating when we get too hot

The CNS works against our will in the human body, which means that its function takes place in the subconscious mind (controlled by the brain). The CNS is divided into two branches in the **sympathetic** and **parasympathetic system** [27].

The **sympathetic** system generally plays an important role in stressful or emergency situations. It is the system that takes action when the body needs to be awake or more specifically in "FFF and sex mode (Freeze-Flight or Fight and sex mode)"[28]. In this case, symptoms appear in the body such as: tachycardia, hypertension, bronchodilation, hair straightening, inhibition of digestive function, mydriasis.

The **parasympathetic** system, on the other hand, controls the basic functions of

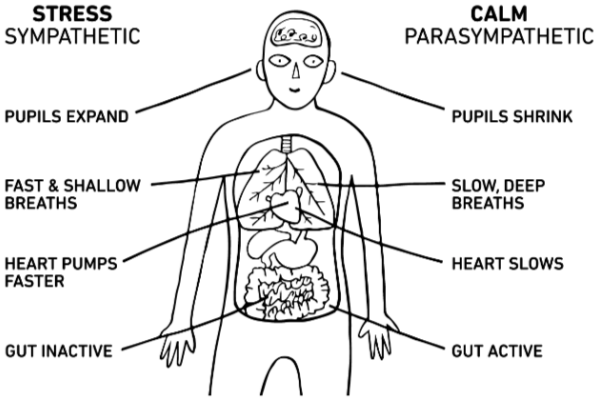


Figure 2.2: The sympathetic and parasympathetic nervous systems

the body when it is at rest. In particular, it takes action when the body is at rest. It also restores the body's functions to a normal rhythm after situations of tension. It is also responsible for regulating digestive and homeostatic functions such as thermoregulation, thirst, hunger, etc. Finally, the coordination of the action of the two systems precisely regulates the involuntary functions of the myocardium, smooth muscle and glands.

Thus, it is important to note that the brain plays a central role in allostasis. It is achieved by simultaneously controlling all mechanisms through the brain, which can impose a high level of oversight and incorporate powerful factors such as experience, memories, anticipation and reassessment of needs to orchestrate appropriate responses. The mediators of allostasis act on the various tissue receptors, in order to produce short and advantageously adaptively results. If the mediators of allostasis do not stop their activity or if they are insufficiently regulated, then the 'allostatic state' is produced which can cause damage as a result. The central idea is that physical or mental costs may arise if the allostatic state is maintained for a sufficient period of time.

2.1.2 Stress and Anxiety

Stressors: A stressor is any stimulus that occurs in the outside world and contributes to the inhibition of the balance of homeostasis. In essence, the human body reacts to such stimuli in order to return to homeostatic balance. In particular,

stress-induced stimuli fall into four categories:

1. **Physical stimuli** e.g. cold, heat, vibration, noise, which usually have a negative psychological impact

2. **Psychological stimuli** that cause behavioral changes, such as anxiety and fear

3. **Social stimuli** that reflect a disturbed relationship between two or more people e.g. divorce, unemployment

4. Those that alter **cardiovascular** and **metabolic homeostasis** such as exercise, orthostasis, hypoglycemia and bleeding.

Finally, a stressor is any stimulus that achieves the inhibition of homeostatic balance function. In addition, the body's response to stress is the process by which the body tries to restore allostasis. More specifically, in the science of biology, the term 'homeostasis' [29] implies the state of stable internal, physical and chemical conditions maintained by living systems. This dynamic equilibrium condition is the condition of optimum operation for the organism and involves many variables, such as body temperature and fluid balance, maintained within certain predetermined limits, and (homeostatic range). Other variables include the pH of the extracellular fluid, the concentrations of sodium, potassium and calcium ions, as well as those of the blood sugar level, which must also be regulated despite changes in environment, diet or activity level. Each of these variables is controlled by one or more regulators or homeostatic mechanisms, which together maintain life.

Allostasis: Originally intended as a theory that would replace homeostasis, but is now considered part of homeostasis, although as concepts have great differences. Thus allostasis is defined as the adaptive process for the active maintenance of stability or homeostasis of the organism, through the change of the function of various organic systems and the expression of their mediators [27].

2.1.3 On Stress and Machine Learning

Figuring out when a person is stressed, can be proven very helpful due to reason that it is something that can contribute to earlier prevention of many health issues. Thus, in order to identify stress, there exist remarkable changes in various bio-signals that indicate stress markers. Using such signals we are able to identify these changes and figure out stress levels. Therefore, at this point the field of **Machine Learning** is introduced, which can offer techniques and tools that enable us to model those signals that arise from homeostatic changes. Based on these signals, data sets can be created that we can use to utilize machine learning tools in order to construct prediction models. Next, a reference is made to the machine learning methods and techniques used in this work.

2.2 Machine Learning

2.2.1 Machine Learning in General

Nowadays, we live in the age of the so-called big data, where all kinds of electronic data are constantly produced and thus the need has arisen to create, process and store this data. For this reason, in order to utilize these data we need tools and techniques. Therefore, the research field of machine learning can be used as a means to analyze the knowledge derived from these data with the ultimate goal of drawing conclusions. Machine learning is a combination of statistical science, artificial intelligence and computer science. The term machine learning in computer systems is called the creation of models from a set of data. More specifically, a stricter definition according to Mitchell [30] is:

Definition 1. *'A computer program is considered to be learning from experience E in relation to a class T task and a performance measure P if its performance in class T tasks as valued by P are improved by experience E.'*

Machine learning is essentially a field of artificial intelligence, which involves algorithms and models that allow computer systems to learn using statistical methods that aid for data analysis. In essence, it is a collection of methods and techniques

that can automatically identify various patterns in the data and then based on them, make predictions about the future of the results or make decisions based on specific situations. Although there are variations in the definition of the types of machine learning, it is usually possible to divide them into categories depending on the nature of the problem, falling into one of the following categories:

- **Supervised Learning:** In this type of learning the computer system is called to 'learn' a function which is called a target function and is a description of the model produced from the data.

  In addition, the training data that are given as an input to each algorithm contain the desired solutions. These data are called 'labels'. A typical process that takes place in this type of learning is classification.

- **Unsupervised Learning:** In this type of learning, the computer is called upon to discover relationships in an unlabeled set of data, creating patterns of unknown number and existence. In essence, we want to be able to separate the data into structures, drawing conclusions about their characteristics. The most common operation which occurs in this type of learning is clustering. For instance, a usage example is the 'market basket analysis'. This example of usage is trying to evaluate products that a customer might buy based on previous purchases they have made.

- **Reinforcement Learning:** In reinforcing learning, we use data sets that have not been tagged. For this reason, an evaluation measure has been defined in order to see the quality of the results. The operation of this type of learning is via a mechanism, called 'reward mechanism'. Each time the algorithm gives the desired result, it receives a 'reward' and its goal is through a repetitive process to maximize the reward it receives. In this way the training process is carried out in order to solve the requested problem. Reinforcement learning usually finds applications in games, where for example an agent learns the strategies and moves of a game such as chess or table tennis, constantly testing moves which, if successful, increase the degree of reward. In this way, we can advice an agent when it wins and when it loses by simply using an evaluation

function that gives reasonable estimates of the probability of winning from any position.

2.2.2 Supervised Learning

Even though, a wide variety of tasks exist that could be solved using machine learning techniques, supervised machine learning aims to solve two kinds of problems:

1. **Classification Problems**
2. Regression Problems

In the scope of this thesis we focus only on the supervised learning and more specifically on classification problems. Therefore, at this stage, the current machine learning algorithm, trying to get the right information, which will allow for the control stage to investigate and identify the classification of each activity. This information is essentially the data to be trained which should contain the correct answer, which is known as the 'target'. The algorithm then finds patterns in the data to be trained that correspond to the characteristics of the input data in the target attribute, i.e. the prediction response. The learning algorithm results in a statistical prediction model, the purpose of which is to identify these patterns. Therefore, each model can be used to predict new data for which the target is unknown. The training process is illustrated in the following image.

2.2.3 Classification

Classification is of the most important tasks in the area of machine learning. Its main purpose is to provide a categorized label of new case categories based on previous observations. Classification turns down into two major categories which include binary classification, where the algorithm learns a set of rules in order to distinguish labels between two possible categories. On the other hand, the algorith learn a set of rules in order to classify data into more than two classes. This type of classification is called multiclass classification. Classification can be described as a procedure that contains two stages. The first stage refers to the data pre-preprocessing and model training, while the second stage refers to model testing and evaluation. Now, let

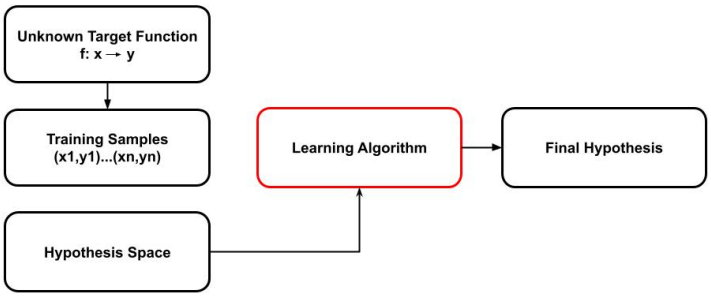


Figure 2.3: Statistical model training procedure

us discuss some of the most widely used classification algorithms and techniques in details.

2.2.3.1 Support Vector Machines

Support Vector Machines (SVM) [31] are feed-forward networks that can be used for classification and function estimation problems. Support Vector Machines operate as one of the best approaches for data modeling. It is a popular machine learning technique, which has been successfully applied to many real classification problems from various fields. Their basic functionality is the hyper-plane construction that can be used, in order to decide to which class belong the input samples. This hyper-plane is built, so as to maximize the margin of separation between positive and negative samples. In addition, it must be noted that the hyper-plane is not constructed at the input space, where the problem may not be solved linearly, but in the feature space where it has been directed. A hyper-plane that has been built that way is generally referred as the 'Optimum Hyper-Plane', an acquired property, as support vector machines are an accurate implementation of the structural risk minimization method. Despite the fact that they incorporate knowledge, adapted to the subject area in which they are applied, they provide remarkable generalization ability. Their main feature is an inner product kernel between the input vector and

one of the support vectors. The set of the latter is defined as a subset of the vectors that make up the training set and are extracted based on an optimization algorithm. Based on these vectors the optimum hyper-plane is constructed. The SVM kernel is built based on various techniques, leading to numerous SVM types. The most basic and widely used kernel functions used to address nonlinearity problems in real data include:

- **Linear Kernel:** The linear kernel function is only suitable for linear separable data
- **Polynomial Kernel:** The application of a polynomial kernel dramatically increases performance. The higher the degree of the polynomial, the more flexible the decision function becomes.
- **Sigmoid Kernel:** The sigmoid kernel or tangent kernel, although is mainly used in neural networks, often gives remarkable results in the application of SVMs.

2.2.3.2 Decision Tree

The decision tree [32] is a supervised learning algorithm that can solve classification problems, where they are used to predict which class a sample belongs to, and regression problems, where they are used to predict the numerical values of a sample. It is a tree-shaped graph, where each inner node represents a feature, each branch represents the possible output, and each leaf represents a class. The path from the root to the leaf represents the classification rules. Initially, the tree receives a training set of data, containing various samples that characterize it. Input features can be discrete or continuous, as the characteristics of output values. The most important thing is that the tree that has been created should not be overloaded. Some of the best known algorithms for decision trees are: ID3, C4.5, CART and SPRINT.

2.2.3.3 Naive Bayes

Naive Bayes classifiers [33] are a family of algorithms, which are based on Baye's theorem and a hypothesis that the features are independent of each other. Suppose there is a data set D that is represented by an N-dimensional vector $X = (x_1, x_2, \ldots, x_N)$ which is the measurement at each sample of the set of N features. Let us assume that there are K categories $c_1, c_2, \ldots, c_K$ of a class c, where we need to predict its categorization. In order to categorize a sample X, we calculate the probabilities $P(c_1|X), P(c_2|X), \ldots, P(c_K|X)$. These probabilities denote the probability that a sample X_i belongs to class c_1 or c_2 or c_K respectively. Sample X_i is categorized into that category whose probability $P(X_i|c)$ is the maximum probability. In order to examine the probability whether a sample belongs to a specific class we use the Baye's theorem, so we have:

$$P(c_i|X) = \frac{P(X|c_i)P(c_i)}{P(X)}$$

Naive Bayes classifiers have the advantage that they need a small amount of training data and can be trained effectively using the method of supervised learning. Attributes are represented as vectors, with the principle that the value of each attribute is independent of the the rest. In this way, instead of calculating the time consuming combination of all the conditional probabilities, we calculate only their product, for this reason Naive Bayes algorithm is called a 'greedy algorithm'.

2.2.3.4 K-Nearest Neighbor

The k-nearest neighbor [34] (k-NN) algorithm is a non-parametric method used in classification and regression problems. It is also called a 'lazy algorithm' since it memorizes the entire training set in memory. The basic idea of the algorithm for classifying an element is that the properties of each specific element entered in the algorithm should be similar to those that have other points, based on a certain distance from the newly entered element. This distance is often called 'the neighborhood'. We need to mention that 'K' is an integer that is used as a reference to the number of the nearest neighbors, a decision factor for the classification. The implementation steps of the algorithm are as follows: Initially we need to select

the number of neighbors and the metric distance. Then the neighboring element will be found, based on which the input element must be categorized. Finally, the new item is categorized. In order to measure closeness, we need to define a metric. Therefore, we need a metric based on the distance between points. Such distance measures are Euclidean distance, Manhattan distance, Minkowski distance and Hamming distance. The aforementioned distance metrics are calculated as follows:

1. Euclidean distance or L_2 Norm:

   For $p = 2$ the distance is called Euclidean and is calculated using the following formula: $d(x_1, x_2) = \sqrt{\sum_{i=1}^{N}(x_{1i} - x_{2i})^2}$

2. Manhattan distance or L_1 Norm:

   For $p = 1$ the distance is called Manhattan and is calculated using the following formula: $d(x_1, x_2) = \sum_{i=1}^{N} |x_{1i} - x_{2i}|$

3. Minkowski distance:

   Is a generalization of Euclidean and Manhattan distance and is calculated using the following formula: $d(x_1, x_2) = \sqrt[p]{\sum_{i=1}^{N} |x_{1i} - x_{2i}|^p}$

4. Hamming distance:

   Hamming distance can be proven very useful for discrete attribute calculation. This distance is calculated using the following formula: $d(x_1, x_2) = \sum_{i=1}^{N} 1_{x_{1i}} \neq x_{2i}$

One of the advantages of (k-NN) algorithm is the direct categorization in the new training data. On the other hand a major disadvantage of this method is that the introduction of new data linearly increases computational complexity.

