Logging Stress and Anxiety Using a Gamified Mobile-based EMA Application, and Emotion Recognition Using a Personalized Machine Learning Approach

Smart Systems
Master's Degree Programme in Information and Communication Technology
Department of Computing, Faculty of Technology
Master of Science in Technology Thesis

Author:
Ali Kargarandehkordi

Supervisors:
Assistant Professor, Peter Yiğitcan Washington (University of Hawaiʻi at Mānoa, USA)
Assistant Professor, Matti Kaisti (University of Turku, Finland)

June 2023

Master of Science in Technology Thesis
Department of Computing, Faculty of Technology
University of Turku

Subject: Smart Systems
Programme: Master's Degree Programme in Information and Communication Technology
Author: Ali Kargarandehkordi
Title: Logging Stress and Anxiety Using a Gamified Mobile-based EMA Application, and Emotion Recognition Using a Personalized Machine Learning Approach
Number of pages: 67 pages, 6 appendix pages
Date: June 2023

Abstract.

According to American Psychological Association (APA) more than 9 in 10 (94 percent) adults believe that stress can contribute to the development of major health problems, such as heart disease, depression, and obesity. Due to the subjective nature of stress, and anxiety, it has been demanding to measure these psychological issues accurately by only relying on objective means. In recent years, researchers have increasingly utilized computer vision techniques and machine learning algorithms to develop scalable and accessible solutions for remote mental health monitoring via web and mobile applications. To further enhance accuracy in the field of digital health and precision diagnostics, there is a need for personalized machine-learning approaches that focus on recognizing mental states based on individual characteristics, rather than relying solely on general-purpose solutions.

This thesis focuses on conducting experiments aimed at recognizing and assessing levels of stress and anxiety in participants. In the initial phase of the study, a mobile application with broad applicability (compatible with both Android and iPhone platforms) is introduced (we called it STAND). This application serves the purpose of Ecological Momentary Assessment (EMA). Participants receive daily notifications through this smartphone-based app, which redirects them to a screen consisting of three components. These components include a question that prompts participants to indicate their current levels of stress and anxiety, a rating scale ranging from 1 to 10 for quantifying their response, and the ability to capture a selfie. The responses to the stress and anxiety questions, along with the corresponding selfie photographs, are then analyzed on an individual basis. This analysis focuses on exploring the relationships between self-reported stress and anxiety levels and potential facial expressions indicative of stress and anxiety, eye features such as pupil size variation and eye closure, and specific action units (AUs) observed in the frames over time. In addition to its primary functions, the mobile app also gathers sensor data, including accelerometer and gyroscope readings, on a daily basis. This data holds potential for further analysis related to stress and anxiety. Furthermore, apart from capturing selfie photographs, participants have the option to upload video recordings of themselves while engaging in two neuropsychological games. These recorded videos are then subjected to analysis in order to extract pertinent features that can be utilized for binary classification of stress and anxiety (i.e., stress and anxiety recognition). The participants that will be selected for this phase are students aged between 18 and 38, who have received recent clinical diagnoses indicating specific stress and anxiety levels. In order to enhance user engagement in the intervention, gamified elements - an emerging trend to influence user behavior and lifestyle - has been utilized. Incorporating gamified elements into non-game contexts (e.g., health-related) has gained overwhelming popularity during the last few years which has made the interventions more delightful, engaging, and motivating.

In the subsequent phase of this research, we conducted an AI experiment employing a personalized machine learning approach to perform emotion recognition on an established dataset called Emognition. This experiment served as a simulation of the future analysis that will be conducted as part of a more comprehensive study focusing on stress and anxiety recognition. The outcomes of the emotion recognition experiment in this study highlight the effectiveness of personalized machine learning techniques and bear significance for the development of future diagnostic endeavors. For training purposes, we selected three models, namely KNN, Random Forest, and MLP. The preliminary performance accuracy results for the experiment were 93%, 95%, and 87% respectively for these models.

Keywords: Mobile application, anxiety, stress, personalized machine-learning, gamification, emotion recognition, model.

Table of Contents

1 Introduction

1.1 Research Background and Significance

The field of remote mental health monitoring has been transformed by recent advancements in computer vision and machine learning. These advancements have paved the way for scalable and easily accessible solutions through web and mobile applications. The contemporary mobile application market is saturated with a diverse array of offerings. Simultaneously, the rising awareness of mental health issues within societies has led to the proliferation of technology-based mobile health (mHealth) applications, which have become integral to individuals' daily lives and gained significant popularity [1, 2].