2.2.4 Evaluation Metrics

When evaluating algorithms and classification, an index and tool for measuring success accurately is necessary. Some of the most common in the machine learning community are defined below. So in order, for better understanding the next terms, we need to introduce the following terms and definitions.

2.2.4.1 Confusion Matrix

Confusion matrix is a matrix that describes the complete performance of a statistical model. It provides the appropriate information about the actual and the predicted classifications performed by a categorization system. Performance is evaluated based on the matrix data, which are symbolized as follows:

Term	Description
True Positive(TP)	Correctly classified event
True Negative(TN)	Correctly rejected event
False Positive(FP)	Type I error
False Negative(FN)	Type II error

Table 2.1: Definition of terms for classification performance

For instance, suppose an M-classes classification problem. It is important to examine if there is any number of classes that tend to show a great tendency to confuse. Let, a confusion matrix $A = [A(i, j)]$ be defined such that an element $A(i, j)$ is the number of points, given which the actual class label was in class i and was classified in class j. Based on the confusion matrix we can compute precision, specificity, sensitivity (or recall), accuracy and f1-measure.

2.2.4.2 Precision, Recall, F1-measure

Precision(P): Precision can be seen as the positive predicted value rate. Of all samples classified as class i, how many were truly belonging to class i. This metric can be very informative about the discrimination of class i, related to other classes.

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

Recall (R): Measuring recall is sometimes called the true positive rate, resembled by the following formula:

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

This metric determine the rate of how many instances of class i where correctly classified as i, and how many were misclassified. By using this metric, we are able to detect how well class i is recognized.

F1 score (Macro Averaged): F1 score is actually an approach in order to represent accuracy, utilizing both precision and recall, using the average of the aforementioned. This metric represents the harmonic mean, resembled by the following formula:

$$F_1 = 2\frac{PR}{P+R}$$

Support: Support depicts the number of true occurrences of each class, as well as the number training samples per label class.

2.2.4.3 Sensitivity and Specificity

In order to determine reliability of the results, metrics have been devised to help investigate the relationship between tests and whether or not an activity is recognized.For this reason, statistical analysis is performed using diagnostic accuracy measures. Key measures for quantifying the accuracy of a test include sensitivity and specificity. Sensitivity of a statistical test is the probability of the true positive diagnosis of an activity. On the other hand, specificity of a statistical test is the probability of the false positive diagnosis of an activity.

The calculation of sensitivity is given by:

$$Sensitivity = TPR = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - FNR$$

The calculation of specificity is given by:

$$Specificity = TNR = \frac{TN}{N} = \frac{TN}{TN+FP} = 1 - FPR$$

2.2.4.4 ROC Curves

Making predictions is an important concern for any scientific field. It is therefore necessary to ensure predictive accuracy in the design and comparison of models, algorithms and technologies that produce predictions. In order to ensure the desired

accuracy in the predictions as well as in the selection of classifiers based on their performance, we introduce a very useful tool, which is called 'Receiver Operating Characteristic - ROC'. ROC work in a simple way, calculating the false positive (FP) and true positive(TP) values by shifting the decision threshold of each classifier. ROC curve, although it examines all the thresholds for a given classifier, does not return accuracy and recall, but shows the false positive rate versus the real positive rate. The diagonal of the curve is called random hypothesis and the models below it are considered worse than the models above it. Any point on this line means that the proportion of the correctly classified class samples is the same as the proportion of the incorrectly classified samples that do not belong to the class. Some points on the ROC space are of particular interest. Initially the lowest point on the left, the point (0,0) represents the strategy of never issued a positive classification. This means that a classifier produces neither false positives (FP) nor true positives (TP). On the other hand, regardless of circumstances only positive classifications are produced represented by the upper right point (1,1). The point (0,1) represents the perfect classification. Therefore, the closer to the upper left corner of the curve there is a classifier the better it is as it has a true positive value of one and a false positive value of zero. Conversely, the closer a classifier is to the diagonal, the less accurate it is. In essence, a ROC graph summarizes all of the confusion matrices that each threshold produced. Moreover, we can find the optimal threshold depending on how many false positives we are willing to accept.

2.2.4.5 AUC Area

AUC is an acceptable performance measurement system for a ROC curve. ROC curves can be a measure of presenting the family of the best decision limits for the relative costs of TP and FP. The area defined below the curve is a measure of the quality of noise-signal separation and is often used in the statistical inference of ROC curves. The AUC makes it easy to compare one ROC to another. The AUC is also closely related to the Gini index which is twice the space between the diagonal and the ROC curve.

$$Gini = 2 \times AUC - 1$$

2.2.5 Cross Validation

One of the most important factors in terms of the performance of the models we are called to construct is the way in which our data is divided into the training set as well as the testing set. More specifically, the initial set, for which we know for each observation the class label to which it belongs, is divided into a training set and a testing set. These two sets are subsets of the original data set. This separation is called partitioning and can be achieved in a number of ways which are explained below.

2.2.5.1 K-Fold Cross-Validation

Cross validation is a way to get a reliable estimate of the performance of a machine learning model. What we want to achieve is that we ask the model to adapt as best as possible to the training data. In essence, cross-validation is a statistical procedure of evaluating and comparing machine learning algorithms with data partitioning. The basic form of cross-validation is the **k-fold cross-validation**. Thus, in this type of cross-validation, the training and testing sets must be crossed in successive rounds so that each sample has a chance to be ratified. The process of applying cross-validation is relatively simple: Initially, the data is partitioned into k aspects of equal or nearly equal size. Then k training and testing repetitions are performed, so that in each repetition a different one aspect of the data is sufficient for validation while the remaining k-1 aspects are used for learning. So, having k training sets, k experiments are performed. Finally, the performance of each model is extracted by calculating the average of the yields measured in each experiment. The usual values given in k are the values 5 and 10. A common phenomenon is that cross-validation is considered to give a better assessment of a model's performance in practice than simply evaluating a system on a small portion of available training data, making cross-validation a popular way of evaluating machine learning models.

2.2.5.2 Leave One Out Cross Validation

This method is a special case of K-fold cross validation where the value of k (assuming the number of observations is n) is equal to the number of observations.

Thus, n-1 samples are used for training each time, while only one sample is used for testing. This procedure is repeated n times. This validation method is still widely used, especially when the available data set is very small in size.

2.2.5.3 Split Validation

The split method is the most classic and popular approach to data separation in training and test sets respectively. It is used to evaluate the performance of machine learning algorithms when used to make predictions for data not used for model training. The process is relatively involves taking a data set and dividing it into two subsets. The first subset is used to fit the model and is referred to as a training data set. The second subset is not used for model training. In this work we randomly selected 67% of the data set to be used as the training set and the remaining 33% for the test set.

2.2.6 Handling Class Imbalance

Basic machine learning algorithms fail to deal with unbalanced data problems, in the sense that they cannot classify the data set with great success. This is because the classification error in the majority class outweighs the classification error in the minority class. This dominance leads to the promotion of the decision function away from the majority class so as to reduce the classification error during the weight adjusting process. Therefore, the test set data is more often incorrectly classified in the minority class than those belonging to the majority class. This is a problem due to the reason that the machine learning algorithms work best when the number of occurrences in each class is about the same. There various techniques in order to handle class imbalance issues. These techniques can be summarized in the following levels:

- Data level
- Algorithmic level
- Cost sensitive level

- Feature selection level
- Enselmble level

In general, in order to handle class imbalance, as a rule of thumb we can perform the next steps that are depicted in the following scheme. When issuing class imbalance problems, we should focus either on the minority or the majority classes. If we choose the minority class path, we need to over-sample its data, in order to strike a balance. In the exact opposite case, we need to under-sample the overall data. In this work, class imbalance came up due to 'Lack of density'. That is a major issue that can arise in classification problems, for the reason that we are called to classify small amounts of data. This issue relates to lack of density or lack of information, where the machine learning algorithms do not have all the necessary data in order to make generalizations about the samples distribution.

2.2.6.1 Sampling Techniques

As treatment of class imbalance problem is resampling. Resampling works by changing the balances in the data set either by increasing the number of samples in the minority class or by reducing the number of samples in the majority class. The resulting data obtained after resampling is more balanced. These sampling methods therefore use such heuristics, which try to approach the optimal distribution of the samples, so that we can process and to draw reliable conclusions about the data. Although there are various resampling techniques and methods, we preferred to apply **'Random Oversampling'** and **'Random Undersampling'** methods [35, 36] in this work. More specifically, random undersampling tries to randomly delete examples in the majority class, where random oversampling attempts to randomly duplicate examples in the minority class. They are referred to as 'naive resampling' methods due to the reason that they do not use heuristics and there is no intuition about the data.

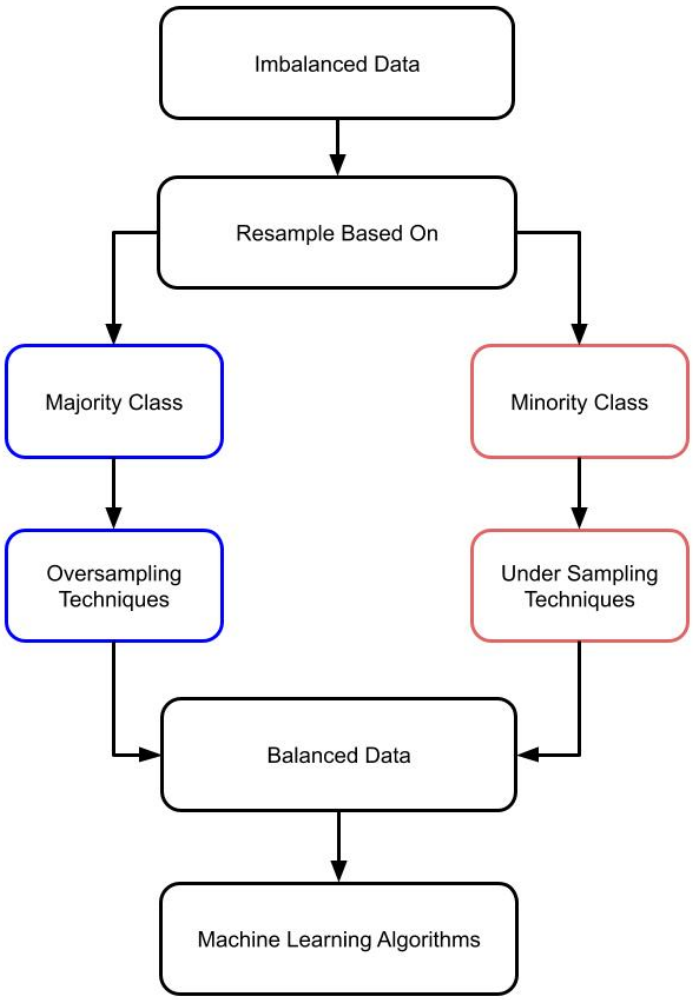


Figure 2.4: A rule of thumb

2.3 Human Activity Recognition

Human Activity Recognition (HAR) is a computer vision active research field that aims to classify human behaviour. More specifically, HAR plays a significant role in everyday human interactions and interpersonal relationships. It is a task of high interest within the field of ubiquitous sensing, with applications ranging of medical informatics, information security, etc. Therefore, recognizing daily life activities becomes quite useful in order to provide feedback to the user about their behaviour. Essentially, we can utilize this research area as part of a self-tracking culture in order to facilitate the individual understand patterns of activities he/she performs in

his/her daily life. In addition, the recognition of behavior can significantly contribute to the monitoring of various abnormalities in people suffering from mental illness with the ultimate goal of preventing side effects. HAR is categorised into **two** main categories [37]:

1. **Unimodal** Activity Recognition

   This category depicts human activities from a single modality (e.g. camera images). This category is further analyzed into the following categories:

   - Space-time methods
   - Stochastic methods
   - Rule-based methods
   - Shape-based methods

2. **Multimodal** Activity Recognition

   This category is a combination of information sources from multiple modalities (e.g a fusion of information coming from multiple sensors of different types). Multimodal Activity Recognition is divided into the following categories:

   - Affective methods
   - Behavioral methods
   - Social-networking methods

In this work, we focus on multimodal activity recognition methods and more specifically we aim to combine affective and behavioral methods, due to the reason that the first tend to recognize human activities based on the person's affective state and the latter due to the reason that tends to recognize non-verbal communication such as body movements. Then, according to O´scar Lara et al.[38] we present a formal definition for HAR.

Definition 2. *'Given a set* $W = \{W_0, W_1, \dots W_{m-1}\}$ *of m equally sized time windows, totally or partially labeled, and such that each* W_i *contains a set of time series* $S_i = \{S_{i,0}, S_{i,1}, \dots S_{i,k-1}\}$ *from each of the k measured attributes, and a set*

$A = \{a_0, a_1, \ldots a_{n-1}\}$ of activity labels, the goal is to find a mapping function $f : S_i \mapsto A$ that can be evaluated for all possible values of S_i such that $f(S_i)$ is as similar as possible to the actual activity performed during W_i.'

Chapter 3

Dataset and Proposed Architecture

This chapter describes in detail the criteria and the data collection process followed in this work.