Continual exposure to different stressors and recurring experiences of anxiety can significantly impair one's quality of life and overall life expectancy [3]. Many diseases, including cardiovascular and immune-related conditions, are known to be connected to these mental states, as understood presently [4, 5]. Although there have been extensive analytic and treatment efforts done, still the level of such disorders within a large scale of the population worldwide remains greatly unknown or undertreated [6, 7]. Within the last few years, as a potential solution to detect anxiety, advanced wearable devices offer cognitive state measurements. Although this technology is under consistent development, still wearable devices are facing significant challenges to measure anxiety in a stable way [8].

The widespread use of mobile phones has brought about numerous advantages, as these portable devices have become an essential aspect of daily life in contemporary society [9, 10]. It helped people to overcome commonly known issues regarding health services including availability, affordability, and time constraints [11, 12]. Implementing mHealth into mobile applications would enable individuals to access mental health services who would otherwise not seek them in clinical environments [12, 13]; because most often in addition to the mentioned barriers above, attitudinal factors, on the other side can play a more important role in life management [11].

Stress and, particularly, anxiety are among subjective matters that require methodological and practical experts to be analyzed. Therefore, measuring such states through objective means can be quite demanding, and most of the time, not accurate. Typically, the process of collecting the necessary data to detect such states involves administering questionnaires to users, which can be perceived as tedious and monotonous. This method, designed by psychological experts, often leads to a negative impact on the final result due to its potential to induce boredom in users [14]. Expanding usage of AI and machine learning techniques in recent research, throw light on complexities and mechanisms behind recognizing these subjective experiences. However, achieving an accurate ML model to detect stress and anxiety, could be challenging on its own, and craves hard work and creative ideas behind it. Accurate data collection and model selection are key factors in improving the efficiency of machine learning techniques for recognizing stress and anxiety. Some studies have suggested promising methods to increase this accuracy, including 1) The utilization of a personalized ML approach. This approach involves tailoring an individual ML model to each participant, thereby improving the diagnostic procedure [15, 16], and 2) Engaging respondents completely through an emerging trend in the field of human-computer interaction: the use of gamified web and mobile applications. Utilizing these platforms and incorporating gamified elements can have a positive influence on the psychological and behavioral attributes of the target users. This, in turn, can lead to a notable enhancement in user engagement and data efficiency [17-19].

Numerous studies provide evidence that anxiety exerts a direct influence on facial expressions, which can be discerned even during routine activities. Anxiety primarily affects facial expressions, leading to observable changes in the eyes, including fluctuations in pupil size, eyelid closure, and blinking frequency. It also involves alterations in mouth, such as specific changes in the lips and cheeks, as well as variations in head movements and speed [20-22].

The second chapter of this thesis introduces a comprehensive overview of the techniques employed in building a cross-platform mobile application for the purpose of Ecological Momentary Assessment (EMA), which serves as the primary focus of the first phase of the thesis. The chapter delves into the development process, outlining the various application components and discussing the future recruitment procedure. Moving on to the third chapter, it elucidates the methodology for conducting emotion recognition tasks targeting six discrete emotions. A similar approach will be adopted for the stress and anxiety recognition task once an ample amount of data has been collected. The fourth chapter concludes the research work, while the fifth and final chapter explores potential future technologies for further investigations.

1.2 Literature Review

In this section, we present a comprehensive literature review that serves as the foundation for the integration of techniques employed in constructing our study and developing the STAND mobile application. By drawing inspiration from a range of relevant sources and their features on recognizing various mental states, we aim to build upon existing knowledge and incorporate the most effective strategies into our research framework.

1.2.1 Ecological Momentary Assessment (EMA)

Ecological momentary assessment (EMA) has emerged as a promising approach for capturing momentary moods in natural and real-life settings. This method, also known as the Experience Sample Method or Ambulatory Assessment, involves repeated observations to gather data on mood fluctuations over time. It provides a valuable tool for studying mood dynamics in everyday environments [23-25]. EMA allows individuals to provide real-time reports of their current mental state in a familiar and immediate manner. In most of the conventional methods, with the aid of standardized instruments, clinicians examine the reports collected retrospectively in unfamiliar settings [25-27]. By EMA, the principle is to maintain regular life patterns while attending an intervention. Patients, self-report their mood or anxiety on a regular basis (e.g., multiple times a day) while performing their ordinary daily tasks over a specific period of time. Depending on study objectives, such reports can be collected in the short/long term through a variety of instruments such as smartphones, wearable devices, and other digital assistants [28]. Utilizing EMA in digital healthcare interventions brings a range of methodological strengths including generalization enhancement with more ecological (real-world environments) validity, lower retrospective bias and errors due to the "momentary" aspect of the method, and the capability to gather the data at regular intervals and in diverse real-world scenarios [25].

1.2.2 Conventional Approaches and Techniques

Analyzing the data collected from the patients through clinician interviews or self-reports (scale-based responses to a standardized list of questions) are the two most common ways to assess anxiety disorders [29]. In these protocols, individuals are requested to provide self-reports on the intensity and severity of their physical symptoms and emotions associated with anxiety in their daily lives. While the effectiveness of these approaches is acknowledged, there are concerns regarding the subjective nature of the gathered responses, which can potentially reduce the accuracy of the data and introduce biases in the assessment of stress [30]. Reluctancy or inability to answer the questions accurately because of recollection inaccuracies, ambiguity in the questions [31], psychopathology and biases [32, 33], or discomfort during interviews with clinicians, are commonly encountered issues that could potentially undermine the effectiveness of the data.

Evaluating the levels of cortisol produced by the HPA axis is another classic physiological method to measure stress/anxiety. This procedure involves a demanding process of analyzing cortisol extracted from different sources of the body e.g., blood, urine, hair, and saliva [34, 35]. There are other indicators for physiological measures of stress that are based on assessments of activities of the brain (EEG features associated with arousal, such as asymmetry [36, 37], and the ratio of beta to alpha waves [38]), heart (ECG), blood (PPG), muscle (EMG), and responses from the skin (galvanic skin/electrodermal response), and respiratory system [14, 39, 40]. Skin conductance will be increased during stressful moments as anxiety regulates active sweat glands [41]. There is also a correlation between breathing patterns and emotional status useful for anxiety detection, in which there are several reports that the respiration rate increases considerably while someone is under stress [42, 43]. Finally, speech characteristics are affected by anxiety [44]. Investigating the changes in fundamental frequencies of the voice are among the most attractive topics for researchers working in this field [20, 45].

1.2.3 Facial Signs and Expressions

In the late 1800s, Charles Darwin made a significant discovery in his book "The Expression of the Emotions in Man and Animals," highlighting the connection between emotional states and facial expressions [46]. In subsequent research conducted during the 1900s and early 2000s, as more sophisticated techniques like evaluating the HPA axis and cardiovascular responses to stress emerged, additional evidence was presented to support this association. In recent years, there have also been several studies reporting that facial signs can be considered as a potential way to provide insights when it comes to the classification of anxiety [20, 21]. Anxiety primarily affects facial expressions through various changes. These alterations encompass the eyes, including variations in pupil size, eye aperture, blinking rate, and gaze distribution. Additionally, there are observable effects on mouth activity [47], alterations in the shape and movement of the lips and cheeks [48], as well as the movements and velocity of the head [49, 50]. More specifically, considering the eye-related features, for instance, blinking rate, there are other factors that might affect the expression such as lying[51], depression, Parkinson [52], schizophrenia [53], and certain external factors like lighting, humidity, and temperature [54]. There is a notable association between the closure of eyelids caused by a light stimulus and the individual's anxiety level [55].

One of the other promising associations between eye-related features and anxiety is pupil diameter. This parameter has been considerably documented in various studies as a potential index indicating the levels of anxiety [56, 57] and has proved to be absolutely effective in the confidentiality rate and higher accuracy of the stress detection procedure [58]. This study assessment was first introduced in 2002 by Partala and Surakka in which a very distinct correlation was discovered between the pattern of eye dilation and emotional states [59].

1.2.4 Data Mining and Machine Learning

In recent times, there has been a growing interest in exploring the applications of Data Mining and Machine Learning (ML) in various healthcare domains, including therapy planning/management, disease diagnosis, and illness prediction. These technologies have demonstrated substantial effectiveness in enhancing the efficiency and quality of medical care, making them highly promising tools for healthcare professionals [60, 61]. In the field of mental diseases such as predicting psychological and perceived anxiety, there has been a successful utilization of machine learning models, however, challenges are always raised when it comes to reliability. Limitations on real healthcare scenarios (available for research and development) data collection [62], a limited availability of research-only resources [63], concerns regarding data confidentiality and restrictions on data disclosure [64] as well as the absence of suitable commercial licenses [65] are the main factors that pose significant hurdles in the widespread adoption of machine learning approaches in various domains.