3.1 Overview

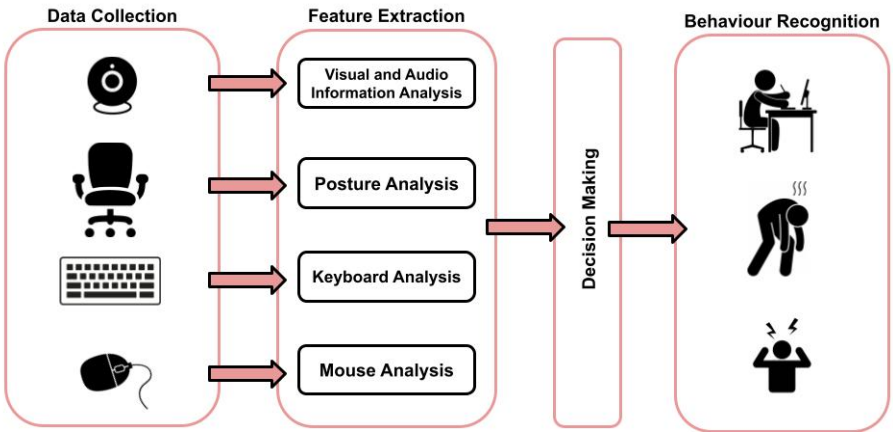


Figure 3.1: The Proposed Architecture

The diagram above depicts the proposed architecture. In order to achieve our goal, we propose a methodology that consists of four phases. First and foremost, we had to define a data collection process with the scope of data acquisition. Therefore, we had to build our own setup. To that end, aiming to develop a multimodal data collection methodology, we chose to use four modalities including:

- Web camera
- Force sensitive resistor sensors mounted on a chair
- Keystroke Dynamics
- Mouse

In the second phase, features are extracted for each modality. More specifically, we extract only features for the data coming from the sensors of the chair, keyboard and mouse. Although there are many algorithms for image and audio processing and analysis, the camera data was only used to verify the measurements that took place. Furthermore, based on FSR sensors, we are able to see the distribution of the user's weight and therefore find the posture of his/her body in the chair during working time. More than that, using keystroke and mouse dynamics we can indicate if an employee is experiencing stress, due to the reason that these types of dynamics are stress markers that emphasize the user's writing, speed, button force pressure and movement patterns. Thus, we conduct unimodal feature extraction for each modality. However, after the features are extracted, we aggregate on the timestamp that the data were collected. We should mention that the feature extraction is performed for a segment size of ten seconds in duration. In addition, we preprocess our dataset, due to class imbalance issues. In phase three, we select various classification algorithms that might adapt the system to address our problem for which the learning is being applied. In the last phase, we evaluate our architecture in relation to how well it manages to categorize the activities performed by an employee, in order to recognize their behaviour.

3.2 Components

3.2.1 Microcontrollers

A microcontroller is an integrated circuit that includes a processor, memory, and input and output which are programmable. Memory as well as RAM are built into the chip. The microcontroller is applicable to systems and automatic control devices such as self-propelled machines, light sensing and controlling devices, temperature

sensing and controlling devices, fire detection and safety devices and many others. A microcontroller has a low production cost and is small in size and for this reason can be numerically controlled in many devices and processes. Some use four-bit words and operate at clock speeds of 4 khz for low power consumption. Furthermore, there is the possibility of being idle and coming back into operation at the touch of a button. Something which makes them very useful because they can operate for long periods with the use of a battery. For this dissertation the model chosen is the Arduino Uno.

3.2.2 What is Arduino

Arduino [1] is a family of open-source electronic microprocessors, where programming is easily implemented and interaction with the external environment is possible. Specifically, Arduino is an open source computing platform with integrated microcontroller and inputs/outputs that are programmed in the Wiring language. The main advantages of using Arduino is that this platform is an economical solution because it has a low cost. In addition, it is architecturally open and can be developed by anyone. Furthermore, this platform is fully portable which allows it to be programmed on most operating systems. It is a fully scalable platform for the reason that the hardware and software are open to everyone.

3.2.2.1 Arduino Models

The Arduino models on the market are presented in the table below. It is important to mention that, although there are a number of options for the Arduino microcontroller, each choice depends on the type and nature of the problem being implemented.

- Serial Arduino
- Arduino DUE
- Arduino Esplora

[1] `https://www.arduino.cc/`

- Arduino Leonardo
- Arduino MEGA 2560
- Arduino MEGA
- Arduino UNO
- Arduino Duemilanove
- Arduino Diecimila
- Arduino Bluetooth
- Arduino NG plus
- Arduino NG
- Arduino Nano
- Arduino Mini
- Arduino Extreme

3.2.2.2 Arduino UNO Characteristics

Below we briefly present the characteristics of Arduino UNO.

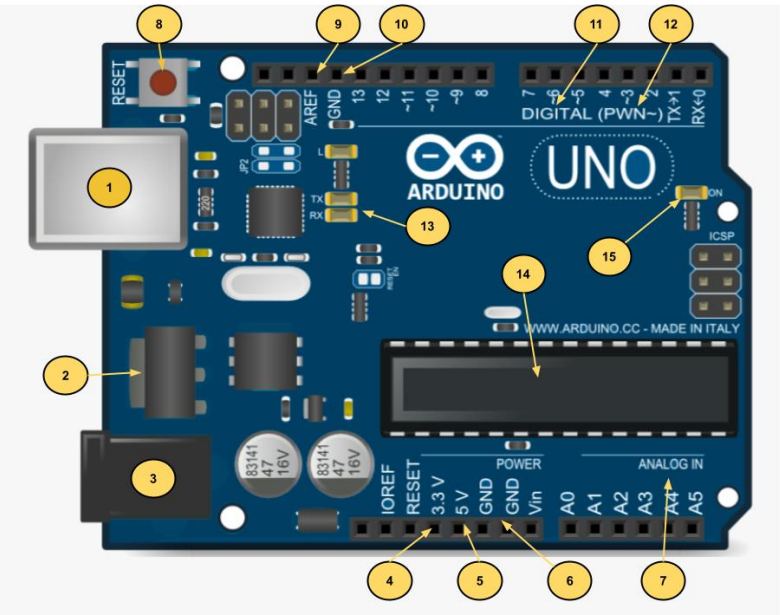


Figure 3.2: Arduino UNO board analysis

1. **USB connection:** Used for programming or power supply
2. **Voltage stabilizer**
3. **Jack DC socket:** Used to power Arduino from a power supply
4. **3.3V pin:** Provides 3.3 Voltage
5. **5V pin:** Provides 5 Voltage
6. **Ground pin**
7. **Analogue pins:** These terminals can read analogue signal
8. **Reset button:** Restarts the code that is loaded in the Arduino board
9. **AREF:** Supports 'Analog Reference' and is used to set an external reference voltage
10. **Ground terminal**
11. **Digital input/output:** Pins 0-13 can be used for digital input or output
12. **PWM:** Pins marked with the symbol (~) can simulate the analog output
13. **TX/RX:** Transmit and receive LED indicator
14. **ATMEGA 328P Microcontroller**
15. **Power light:** LED lights every time the board is connected to a power source

3.2.3 Force Sensitive Resistor (FSR) Sensors

Pressure sensors are designed to measure the presence and relative magnitude of pressure in an area. The resistance of the sensor increases and decreases while it depends on the force exerted. When no pressure is applied to the sensor, its resistance is greater than MΩ. As we increase the pressure exerted on the sensor, the resistance between the two terminals of the sensor decreases [39] These sensors, which are extremely simple to use for both the connection mode and programming which require in combination with a microcontroller that reads the sensor values.

Sensors of this type are made of polymer ink which is a good conductor of electricity. The region recognized by the pressure comprising two types of particles: particles that are good conductors of electricity and particles which are poor conductors of electricity. In particular, the force pressure sensor is divided into two different layers, between which there is an additional separation surface. When pressure is applied, the particles move closer to each other, thus changing the resistance between the sensor terminals. Essentially, force pressure sensors are therefore a variable resistor that varies in value depending on how much pressure is applied to the sensor. The disadvantage of the sensor is that the difference between the measurements can be greater than 10 %. Also, a feature of these sensors is that the interval under which the sensor perceives pressure can be adjusted. That is, the interval is the smallest and maximum pressure value that the sensor can perceive. The shorter the interval, the more sensitive and accurate it will be in the measurements. No pressure can be measured outside the space and there is a possibility of sensor failure. The sensor we use can perceive from 100 grams to 10 kg, but can not understand the difference between 11 kg and 20 kg. In the diagram below, we observe the relationship between pressure and resistance. The ratio between the two is mostly linear of 50 grams or more. However, to activate the sensor, a minimum power value of about 100 grams is required in order for the resistance to become linear and fall below 10 kΩ.

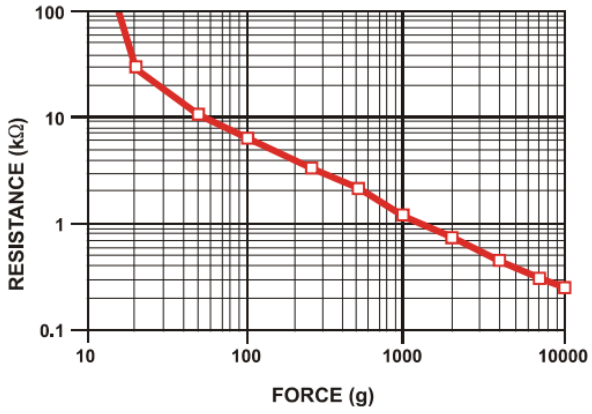


Figure 3.3: Resistance versus Force

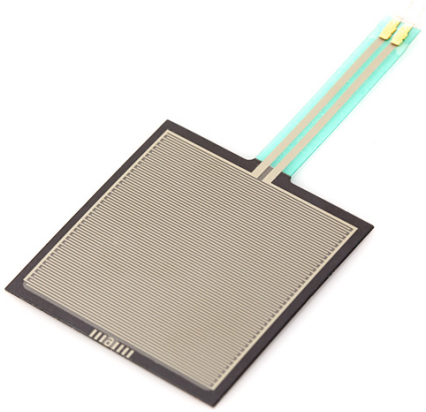


Figure 3.4: An FSR sensor

3.2.3.1 Why use Force Sensitive Resistor Sensors

FSR sensors are probably the best choice in terms of the quality of results and cost, due to the reason that this type of sensors are low cost equipment, especially if we take into considerations their small size and simplicity. The fact that they are very small, flat and require minimal wiring, is an extremely important factor, as these sensors will mounted on a chair surface and therefore should be as invasive as possible. The only drawback we found, is that sensors of this type sometimes respond slowly to the force applied to their surface, which increases the likelihood of incorrect pressure measurement in real time.

3.2.4 Experimental Setup

3.2.4.1 Computer Peripherals

In order to set up our experiments, we used a set of a simple keyboard and mouse, a simple web camera and the FSR sensors mounted on a chair.

3.2.4.2 Smart Chair for Measuring Stress and Anxiety Levels

In conditions where employees perform office work it is observed that musculoskeletal health problems often occur such as backache, back pain and neck pain, due to poor posture. In addition, problems occur in muscles, nerves, ligaments, tendons, and

Figure 3.5: Our Setup

articular cartilage, which can aggravate pre-existing conditions such as tendonitis, muscle fractures, and chondropathies, turning them from acute to chronic conditions, which are now difficult to treat.

The term 'posture' refers to the position of the body: how a person holds himself when he is standing, sitting or lying down. In orthopedics, posture also refers to how muscles, joints, and the spine work together to keep a person upright. Poor posture is considered an asymmetrical body position. For example, the occurrence of an excessive curve in the lumbar spine of a person, which is considered a bad posture. Poor posture occurs when a person's daily activities lead some muscles to become strong while others become weak. This imbalance of muscle strength can lead the body to a bad posture. In essence, some employees develop posture problems as a result of an illness. However, frequently, changes in attitude resulting from stress and overexertion to daily activities. One of the factors which are responsible for causing problems in posture is the working places and environments which contribute to the creation of biochemical reactions in the body such as increased cortisol levels. This chain reaction can lead to high levels of stress and anxiety.

Based on the above, a smart chair was developed in a standard form, in order to continuously monitor the stress rate of employees in working environment conditions. For the implementation of this project, five FSR sensors were used, which were placed in a common office chair and connected to an Arduino UNO board. Special software was then developed to record continuous pressure measurements over time, in order to draw conclusions about the degree of stress an employee experiences.

3.2.4.3 Construction Materials

The construction of the smart chair required the following materials, which are summarized in the following table:

Description	Quantity
Carbon resistors of 10Ω	5
Jumper Wires male to female (30cm)	40
10mm LED lamps (different colors)	5
Breadboard (830)	1
Jumper Wires male to male	10
USB cables (USB A to USB B)	1
Arduino UNO	1
Force Sensitive Resistor Sensosr	5

Table 3.1: Construction materials

3.2.4.4 FSR Installation

In order to position the sensors optimally, we had to take into account the posture of a user based on the following situations:

1. What is the posture of the average user in a period when he is at rest (does not experience stress)
2. What is the average user posture when in situations experiencing intense stress
3. What is the posture of the average user when he is in situations where there is work fatigue
4. Combination of all the above

3.2.4.5 Experimental Tests

In the end, we came to the conclusion that the maximum number of FSR sensors that had to be installed was five sensors. In particular, sensors were placed at each end of the chair (front left, front right, rear left, rear right), as well as in the middle

Figure 3.6: Smart chair prototype

of it. The central idea is that once there is a user sitting in the chair, a weight distribution (Newton) is created that comes from the measurements of the sensors being recorded. On the substance, based on the maximum weight or the maximum values recorded by each sensor, we can know how the user sat down. For instance, if a user is sitting in the front of the chair then the measurements of the sensors located front left and front right will have higher measurement values for the time during which the user is sitting at that point in the chair. Thus, we construct the weight distribution diagram which will show us the concentrations of the highest values of the sensors for that interval. These values can tell us that the user is in this position, because something has caught his attention and he is sitting crouched in front of the computer screen, which can indicate stress-anxiety. More specifically, our goal is, since we take measurements from sensors, to model the posture of an employee. This means that the process of data collection (measurements from the sensors) must take place in order to create the necessary set of data which will simulate some predetermined postures of an employee's body. Then, our goal is to train machine learning algorithms, in order to be able to develop software that can 'guess' an employee's posture based on the assumptions that have been modeled (for example if there is a weight distribution in the front of the chair the employee may be experiencing stress).