Several popular machine learning algorithms, such as Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Random Forest Tree, and Naive Bayes (NB), have been employed in the prediction of anxiety-related risk factors. For instance, Wang et al. conducted a study utilizing the XG-Boost machine learning model to assess the severity levels and prevalence of anxiety among undergraduate students during the COVID-19 pandemic [66]. Richter et al applied ML models to the data collected from adult patients in order to differentiate the symptoms of anxiety [67]. Similar research was done using a number of 5 machine learning algorithms by Priya et al with the aim of predicting stress and anxiety between employed and unemployed individuals [68]. There were reported average accuracy rates of 71% and 73% for random forest trees and naive Bayes algorithms for anxiety and depression symptoms, in two other studies [68, 69]. Furthermore, performance accuracies of 71.4% (anxiety), and 72.3% (stress) were achieved in [68], employing decision tree, SVM, naive Bayes, random forest, and KNN (k-nearest neighbor) algorithms on a sample of adults aged 20-60 years, it was found that the random forest (RF) algorithm achieved the highest accuracy. This study demonstrates the effectiveness of machine learning techniques in predicting mental health risk factors, with the random forest classifier showcasing superior performance accuracy [68]. ML models possess significant potential for diverse applications in healthcare, including pain assessment [70] and helping individuals with autism [71].

The performance of general ML models for detecting emotional states tends to be relatively poor [72]. One of the main reasons behind this issue is perhaps due to the existence of high physical and characteristics variability among different individuals. Cutting-edge research indicates that personalized machine learning is a more effective approach, demonstrating its suitability for advancing precision health machine learning. This is particularly relevant when working with high-dimensional data and making predictions that are subjective and specific to individuals [72, 73]. This approach highlights the need for conducting additional research on interpersonal mental health variations. It involves training and testing each model using individual-specific data to investigate the impact of this variability on accurately detecting emotional transitions and states.

1.2.5 Smart-sensing and Passive Monitoring

Assessing mental disorders through the observation and assessment of patients' behavior as an indicator of changes in symptom severity can be seen as an alternative approach. Incorporating mobile phones as sensing devices in such initiatives allows researchers to passively capture specific behaviors associated with different mental disorders. These portable devices are particularly advantageous for two main reasons: 1) They possess a variety of powerful sensors and software that can evaluate behaviors and contextual information, and 2) Individuals typically carry them throughout the day, ensuring continuous monitoring and data collection [74-76]. Apart from relying on subjective information obtained through interviews or self-reports, mobile phones can serve as a means to gather objective data from the same individuals. This approach enables researchers to employ an objective evaluation of mental health by utilizing the data collected during the data collection process [77].

1.2.6 Reminders

Including alert-based elements (reminders) in digital health has been proposed in previous studies as a means to improve levels of engagement and promote adherence to interventions. These reminders can be delivered in various formats (such as short motivational messages or quotes) through emails, SMS, or push notifications integrated into mobile applications. Integrating this cost-effective functionality into mental healthcare research has demonstrated its significant benefit and crucial role in user engagement compared to interventions lacking this feature [78-80]. Moreover, it has been shown that incorporating reminders into digital health interventions can lead to increased adherence to the research and reduced rates of non-usage attrition. These are well-recognized challenges associated with online-based interventions [81-83]. Nonusage attrition refers to the phenomenon where participants discontinue the use of a digital intervention while still actively participating in the research through other protocols, such as completing questionnaires or assessments.

1.2.7 Gamification

In recent years, the importance of human-computer interaction (HCI) has grown significantly. Gamification, as a prominent trend in enhancing this interaction, has gained considerable attention in the field of machine learning. Gamification is commonly defined as the incorporation of game design elements into non-game contexts, with the primary objective of increasing user engagement and improving their overall experience in HCI. In an academic setting, the utilization of gamified elements and incentives, along with visually appealing designs, to augment user engagement, motivation, and compliance with research activities is referred to as gamification [84-86]. In recent years, various elements of this emerging trend have been integrated into diverse healthcare services to serve as supporting factors in boosting user activity and enhancing behavioral productivity. Different elements [17, 87] are discussed and analyzed so far including points, leaderboards, challenges, feedback [88-93], rewards, and story/theme [94, 95], achievements/badges [88], progress tracking, and levels [91, 93]. In 2021, Dykens et al introduced a framework called UGM (Unified Gamification and Motivation) for understanding the efficiency of including gamification elements to increase engagement in therapeutic interventions [18]. Through investigating the relationship between gamification and motivational psychology, this model tries to prove the efficiency of using game-like elements in order to make the intervention more noticeable, which also results in greater user engagement and motivation.

1.3 Innovation Points

This study aims to contribute mentioned novel insights and innovative approaches to the field of mental health assessment. By employing a personalized machine learning framework, it seeks to overcome the limitations of conventional methods and offer a more tailored and precise approach to recognizing and assessing these mental states. The integration of smartphone-based Ecological Momentary Assessment (EMA) with AI experiments and image analysis provides a multifaceted and comprehensive approach to capturing and analyzing data. The inclusion of sensor data, selfie photographs, and video recordings adds depth and richness to the analysis, enabling a more accurate recognition of stress and anxiety and holistic understanding of their levels. This research also explores the potential relationships between self-reported levels of stress and anxiety and facial expressions, eye features, and other physiological indicators captured in the collected data. The findings of this study have the potential to advance the field of mental health diagnostics and pave the way for personalized interventions that are tailored to individuals' unique experiences of stress and anxiety.

2 First Part: STAND Mobile Application Development

2.1 Overview

The centerpiece of this research is the STAND EMA app, which serves as a powerful tool for collecting multidimensional data from participants. By employing a personalized machine learning approach, we will analyze the expecting rich dataset collected through this app and extract the most influential features. Our objective is to explore the efficacy of various methodologies discussed earlier in achieving a heightened level of accuracy in recognizing emotional states. Additionally, we aim to investigate the potential correlation between an objective indicator of anxiety using computer vision techniques. The outcomes of this study will serve as a valuable reference for future personalized interventions. To accomplish these goals, the research comprises three distinct experimental phases, as depicted in Figure 1.:

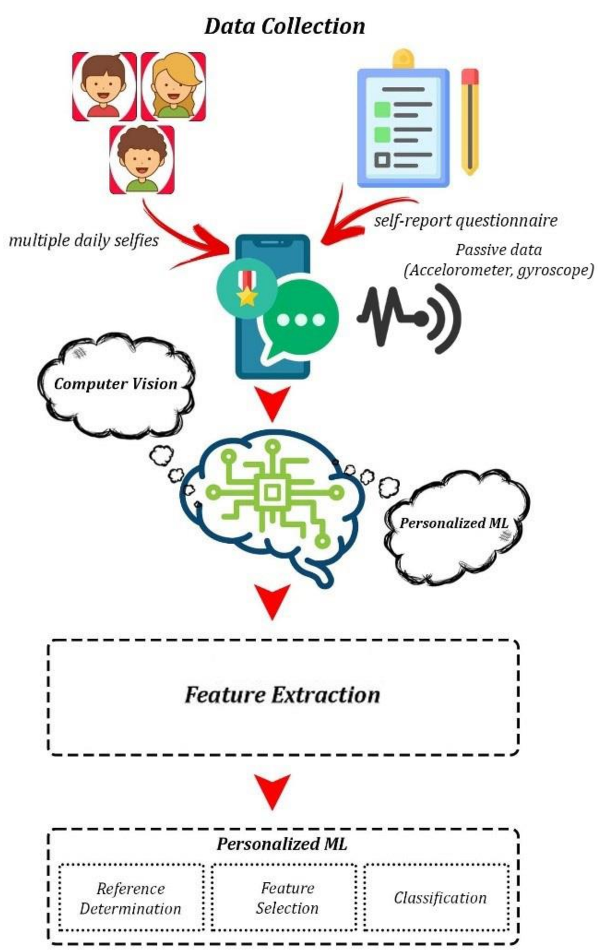

Figure 1: The study workflow for the STAND mobile application, encompassing data collection, feature extraction, feature selection, model training, and performance evaluation.

2.2 Software Development

A cross-platform mobile application called STAND has been developed (both in Android and iOS). This application will be utilized in future research to investigate the stress/anxiety levels of our target users and collect the annotated data required for both stress/anxiety recognition and stress/anxiety level estimation tasks. This data will be collected by relevant questionnaires sent to the users directly through daily notifications. Within each questionnaire, there is an extra feature implemented to enable the users to take selfies and also upload video recordings taken while doing neuropsychological tasks; this process helps us to extract potential facial landmarks and pose coordinates within consecutive frames extracted from video recordings to handle stress/anxiety recognition task. images will be also used for analyzing other facial expressions, such as eye closure and pupil size variation and accordingly investigate the potential relationships between these two features and the stress and anxiety levels of the target users. In the first plan, the frequency of sending questions to the users is ten times a day (each question). To incorporate gamified elements in the research, a star-rewarding system is implemented in the application, and stars are given to the participants based on points they will earn for positive and active participation during the data collection phase (e.g., doing tasks within 30 minutes since the notification is received, and fully answering all questions sent per day while have an active participation in the other tasks during the first phase). Those who earn more stars will have more chances to win the incentives through drawing.