3.3 Problem Definition and Data Collection

The last decade has seen the release of several datasets in the fields of psychology and HAR. These datasets have played a major role in understanding research for HAR. However, there is no dataset that fuses unimodal information in order to extract knowledge. Usually, this is because multimodal datasets (which are typically comprised of images, sensor data, GPS data, etc.) are expensive to collect and annotate, due to the difficulties that arise in integrating, synchronizing, and calibrating the modalities involved. For the reasons above, and for the sake of the evaluation of the proposed method we constructed our own real-life dataset. More specifically the data have been recorded in office conditions. This means that we wanted to simulate a working environment in order for our data to be as realistic as possible. Unfortunately due to the complexity of the setup in combination with the pandemic it was impossible to use different setups with different users. For this reason the data collection was performed by one person. We defined a set of 'office activities' and that was one of the main objectives of the test in order to determine which classes of a user's activity could be recognised. Thus the following classes where selected:

0. **Coding**
1. **Writing Email/Report**
2. **Browsing/Scrolling Social Media**
3. **Communicating**
4. **Absent**

Thus, 25 recordings have taken place in a period of a month and each recording was around 90 seconds of average duration. Each recording have been manually annotated by the user that also performed the recording. The recording process is very simple. The user has two options:

1. Start a bash script that records data for 90 seconds

2. Start a bash script that records for a specific amount of time (for recordings longer in duration)

Initially, tests were conducted on whether it is possible to use recording samples from the chair mechanism in combination with the audio/video, mouse and keyboard streams. In order to achieve good cross-modality data alignment between the FSR sensors and the user's keystrokes and mouse movements, we aggregate the sensor data based on timestamp. The timestamp of the FSR sensors is the exposure trigger due to the reason that chair data is recorded throughout the session and mouse and keyboard data is logged only when used by the user. Given that the chair's exposure time is nearly instantaneous, this method generally yields good data alignment. Therefore, in order to ensure that the mechanism works, a visualization tool was created. The following figures demonstrate the recorded snapshots.

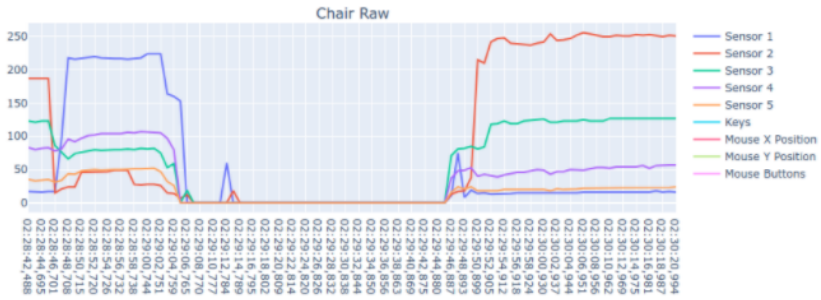


Figure 3.7: Chair raw data

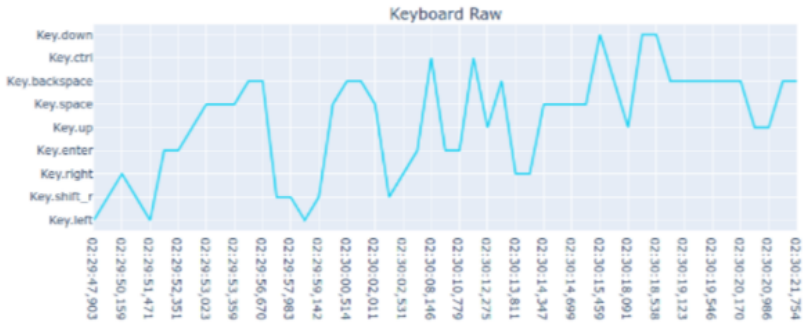


Figure 3.8: Keyboard raw data

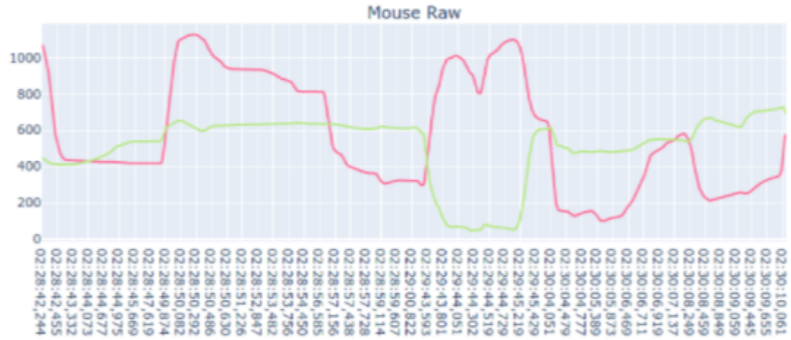


Figure 3.9: Mouse raw data

3.4 Multimodal Representation and Feature Extraction

In this section we provide a brief description of the handcrafted features that we have used in this work. The complete list of features is depicted in the following sections. We should mention that the set of features can be extracted on a specified second segment size. More specifically, we used 10-second segment size. Regarding the methodology for extracting features, we used information that came from the combination of the measurements of the sensors that were mounted on the chair as well as measurements that were based on keystroke and mouse dynamics. With regard to the export of mouse and keyboard functions, we have complied with the general data protection regulation for the protection as well as for ethical reasons, we respect the privacy legislation [40]. Therefore, we do not record all the data flow from the keyboard and mouse, but we record a combination of neutral keys that in no way identify a person. Thus, in order to build the feature vector that serve as the input for the machine learning algorithms, we extract features for each modality separately and finally we aggregate on timestamp. In this way, for each sample, we construct a characteristic 1 * 20 dimension vector in which the first 6 attributes refer to the mouse, the next 4 to the keyboard and the next 10 to the chair.

3.4.1 Extracting Mouse Features

Mouse usage is usually referenced as 'mouse dynamics'. As we mentioned above, mouse dynamics can be a very useful tool in combination with other types of features,

in order to create patterns that recognize the activity of users. In mouse dynamics the time taken to detect a user's pattern is dependent on the number of mouse events, such as the total number of clicks, the speed of mouse cursor and so on. In terms of data collection, characteristics arise either from individual activities or from a combination of activities. For instance, when a user scrolls to the social media, the movements made by the mouse is quite specific and thus can relatively easily produce the corresponding characteristics. But when, for example, the user communicates with one of his/her colleagues, there is the possibility of using the mouse either because for example scrolls to show something to someone else or does not use the mouse, either can write code and at the same time read his/her e-mail.

In order to collect the mouse data, we construct a log file that contains a set of lines and each line depicts a recorded mouse event. More specifically, a mouse event is a tuple that consists of six attributes: The first two fields represent the recorded action timestamp. The action field depicts the mouse event that can take the following buttons and values:

$$\langle \mathbf{Date}, \mathbf{Time}, \mathbf{Action}, \mathbf{PosX}, \mathbf{PosY}, \mathbf{Button} \rangle$$

Mouse event buttons:

1. Left click
2. Right click
3. Middle click

Mouse event button values:

1. CLK: Button click
2. RLS: Button release
3. MV: Move
4. SCRD: Scroll down
5. SCRU: Scroll up

PosX and PosY fields represent the coordinates of the cursor on the screen. Finally, the button field informs us about the button that was pressed during the recording

session. We need to mention that movement of X-axis ranges either from left to right, or right to left respectively. Adding to that, whenever a user moves their mouse to the right or left, X-axis movement will take different pixel values accordingly. On the contrary, is a Y-Axis movement, and it ranges from upside to downwards or from downward to upward. Based on the average of X-Axis movement and Y-Axis movement, we calculate the speed values with the pixels per second for values coming from the X or the Y axis. The extracted features for the mouse are summarized in the following table:

Mouse Features	
Feature Name	**Description**
Clicks	Total number of clicks
R_Clicks	Total number of right clicks
L_Clicks	Total number of left clicks
M_Clicks	Total number of middle clicks
Velocity_X	Speed in X-axis
Velocity_Y	Speed in Y-axis

Table 3.2: Mouse Features

3.4.2 Extracting Keyboard Features

In contrast to mouse usage, keyboard usage is usually referenced as 'keystroke dynamics'. As we have previously indicated, the use of characteristics derived from the keyboard may prove very useful when this knowledge combined with features from other sources. In keystroke dynamics the time taken to detect a user's pattern is dependent on the number of keyboard events, such as the total number of buttons pressed, the typing frequency, the time elapsed between keyboard presses, etc.

In order to collect the keyboard data, we construct a log file that contains a set of lines and each line depicts a recorded keyboard event. More specifically, a keyboard event is a tuple that consists of three attributes:

$$\langle \mathbf{Date}, \mathbf{Time}, \mathbf{Key} \rangle$$

The first two fields represent the recorded action timestamp. However, keystroke

data may contain sensitive information such as passwords, user names, etc. For this reason we log only a set of buttons. The key field depicts the keyboard event that can take the following buttons:

1. Alt
2. AltGr
3. AltR
4. Backspace
5. Space
6. Ctrl
7. CtrlR
8. Down arrow
9. Up arrow
10. Left arrow
11. Right arrow
12. Page down
13. Page Up
14. Enter
15. Shift
16. ShiftR

The extracted features for the keyboard are summarized in the following table:

Keyboard Features	
Feature Name	**Description**
All_keys_N	Total number of arrow, alt, control, shift, backspace, space, enter, page up/down keys
Arrow_keys_N	Total number of arrow keys
Spaces_N	Total number of space keys
Shft_Ctrl_Alt_N	Total number of shift, control, alt keys

Table 3.3: Keyboard Features

3.4.3 Extracting Chair Features

As we mentioned earlier, posture analysis can serve as a metric that assists in HAR. We can achieve our goal by examining the weight distribution of an employee during a workday. This kind of measurements aid to measure the fatigue that an employee experiences, based on the variations in the values received from the sensors. We

believe that the characteristics that arise from these measurements, in combination with the characteristics that have previously been reported, can help us more easily identify the emotional state of the person being monitored by this methodology. Moreover, in order to collect the chair data, we construct a log file that contains a set of lines and each line depicts a recorded chair event. More specifically, a chair event is a tuple that consists of ten attributes:

$$\langle \mathbf{Date}, \mathbf{Time}, \mathbf{A_0}, \mathbf{A_1}, \mathbf{A_2}, \mathbf{A_3}, \mathbf{A_4} \rangle$$

Thus, ten different features are proposed, in order to detect the sitting posture of an employee. We calculate these feature based on the arithmetic mean and the standard deviation for each pressure sensor during a specific amount of time and more specifically during a time segment. The extracted features are summarized in the following table:

Chair Features		
Sensor ID	**Feature Name**	**Description**
0-4	Mean_A{ID}	Average for Sensor_{ID}
0-4	STD_A{ID}	Standard Deviation for Sensor_{ID}

Table 3.4: Chair Features

3.4.4 Fusing Features

Nowadays, fusing data from multiple sensor sources, is therefore becoming a major research trend that directly affects application performance, improves reliability and accuracy of the classification process. Therefore, more and more HAR systems involve multimodal information extraction and fusion. More specifically, the extracted unimodal feature sets can be fused in order to create better representations of high-dimensional (multimodal) feature vectors that serve as input for the classification step. In this work, we applied early fusion methods due to nature of our methodology that involved supervised learning techniques. According to Snoek et al [41]

Definition 3. *'**Early** fusion, is a scheme that integrates unimodal features before*

the learning process.'

A better understanding may be given based on the following scheme:

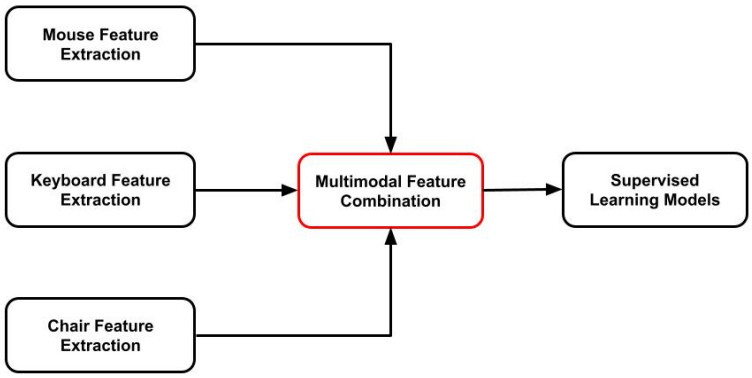


Figure 3.10: Early Fusion

A major disadvantage of this approach is the difficulty in feature combination into a common representation. Moreover, synchronization of the data must be achieved via timestamps for the calculations to be accurate and time aligned. The last challenge is the missing data from one or more sources. Some precautions must be taken such as using probabilistic data fusion algorithms, imputation or removal of the missing time intervals.

3.4.5 Post-processing Features

After combining the exported features from the different sources, and aligning them over time, we create the data set that we will use to train the classifiers. In many cases, as in our case, there were features which take values in different intervals. In this case, we can address this problem by normalizing these features in order to obtain values in similar fields. The method we used to normalize our data is standardization, that is, all the resulting normalized data will have zero mean value and unit dispersion. This was achieved by using the Standard Scaler of Scikit-learn toolbox. So in order to standardize our data set we used the following formula: For available data pf a feature m, we have:

$$\overline{x}_m = \frac{1}{N}\sum_{i=1}^{N} x_{im} \;\; for\, m = 1, 2, \ldots, l$$

$$\sigma_m{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{im} - \overline{x}_m)^2$$

$$\hat{x}_{im} = \frac{x_{im} - \overline{x}_m}{\sigma_m}$$

Moreover, due to the reason that some features may be unavailable during the calculations or their values may be missing, we decided to fill those values with zeros.

Chapter 4

Experiments

4.1 Introduction

In order to study stress in a non-invasive way and using simple sensors as well as the flow of information from the mouse and keyboard, we developed the methodology presented in the third chapter. The development was done using the Python 3.8 programming language. The most notable libraries used are the following:

- **Plotly:** [1]
  Plotly is a clear syntax interactive visualization tool that offers various graph types.

- **scikit-learn:** [2]
  Scikit-learn is an open-source python module, that includes a lot of machine learning algorithms and complex functions for pre-processing and manipulating data.

- **imbalanced-learn:** [3]
  Imbalanced-learn is an open-source python module, that includes tools in order to deal with classification tasks on imbalanced data sets.

[1] https://plotly.com/
[2] https://scikit-learn.org/stable/
[3] https://imbalanced-learn.org/stable/

- **schedule:** [4]

  Python job scheduling for humans. It is the python alternative for Cron.

- **pandas:** [5]

  Pandas is a tool for data management and manufacturing metrics and graphs into a program.

- **NumPy:** [6]

  NumPy is a Python library suitable for management large and multidimensional tables using complex mathematical functions.