The advantage of using the Internet in healthcare services is to overcome some of the limitations posed by traditional methods in this field. Advantages such as availability and accessibility at all times/places, anonymity, flexibility, self-pacing, cost-effectiveness for both participants and clinicians (reduced travel time and costs), and the appeal of interactivity by utilizing visually attractive designs [96-98].

This work is also going to analyze the effects of using animations, informal language in notifications (Figure 2), and other appealing factors in increasing the levels of user engagement in digital interventions (e.g., animations are chosen based on an answer to a question in the registration form, just after opening the app for the first time: Dog person, or a cat person? (Figure 3, 4)).

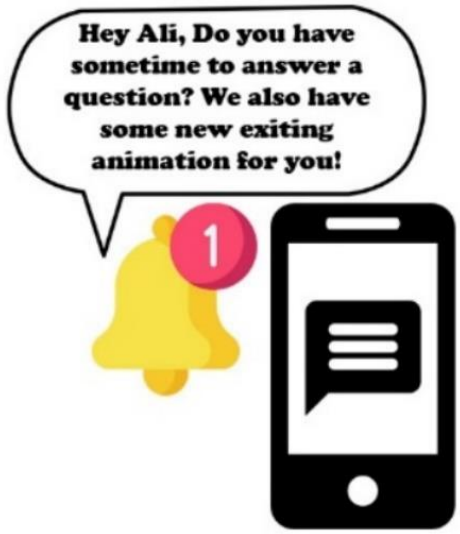

Figure 2. Using informal and motivating language for daily notifications.

The integration of visually appealing design elements, such as customized and dynamically changing animations over time, coupled with the utilization of notifications featuring informal and motivating language, can significantly enhance user engagement within the app. These design choices aim to captivate users' attention, evoke positive emotional responses, and create a sense of personal connection and motivation. By employing visually enticing and interactive animations, the app can create a dynamic and immersive user experience, fostering a sense of delight and enjoyment. Furthermore, the use of notifications that employ informal and motivating language can establish a friendly and supportive tone, encouraging users to actively participate in the assessment process. This approach seeks to establish a positive user-app interaction, fostering sustained engagement and ultimately leading to more reliable and comprehensive data collection.

Gamification elements like rewards and points can create a sense of achievement and satisfaction, fostering intrinsic motivation and encouraging active participation. The star-rewarding system in the application aligns with the principles of behavioral reinforcement, as individuals are more likely to engage in desired behaviors when they are rewarded for their efforts. By earning stars based on their positive and active participation, participants are incentivized to complete tasks within the designated timeframe and fully respond to the questions posed to them (Figure 5).