The open source code [7] that implements the baseline method described in the previous chapter can be also used in the experimentation process. We have experimented with a set of widely used classifiers, deriving from the Scikit-learn toolbox including:

- Support Vector Machines [31]
- Decision Trees[citation] [32]
- K-nearest neighbor classification [34]
- Naïve Bayes (NB) [33]

As mentioned at length in the previous chapter, we designed a data collection process that consisted of a simple keyboard and mouse as well as power sensors mounted on a simple office chair. The purpose of the experiment we performed was to test the performance of the algorithms. In essence, our goal was to be able to examine whether each of the above algorithms was able to distinguish the activities performed by the user each time. The procedure we defined was as follows: The subject was sitting in a chair (in which the FSR sensors were placed), in front of an office on which there was a camera that was at face height as well as a conventional mouse and a keyboard. As previously mentioned, the camera data was used only to evaluate the recordings as well as to assist in being able to time align the recorded data from the other sources. The user performed the aforementioned activities either individually

[4] https://pypi.org/project/schedule/
[5] https://pandas.pydata.org/
[6] https://numpy.org/
[7] https://github.com/amitsou/Multimodal-User-Monitoring

or in combination using a bash script that recorded data for 90 seconds. During the experiment, the user was asked to be as focused as possible on the activities they were performing, without however being particularly rigorous, in order to achieve the most realistic conditions possible. Once a set of measurements was completed, it was not necessary to repeat after a few seconds or minutes, but each subsequent measurement was made at different times within a twenty-four hour or within the next twenty-four hours, in order to change the user's emotional state. With this we wanted the user to feel differently the degree of stress as well as the degree of fatigue.

4.1.1 Experimental Evaluation

At this point we present the results of the experiments we described in the previous section. For the experimental evaluation of the proposed methodology, we performed two series of extensive experiments. More specifically, we performed experiments at recording dependent level, in terms of the first category and for the second category we performed experiments at recording independent level. This essentially means that classification tasks were made for each category, which includes binary classification as well as classification for all classes. In addition, in each type of classification and category, experiments were performed on unbalanced and balanced data sets. This way, we aimed to study the generalization of our approach. The experimental evaluation methodology is depicted in the following diagrams.

In this work, we chose to split the data at segment level of 10-seconds each. Now, let us focus on the way the data was split into training and testing subsets. Regarding the separation of data for the case of the recording dependent level, since we did not have a good heuristic to split the users in train and taking into account that there was a class imbalance problem, we considered that a good solution would be to use the 33% of the data for the training set, while the remaining 67% samples of the rest were kept for the construction of the test set. On the other hand, regarding the separation of the data for the case of the second experimental series, we chose to split the data at segments of 10-seconds each, with the difference that we made the following addition. We separated the data set, in terms of recording independent level. In

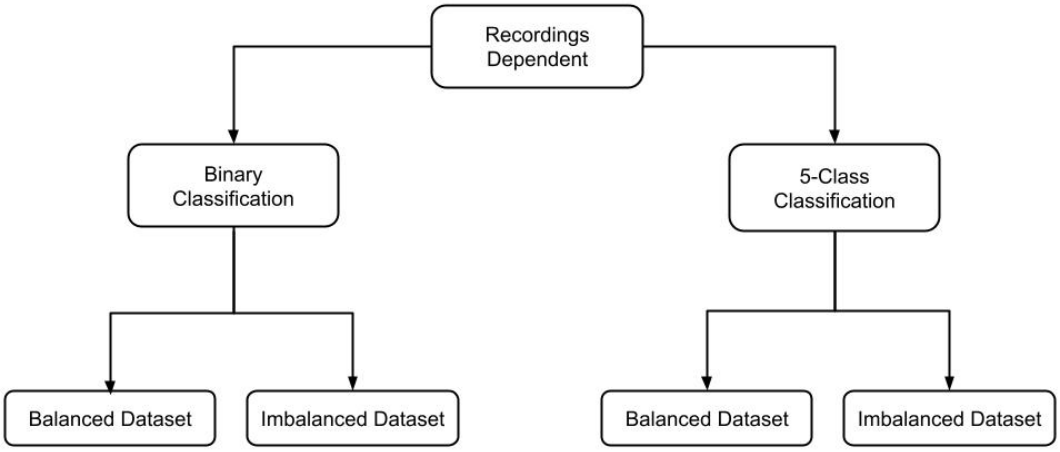


Figure 4.1: Diagram: Recordings dependent experimental series

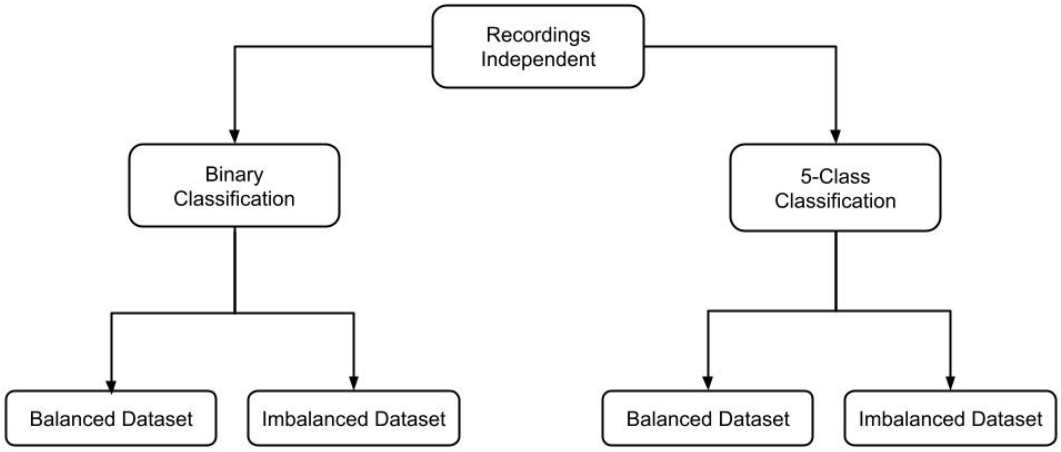


Figure 4.2: Diagram: Recordings independent experimental series

that way, different 10-sec segments of the same recording cannot belong to both the training and test sets at the same time, since that would introduce significant bias in the results, as the classifiers would be 'recording-dependent'. Under that constraint, 50% of the overall data were used for training the training set and the remaining 50% for the test set. More specifically, of the 24 recordings, 12 recordings were used for the training set, while the rest 12 were used for the test set.

4.2 Results

From the aforementioned performance metrics that we presented in chapter two, F1-macro average provides the most suitable evaluation metric of the classification tasks that were conducted, due to the reason that it takes into account the class imbalance problem. Thus, we present the confusion matrices of the best model as well as the ROC curves. Regarding the problem of binary classification, we show only those classes that achieved F1 measure more than 80%. On the other hand, we present all the results for the classification problem of all classes (5-Class Classification). Moreover it is important to note that for every experiment, all the algorithms used had their default parameters and no parameterization was performed.

4.3 Binary Classification: Recording Dependent Evaluation

4.3.1 No Resampling

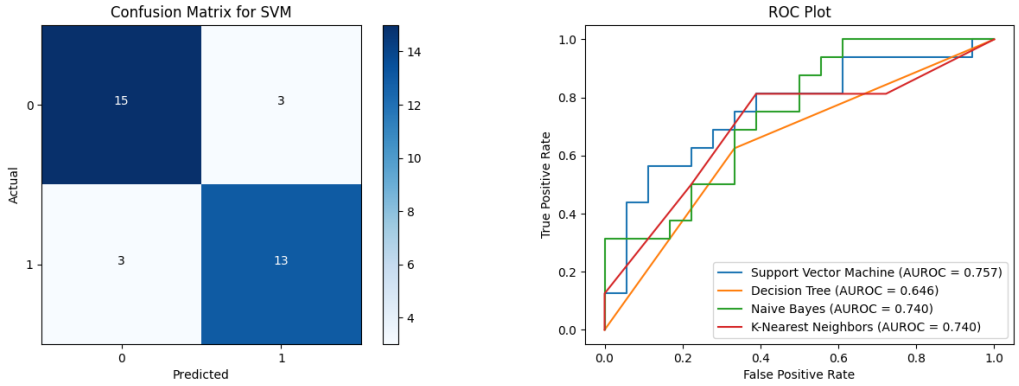
First and foremost, we start our experiments based on the imbalanced data set. The binary classification experiments that appeared to achieve F1-measure over 80% include the following pairs of classes:

4.3.1.1 Task: Coding VS Writing Email/Report

The model that managed to better separate Coding versus Writing Email/Report was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

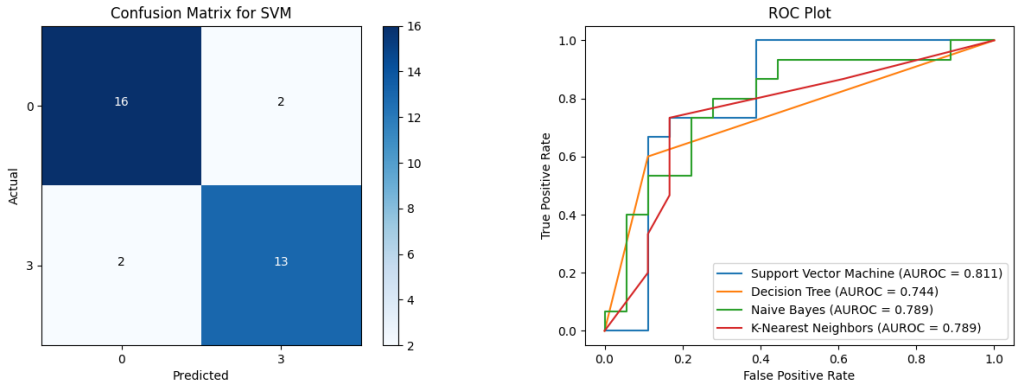
4.3.1.2 Task: Coding VS Communicating

The model that managed to better separate Coding versus Communicating was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.3: Confusion Matrix and ROC Curve for Coding VS Writing Email/Report



(a) Confusion Matrix (b) ROC Curve

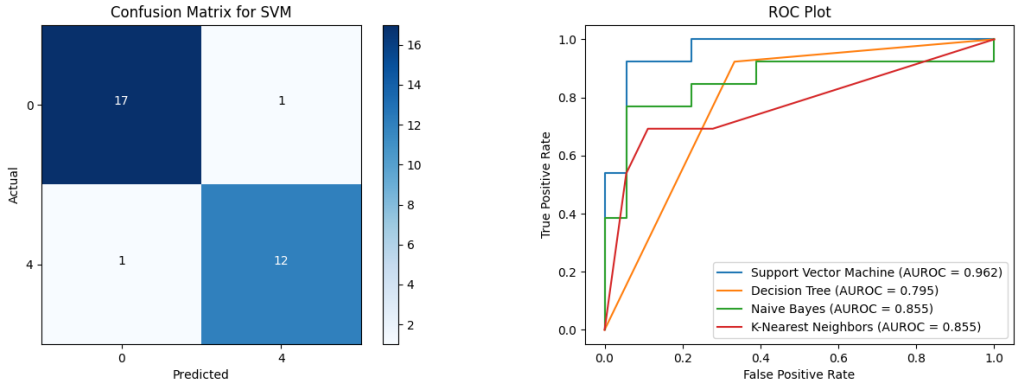
Figure 4.4: Confusion Matrix and ROC Curve for Coding VS Communicating

4.3.1.3 Task: Coding VS Absent

The model that managed to better separate Coding versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

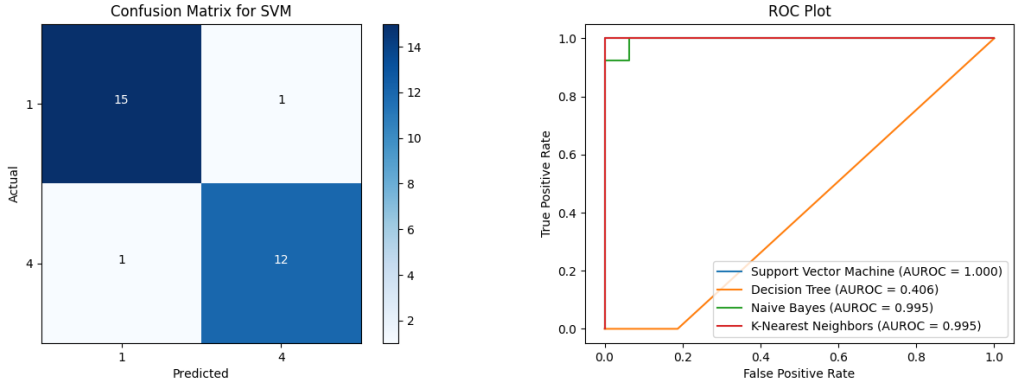
4.3.1.4 Task: Writing Email/Report VS Absent

The model that managed to better separate Writing Email/Report versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.5: Confusion Matrix and ROC Curve for Coding VS Absent



(a) Confusion Matrix (b) ROC Curve

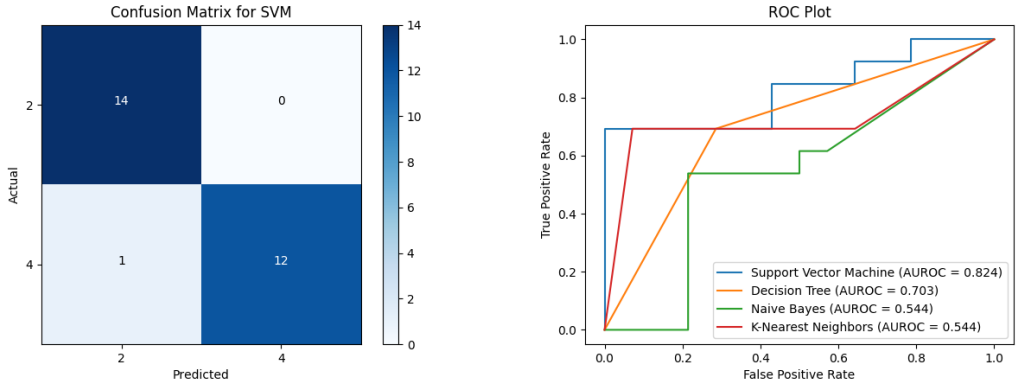
Figure 4.6: Confusion Matrix and ROC Curve for Writing Email/Report VS Absent

4.3.1.5 Task: Browsing/Scrolling Social Media VS Absent

The model that managed to better separate Browsing/Scrolling Social Media versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

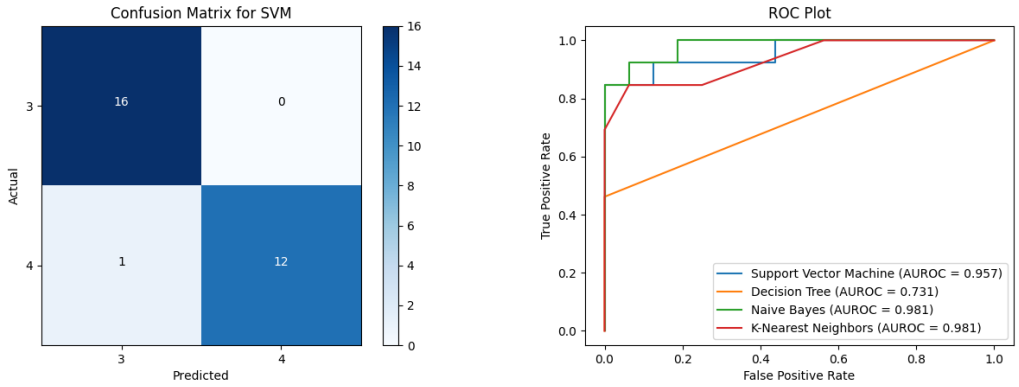
4.3.1.6 Task: Communicating VS Absent

The model that managed to better separate Communicating versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.7: Confusion Matrix and ROC Curve for Browsing/Scrolling Social Media VS Absent



(a) Confusion Matrix (b) ROC Curve

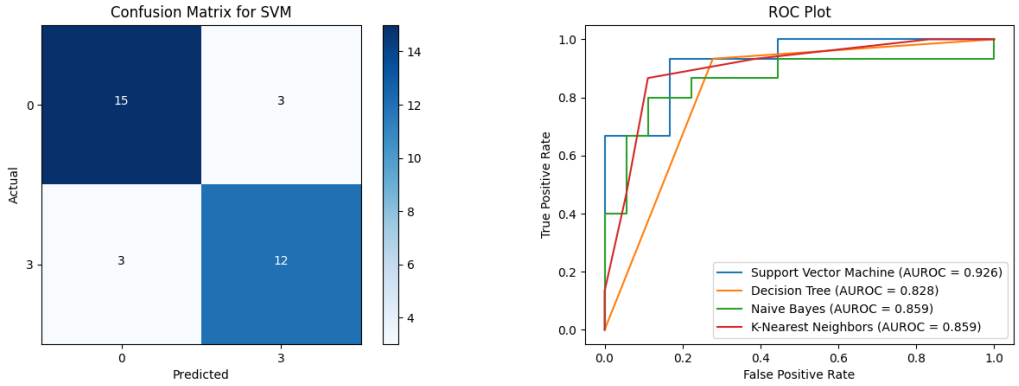
Figure 4.8: Confusion Matrix and ROC Curve for Communicating VS Absent

4.3.2 Random Undersampling

In this section, we are experimenting based on the balanced data set. More specifically, we used the 'random undersampler' from the imblearn API. The binary classification experiments that appeared to achieve F1-measure over 80% include the following pairs of classes:

4.3.2.1 Task: Coding VS Communicating

The model that managed to better separate Coding versus Communicating was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

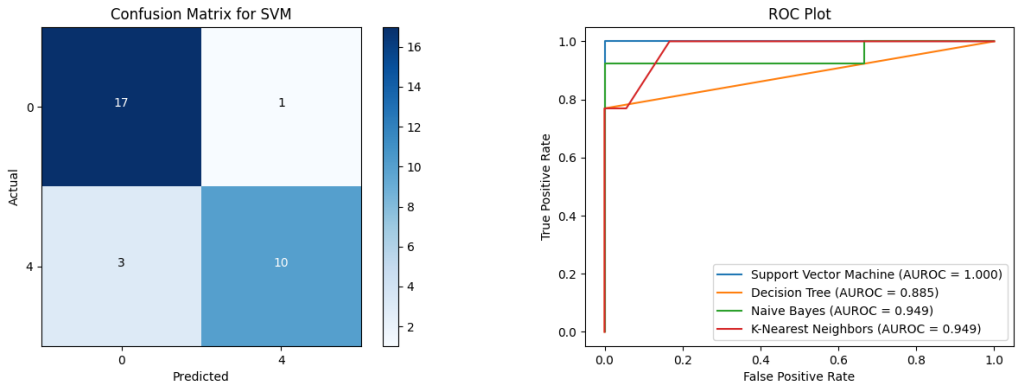


(a) Confusion Matrix (b) ROC Curve

Figure 4.9: Confusion Matrix and ROC Curve for Coding VS Communicating

4.3.2.2 Task: Coding VS Absent

The model that managed to better separate Coding versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

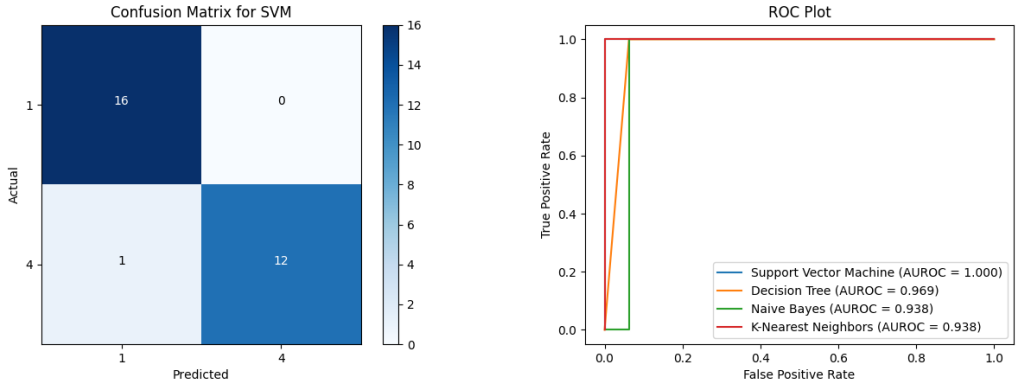


(a) Confusion Matrix (b) ROC Curve

Figure 4.10: Confusion Matrix and ROC Curve for Coding VS Absent

4.3.2.3 Task: Writing Email/Report VS Absent

The model that managed to better separate Writing Email/Report versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

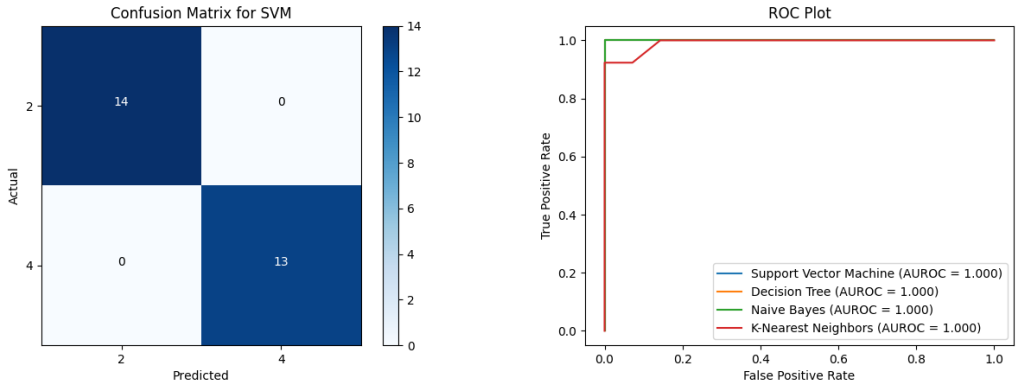


(a) Confusion Matrix (b) ROC Curve

Figure 4.11: Confusion Matrix and ROC Curve for Writing Email/Report VS Absent

4.3.2.4 Task: Browsing/Scrolling Social Media VS Absent

The model that managed to better separate Browsing/Scrolling Social Media versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

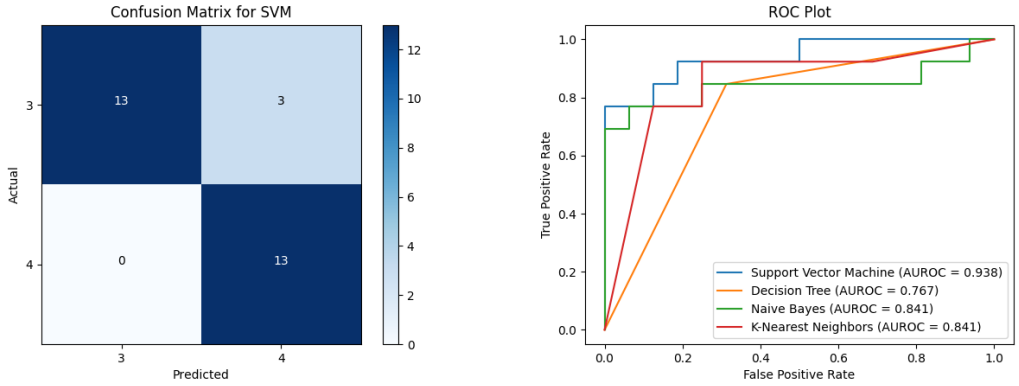


(a) Confusion Matrix (b) ROC Curve

Figure 4.12: Confusion Matrix and ROC Curve for Browsing/Scrolling Social Media VS Absent

4.3.2.5 Task: Communicating VS Absent

The model that managed to better separate Communicating versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.13: Confusion Matrix and ROC Curve for Communicating VS Absent

4.4 5-Class Classification: Recording Dependent Evaluation

In this section, we continue our experiments based on the imbalanced data set. We present the best models in the effort to separate all classes of user activities.

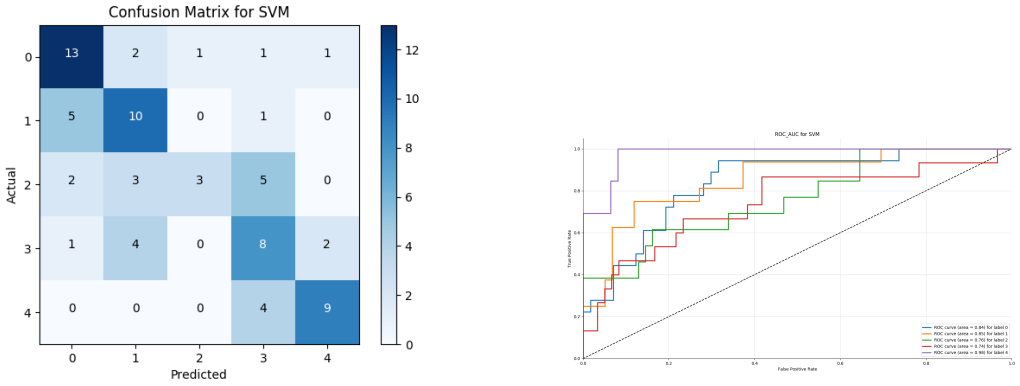
4.4.1 No Resampling

4.4.1.1 Task: All class classification

The model that managed to better separate all the classes was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

4.4.2 Random Undersampling

On the other hand, we continue our experiments based on the balanced data set. More specifically, we used the 'random undersampler' from the imblearn API. We present the best models in the effort to separate all classes of user activities.

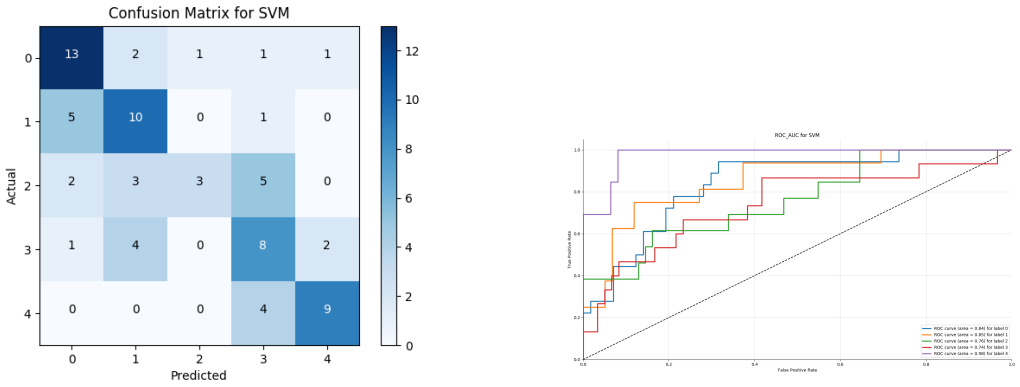


(a) Confusion Matrix (b) ROC Curve

Figure 4.14: Confusion Matrix and ROC Curve for 5-Class Classification

4.4.2.1 Task: All class classification

The model that managed to better separate all the classes was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.15: Confusion Matrix and ROC Curve for 5-Class Classification

At this point we quote the tables with the exact harmonic means for all the aforementioned experiments that are dependent on recordings. First we list the tables for the problems of binary classification based on balanced and unbalanced data sets and then we continue listing the tables for the corresponding classification problem of all classes.