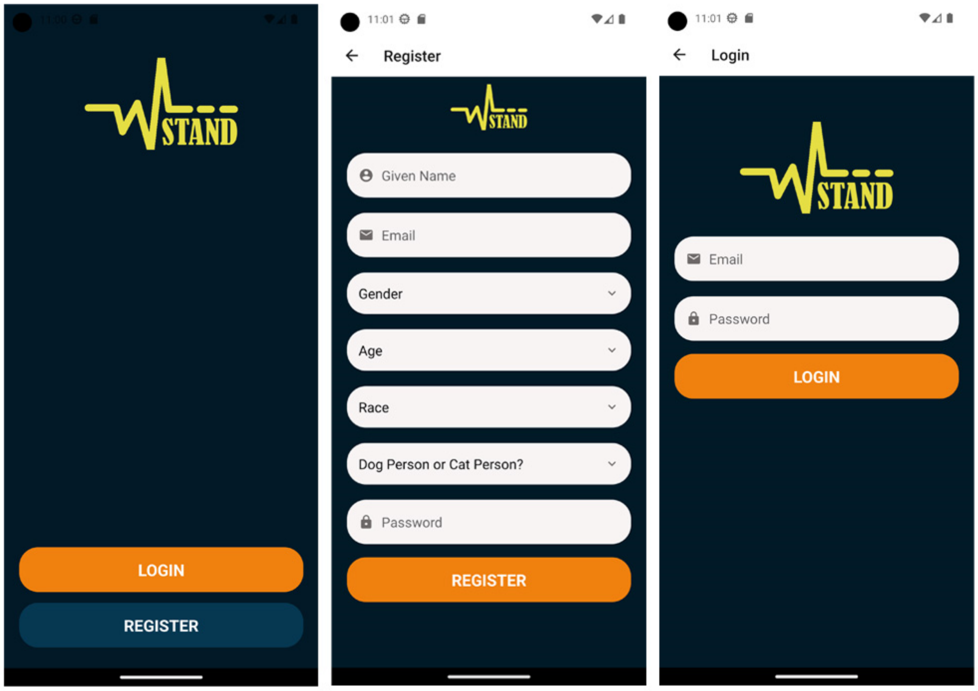

Figure 3. Welcome screen, Login Screen, and Register Screen

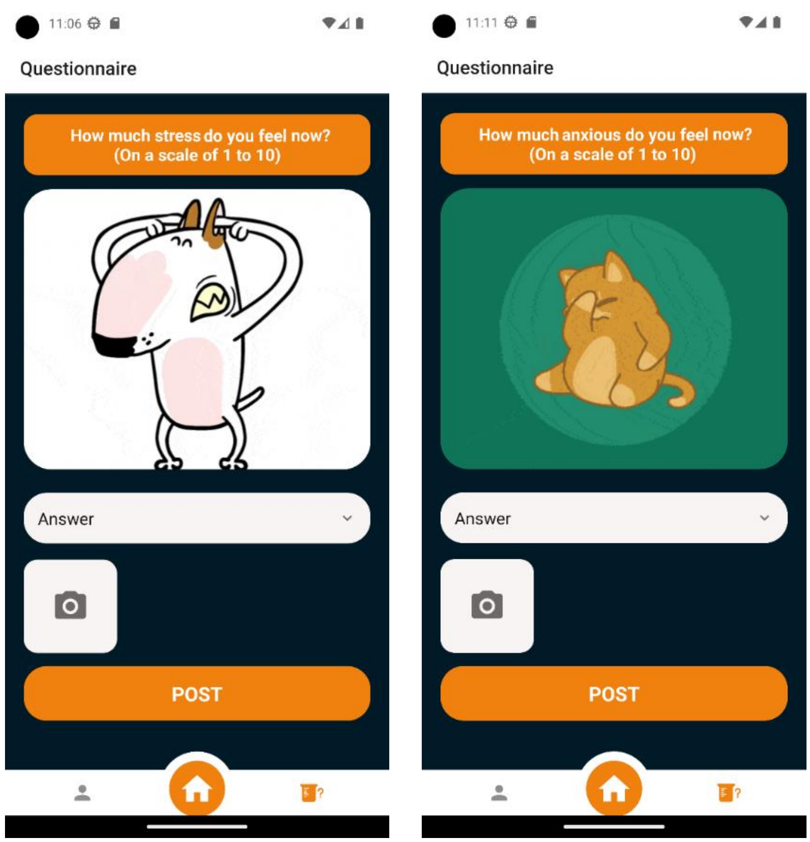

Figure 4. Screenshots of the questionnaire screen built in in the STAND EMA mobile application.