Binary Classification Recording-Dependent Evaluation			
Experiment	**Class1 F1**	**Class2 F1**	**Classifier**
Coding vs Writing Email/Report	83%	81%	SVM
Coding vs Communicating	89%	87%	SVM
Coding vs Absent	94%	92%	SVM
Writing Email/Report vs Absent	94%	92%	SVM
Browsing/Scrolling Social Media vs Absent	97%	96%	SVM
Communicating vs Absent	97%	96%	SVM

Table 4.1: F1-measure scores for binary classification on imbalanced data set

Binary Classification Recording-Dependent Evaluation			
Experiment	**Class1 F1**	**Class2 F1**	**Classifier**
Coding vs Writing Email/Report	83%	80%	SVM
Coding vs Absent	89%	83%	SVM
Writing Email/Report vs Absent	97%	96%	SVM
Browsing/Scrolling Social Media vs Absent	100%	100%	SVM
Communicating vs Absent	90%	90%	SVM

Table 4.2: F1-measure scores for binary classification on balanced data set using imblearn

5 Class Classification Recording-Dependent Evaluation						
Experiment	**Class1 F1**	**Class2 F1**	**Class3 F1**	**Class4 F1**	**Class5 F1**	**Classifier**
All Classes	67%	57%	35%	47%	72%	SVM

Table 4.3: F1-measure scores for 5-class classification on imbalanced data set

5 Class Classification using Recording-Dependent Evaluation						
Experiment	**Class1 F1**	**Class2 F1**	**Class3 F1**	**Class4 F1**	**Class5 F1**	**Classifier**
All Classes	55%	52%	22%	48%	69%	SVM

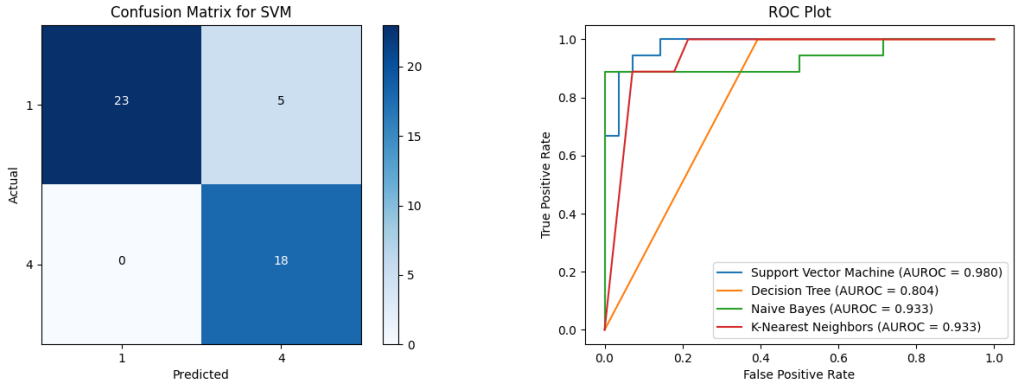
Table 4.4: F1-measure scores for 5-class classification on balanced data set

4.5 Binary Classification: Recording Independent Evaluation

4.5.1 No Resampling

4.5.1.1 Task: Writing Email/Report VS Absent

The model that managed to better separate Writing Email/Report versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix

(b) ROC Curve

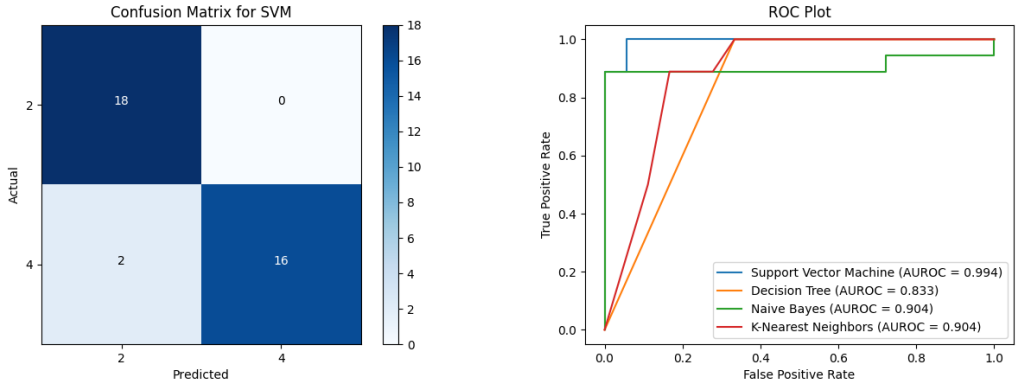
Figure 4.16: Confusion Matrix and ROC Curve for Writing Email/Report VS Absent

4.5.1.2 Task: Browsing/Scrolling Social Media VS Absent

The model that managed to better separate Browsing/Scrolling Social Media versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

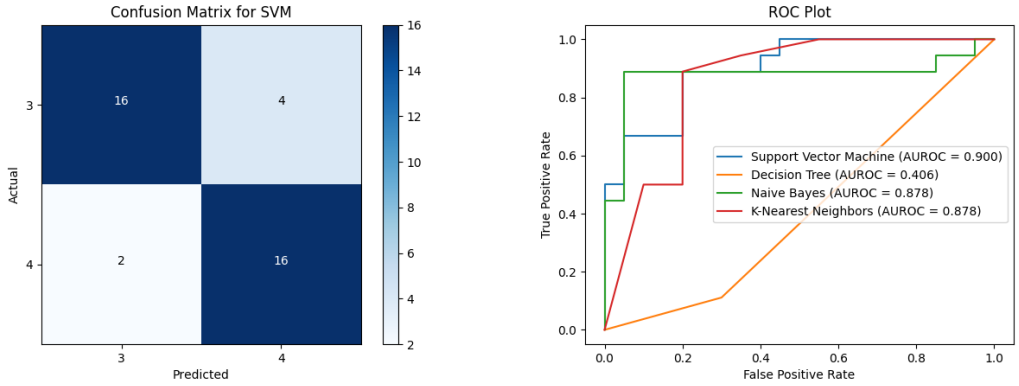
4.5.1.3 Task: Communicating VS Absent

The model that managed to better separate Communicating versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.17: Confusion Matrix and ROC Curve for Browsing/Scrolling Social Media VS Absent



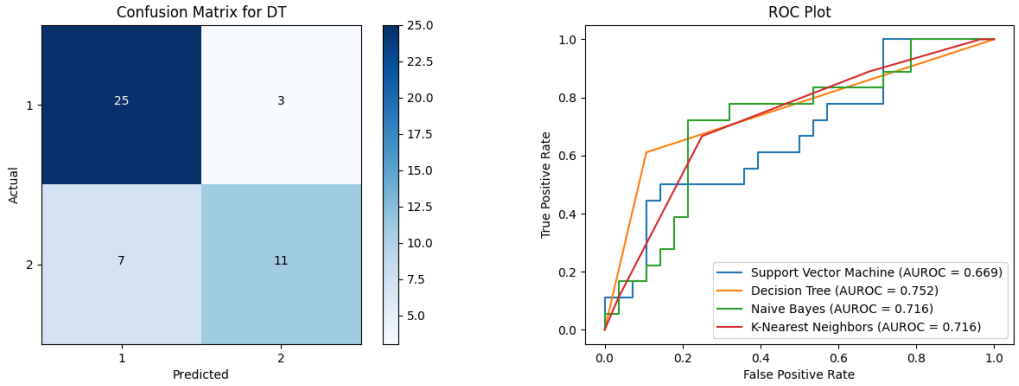
(a) Confusion Matrix (b) ROC Curve

Figure 4.18: Confusion Matrix and ROC Curve for Communicating VS Absent

4.5.2 Random Undersampling

4.5.2.1 Task: Writing Email/Report VS Browsing/Scrolling Social Media

The model that managed to better separate Writing Email/Report versus Scrolling Social Media was the Decision Tree. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

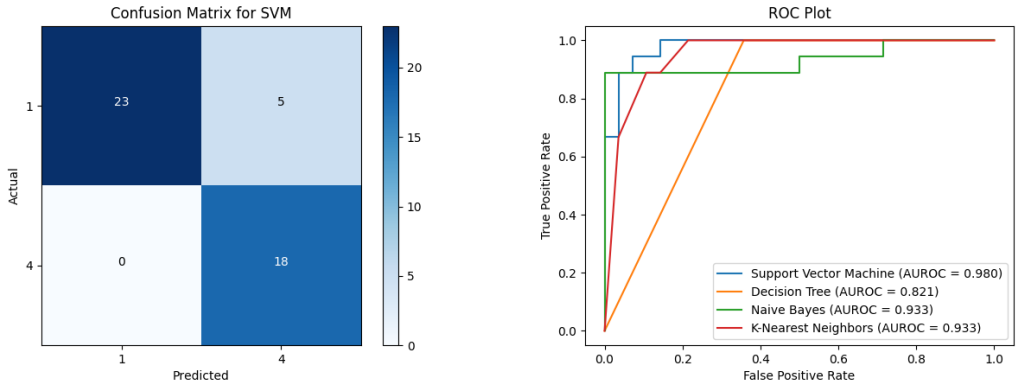


(a) Confusion Matrix (b) ROC Curve

Figure 4.19: Confusion Matrix and ROC Curve for Writing Email/Report VS Browsing/Scrolling Social Media

4.5.2.2 Task: Writing Email/Report VS Absent

The model that managed to better separate Writing Email/Report versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

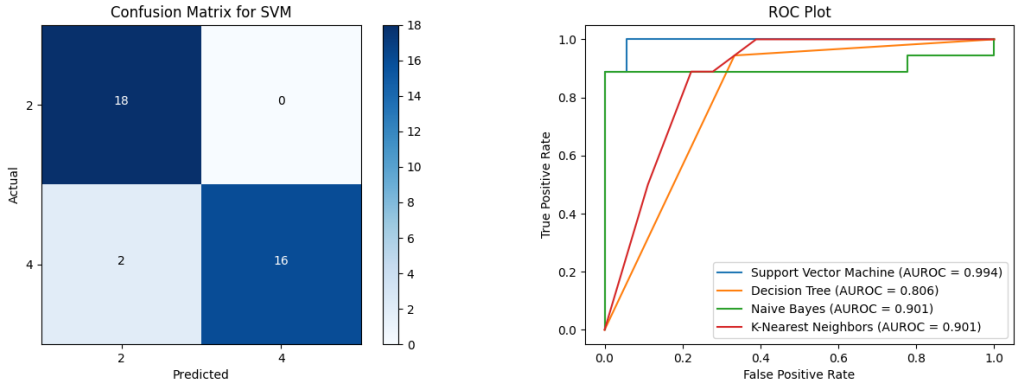


(a) Confusion Matrix (b) ROC Curve

Figure 4.20: Confusion Matrix and ROC Curve for Writing Email/Report VS Absent

4.5.2.3 Task: Browsing/Scrolling Social Media VS Absent

The model that managed to better separate Browsing/Scrolling Social Media versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.

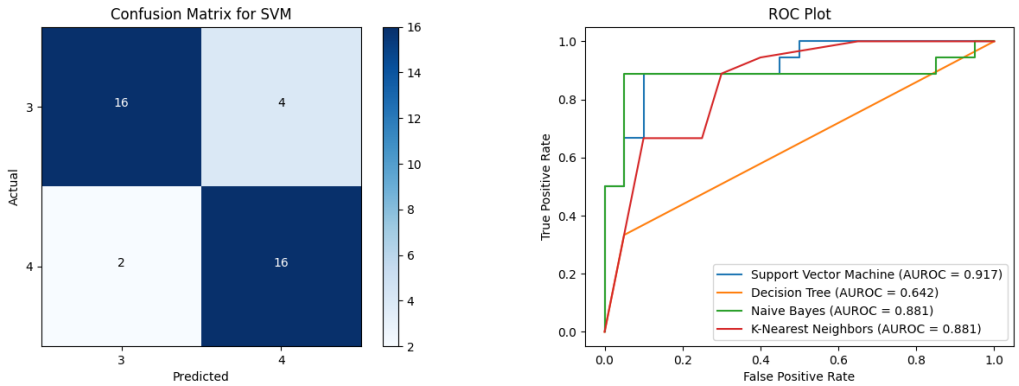


(a) Confusion Matrix (b) ROC Curve

Figure 4.21: Confusion Matrix and ROC Curve for Browsing/Scrolling Social Media VS Absent

4.5.2.4 Task: Communicating VS Absent

The model that managed to better separate Communicating versus Absent was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

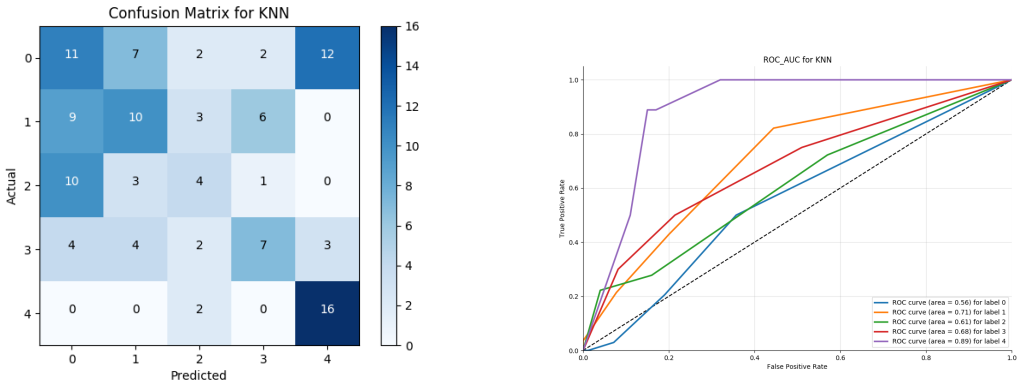
Figure 4.22: Confusion Matrix and ROC Curve for Communicating VS Absent

4.6 5-Class Classification: Recording Independent Evaluation

4.6.1 No Resampling

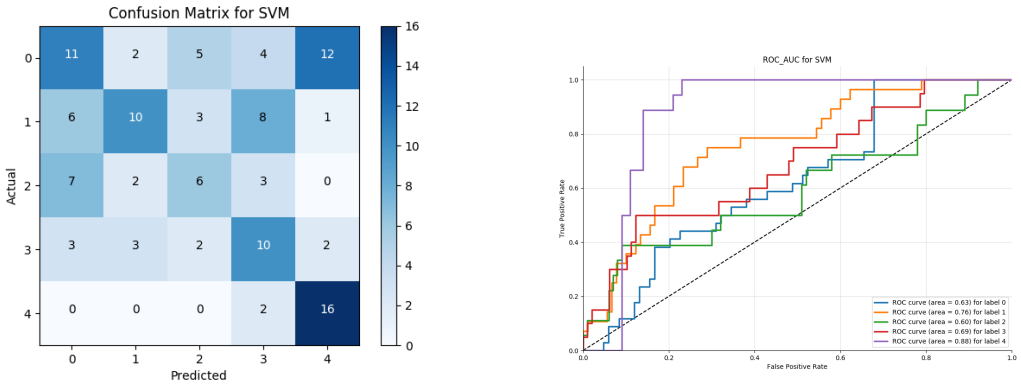
4.6.1.1 Task: All class classification

The model that managed to better separate all the classes was the Support Vector Machine and the K-Nearest Neighbor. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.23: Confusion Matrix and ROC Curve for 5-Class Classification



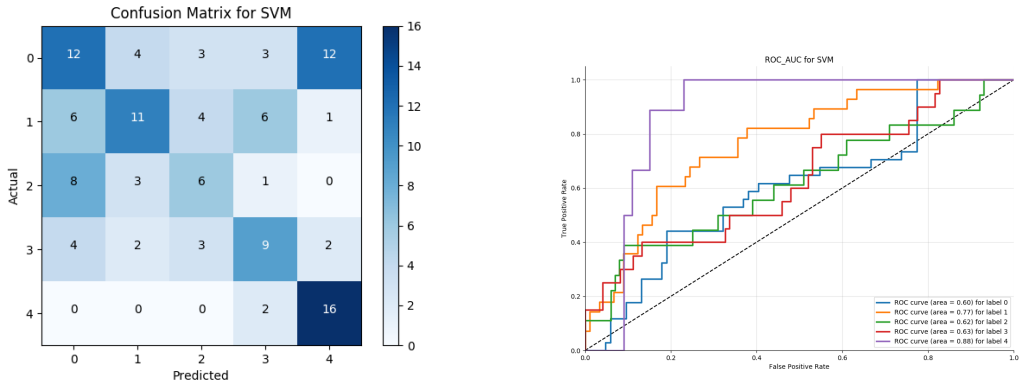
(a) Confusion Matrix (b) ROC Curve

Figure 4.24: Confusion Matrix and ROC Curve for 5-Class Classification

4.6.2 Random Undersampling

4.6.2.1 Task: All class classification

The model that managed to better separate all the classes was the Support Vector Machine. Next we quote the confusion matrix as well as the roc curves for this specific experiment.



(a) Confusion Matrix (b) ROC Curve

Figure 4.25: Confusion Matrix and ROC Curve for 5-Class Classification

At this point we quote the tables with the exact harmonic means for all the aforementioned experiments that are independent on the recordings. First we list the tables for the problems of binary classification based on balanced and unbalanced data sets and then we continue listing the tables for the corresponding classification problem of all classes.

Binary Classification Recording-Independent Evaluation			
Experiment	**Class1 F1**	**Class2 F1**	**Classifier**
Writing Email/Report vs Absent	90%	88%	SVM
Browsing/Scrolling Social Media vs Absent	95%	94%	SVM
Communicating vs Absent	84%	84%	SVM

Table 4.5: F1-measure scores for binary classification on imbalanced data set

Binary Classification Recording-Independent Evaluation			
Experiment	**Class1 F1**	**Class2 F1**	**Classifier**
Writing Email/Report vs Browsing/Scrolling Social Media	83%	69%	DT
Writing Email/Report vs Absent	90%	88%	SVM
Browsing/Scrolling Social Media vs Absent	95%	94%	SVM
Communicating vs Absent	84%	84%	SVM

Table 4.6: F1-measure scores for binary classification on balanced data set using imblearn

5 Class Classification Recording-Independent Evaluation						
Experiment	**Class1 F1**	**Class2 F1**	**Class3 F1**	**Class4 F1**	**Class5 F1**	**Classifier**
All Classes	36%	44%	35%	43%	65%	SVM

Table 4.7: F1-measure scores for 5-class classification on imbalanced data set

5 Class Classification Recording-Independent Evaluation						
Experiment	**Class1 F1**	**Class2 F1**	**Class3 F1**	**Class4 F1**	**Class5 F1**	**Classifier**
All Classes	38%	46%	35%	44%	65%	SVM

Table 4.8: F1-measure scores for 5-class classification on balanced data set

Chapter 5

Conclusions and Future Work

5.1 Conclusion and Future Work

In this work the problem of recognizing activities performed by employees during a working day was studied. The mechanism of stress was first introduced, as well as the most common ways to detect it. The purpose of this work was to implement a technical measurement of a user's activities in a work environment, which requires the use of a webcam, mouse, keyboard, FSR sensors on the surface of a chair, as well as a simple personal computer. The technique we implemented is based on features exported from the aforementioned computer peripherals (mouse, keyboard), as well as the FSR sensors. As far as we know, this is the first attempt to combine these three ways. For each of them a time series is constructed, which depicts the y-component in terms of time. Then we divide these time series into segments of 10 seconds and then synchronize them in order to construct the set of data that will be used to apply techniques from the field of machine learning. Special emphasis was given to the study of the research area of recognizing the work activities of users from multiple sources of information. We were able to create a small proof of concept using some limited data and create a classifier with the aim of distinguishing between five user activities. The method implemented led to the conclusion of useful conclusions for possible research extensions. On average the proposed feature extraction and classification approach, achieves an average **F1** of:

1. 55.6% for the case of the recording dependent 5-class classification experiment on imbalanced data
2. 49.2% for the case of the recording dependent 5-class classification experiment on balanced data
3. 44.6% for the case of the recording independent 5-class classification experiment on imbalanced data
4. 45.6% for the case of the recording independent 5-class classification experiment on balanced data

We believe that there are significant margins improvement in the extraction of features, where it is possible to explore various other approaches such as e.g. the configuration of algorithms as well as the introduction of techniques from the field of deep learning, which have been used in the literature with great success. Due to the nature of the task, most of our work revolved around building up the data collection process. Despite our best efforts, we had to deal with unexpected issues that made the evaluation of our data set impossible. Moreover, it was hard to gather a large enough data set in order to deal with the class imbalance problem or use some state-of-the-art tools and techniques. The two problems mentioned above are an important opportunity to improve this work. Another point that could improve the performance of our architecture, would be to make advantage of the web-camera that records both video and audio streams. This could prove to be very important as new features could be extracted from the image and sound, in order to enrich the feature vector being exported.

References

[1] Tal Yarkoni. Psychoinformatics: New Horizons at the Interface of the Psychological and Computing Sciences. *Current Directions in Psychological Science*, 21(6):391–397, December 2012. Publisher: SAGE Publications Inc.

[2] Luciano Floridi. *The Ethics of Information*. OUP Oxford, October 2013. Google-Books-ID: _XHcAAAAQBAJ.

[3] Linsey M. Barker and Maury A. Nussbaum. Fatigue, performance and the work environment: a survey of registered nurses. *Journal of Advanced Nursing*, 67(6):1370–1382, 2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2648.2010.05597.x.

[4] Christian Montag, Éilish Duke, and Alexander Markowetz. Toward Psychoinformatics: Computer Science Meets Psychology. *Computational and Mathematical Methods in Medicine*, 2016:e2983685, June 2016. Publisher: Hindawi.

[5] Alexander Markowetz, Konrad Błaszkiewicz, Christian Montag, Christina Switala, and Thomas E. Schlaepfer. Psycho-Informatics: Big Data shaping modern psychometrics. *Medical Hypotheses*, 82(4):405–411, April 2014.

[6] Deborah Lupton. Self-tracking cultures: towards a sociology of personal informatics. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*, OzCHI '14, pages 77–86, New York, NY, USA, December 2014. Association for Computing Machinery.

[7] Dvijesh Shastri, Manos Papadakis, Panagiotis Tsiamyrtzis, Barbara Bass, and

Ioannis Pavlidis. Perinasal Imaging of Physiological Stress and Its Affective Potential. *IEEE Transactions on Affective Computing*, 3(3):366–378, July 2012. Conference Name: IEEE Transactions on Affective Computing.

[8] Takuto Hayashi, Yuko Mizuno-Matsumoto, Eika Okamoto, Makoto Kato, and Tsutomu Murata. An fMRI study of brain processing related to stress states. In *World Automation Congress 2012*, pages 1–6, June 2012. ISSN: 2154-4824.

[9] Rafal Kocielnik, Natalia Sidorova, Fabrizio Maria Maggi, Martin Ouwerkerk, and Joyce H. D. M. Westerink. Smart technologies for long-term stress monitoring at work. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 53–58, June 2013. ISSN: 1063-7125.

[10] Kais Riani, Michalis Papakostas, Hussein Kokash, Mohamed Abouelenien, Mihai Burzo, and Rada Mihalcea. Towards detecting levels of alertness in drivers using multiple modalities. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '20, pages 1–9, New York, NY, USA, June 2020. Association for Computing Machinery.

[11] Burcu Cinaz, Bert Arnrich, Roberto La Marca, and Gerhard Tröster. Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing*, 17(2):229–239, February 2013.

[12] Chang Zhi Wei. Stress Emotion Recognition Based on RSP and EMG Signals. *Advanced Materials Research*, 709:827–831, 2013. Conference Name: Advances in Applied Science, Engineering and Technology ISBN: 9783037857182 Publisher: Trans Tech Publications Ltd.

[13] Jonathan Aigrain, Michel Spodenkiewicz, Séverine Dubuisson, Marcin Detyniecki, David Cohen, and Mohamed Chetouani. Multimodal Stress Detection from Multiple Assessments. *IEEE Transactions on Affective Computing*, 9(4):491–506, October 2018. Conference Name: IEEE Transactions on Affective Computing.

[14] Armando Barreto, Jing Zhai, Naphtali Rishe, and Ying Gao. Significance of Pupil Diameter Measurements for the Assessment of Affective State in Computer Users. In Khaled Elleithy, editor, *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, pages 59–64, Dordrecht, 2007. Springer Netherlands.

[15] Davide Carneiro, José Carlos Castillo, Paulo Novais, Antonio Fernández-Caballero, and José Neves. Multimodal behavioral analysis for non-invasive stress detection. *Expert Systems with Applications*, 39(18):13376–13389, December 2012.

[16] Clayton Epp, Michael Lippold, and Regan L. Mandryk. Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 715–724, New York, NY, USA, May 2011. Association for Computing Machinery.

[17] Belinda H. W. Eijckelhof, Maaike A. Huysmans, Birgitte M. Blatter, Priscilla C. Leider, Peter W. Johnson, Jaap H. van Dieën, Jack T. Dennerlein, and Allard J. van der Beek. Office workers' computer use patterns are associated with workplace stressors. *Applied Ergonomics*, 45(6):1660–1667, November 2014.

[18] Klemen Peternel, Matevž Pogačnik, Rudi Tavčar, and Andrej Kos. A Presence-Based Context-Aware Chronic Stress Recognition System. *Sensors*, 12(11):15888–15906, November 2012. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[19] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. AffectAura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 849–858, New York, NY, USA, May 2012. Association for Computing Machinery.

[20] Paul Ekman. Emotions revealed. *BMJ*, 328(Suppl S5):0405184, May 2004. Publisher: British Medical Journal Publishing Group Section: Student.

[21] Vijay P. Patil, Krishna Kant Nayak, and Manish Saxena. *Voice Stress Detection*. 2013.

[22] Andres L. Bleda, Francisco J. Fernández-Luque, Antonio Rosa, Juan Zapata, and Rafael Maestre. Smart Sensory Furniture Based on WSN for Ambient Assisted Living. *IEEE Sensors Journal*, 17(17):5626–5636, September 2017. Conference Name: IEEE Sensors Journal.

[23] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep Learning for Sensor-based Activity Recognition: A Survey. *Pattern Recognition Letters*, 119:3–11, March 2019. arXiv: 1707.03502.

[24] Jingyuan Cheng, Bo Zhou, Mathias Sundholm, and P. Lukowicz. Smart Chair: What Can Simple Pressure Sensors under the Chairs' Legs Tell Us about User Activity? *undefined*, 2013.

[25] BeautyTech.jp. Coming soon: A revolutionary smart mirror by Panasonic, March 2019.

[26] Bert Arnrich, Cornelia Setz, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. What Does Your Chair Know About Your Stress Level? *IEEE Transactions on Information Technology in Biomedicine*, 14(2):207–214, March 2010. Conference Name: IEEE Transactions on Information Technology in Biomedicine.

[27] Robert M. Sapolsky. *Why Zebras Don't Get Ulcers: The Acclaimed Guide to Stress, Stress-Related Diseases, and Coping (Third Edition)*. Henry Holt and Company, September 2004. Google-Books-ID: EI88oS_3fZEC.

[28] Walter B. Cannon. Organization for physiological homeostasis. *Physiological Reviews*, 9(3):399–431, July 1929. Publisher: American Physiological Society.

[29] George P. Chrousos and Philip W. Gold. The Concepts of Stress and Stress System Disorders: Overview of Physical and Behavioral Homeostasis. *JAMA*, 267(9):1244–1252, March 1992.

[30] Jaime G. Carbonell, Ryszard S. Michalski, and Tom M. Mitchell. 1 - AN OVERVIEW OF MACHINE LEARNING. In Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell, editors, *Machine Learning*, pages 3–23. Morgan Kaufmann, San Francisco (CA), January 1983.

[31] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995.

[32] Yan-yan SONG and Ying LU. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2):130–135, April 2015.

[33] I. Rish. An empirical study of the naive Bayes classifier. *undefined*, 2001.

[34] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. KNN Model-Based Approach in Classification. In Robert Meersman, Zahir Tari, and Douglas C. Schmidt, editors, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Lecture Notes in Computer Science, pages 986–996, Berlin, Heidelberg, 2003. Springer.

[35] Jorma Laurikkala. Improving Identification of Difficult Small Classes by Balancing Class Distribution. In Silvana Quaglini, Pedro Barahona, and Steen Andreassen, editors, *Artificial Intelligence in Medicine*, Lecture Notes in Computer Science, pages 63–66, Berlin, Heidelberg, 2001. Springer.

[36] Andrew Estabrooks, Taeho Jo, and Nathalie Japkowicz. A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Computational Intelligence*, 20(1):18–36, 2004. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0824-7935.2004.t01-1-00228.x.

[37] Michalis Vrigkas, Christophoros Nikou, and Ioannis A. Kakadiaris. A Review of Human Activity Recognition Methods. *Frontiers in Robotics and AI*, 0, 2015. Publisher: Frontiers.

[38] Oscar D. Lara and Miguel A. Labrador. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys Tutorials*,

15(3):1192–1209, 2013. Conference Name: IEEE Communications Surveys Tutorials.

[39] J. F. Zhou, Y. H. Song, Q. Zheng, Q. Wu, and M. Q. Zhang. Percolation transition and hydrostatic piezoresistance for carbon black filled poly(methylvinylsilioxane) vulcanizates. *Carbon*, 46(4):679–691, April 2008.

[40] General Data Protection Regulation (GDPR) – Official Legal Text.

[41] Cees G. M. Snoek, Marcel Worring, and Arnold W. M. Smeulders. Early versus late fusion in semantic video analysis. In *Proceedings of the 13th annual ACM international conference on Multimedia*, MULTIMEDIA '05, pages 399–402, New York, NY, USA, November 2005. Association for Computing Machinery.