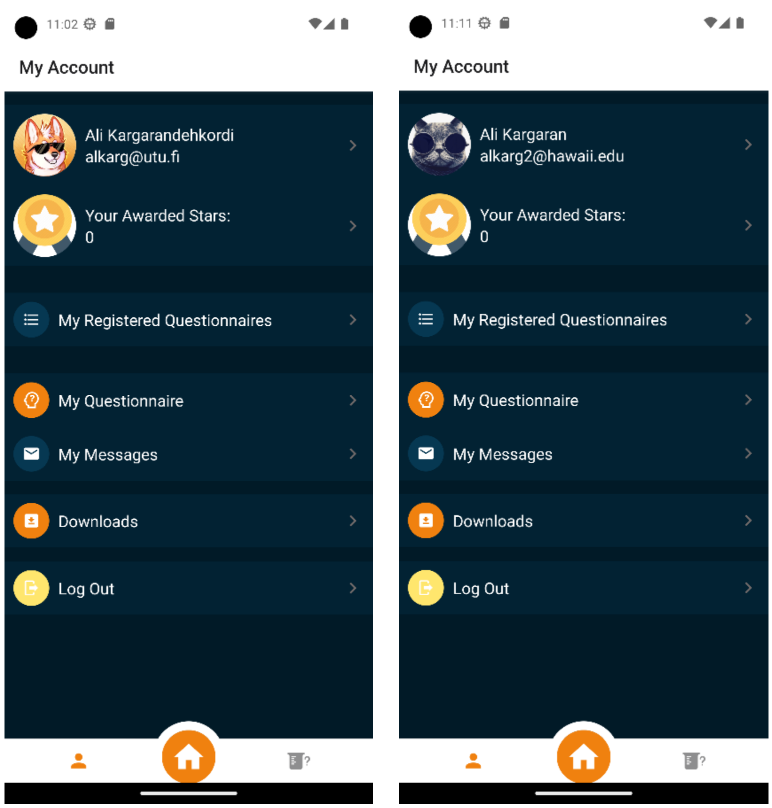

Figure 5. Screenshots of the accounts screen built in in the STAND EMA mobile application for both type of users interested in petting cat or dog.

The implementation of passive sensor data collection, specifically accelerometer and gyroscope data, within the STAND mobile application provides a novel avenue for future exploration and analysis in the realm of stress and anxiety. By continuously gathering data from these built-in sensors, the application can capture subtle movement patterns and physical behaviors that may be indicative of an individual's stress and anxiety levels. The passive nature of data collection eliminates the need for active user input, allowing for unobtrusive and uninterrupted monitoring throughout the day. This wealth of sensor data holds promises for uncovering new insights and patterns related to stress and anxiety, providing valuable information for further research and analysis. The analysis of accelerometer and gyroscope data can potentially reveal correlations between specific physical movements and the presence or severity of stress and anxiety symptoms. By integrating this passive sensor data collection feature, the STAND mobile application offers a comprehensive approach to understanding and monitoring individuals' mental well-being, contributing to advancements in the field of digital health and providing opportunities for personalized interventions and support.

2.3 Future Larger Scale Data Collection and Research Using STAND

The aforementioned procedure will be conducted as part of a larger-scale research initiative conducted by the Digital Health Lab at the University of Hawaii at Manoa [99]. This comprehensive study spans a 1-month duration, during which an ample amount of data will be gathered to meet the requirements of personalized machine learning algorithms. Throughout the designated period, participants will be requested to complete a daily self-report questionnaire aimed at assessing relevant psychometric properties. The questionnaire comprises two distinct questions (listed below) and is intended to be delivered through ten separate notifications each day, with one notification per question. Upon receiving a notification, participants are required to rate their current level of stress and anxiety on a scale ranging from 1 to 10.

- How much stressed do you feel now?
- How much anxious do you feel now?

Figure 2 and figure 3 show a few screenshots of the application, illustrating the different screens (for both cat person and dog person) participants will visit during the data collections phase.

The participants are also asked to attend two neuropsychological tests several times during the same period. Participant are being recorded while doing these tests in order to collect sufficient data for the mental state recognition (stress and anxiety) task. The similar procedure of how the facial landmarks and other pose coordinates that are going to be extracted and selected for personalized ML models, are discussed comprehensively in the second part of this thesis in chapter 3.

In the initial phase of the study, we will extract key features from the selfie images obtained from users through the STAND mobile application, which is specifically designed for collecting data on a daily basis, multiple times a day. The primary emphasis will be placed on extracting two pivotal features: eye closure and eye pupil size variation. Additionally, this study will delve into the investigation of other facial landmarks and changes in Action Units (AUs) across frames as they unfold over time. By leveraging the annotations provided by the users regarding their perceived level of stress and anxiety during the selfie-taking process, the data will be labeled accordingly. This labeling approach enables us to associate the extracted eye features with the levels of stress and anxiety experienced by the users. Apart from determining the stress and anxiety levels among our target users, this procedure also allows us to explore potential connections between the eye features and stress and anxiety. Through this investigation, we aim to gain insights into the relationship between these eye features and the emotional states of stress and anxiety, providing valuable information for understanding and managing stress and anxiety levels in individuals.

In a second procedure, we aim to train three distinct machine learning models using facial landmarks and pose coordinates extracted from video recordings of participants. The data will be obtained through the utilization of the OpenFace tool, which allows for the extraction of data at the frame level. This extracted data will then undergo a feature estimation and selection process to identify the most relevant and informative features for the classification task. The objective of this parallel procedure is to perform binary classification to determine whether the user exhibits symptoms of stress or anxiety (the experiment participants are the student who have been diagnosed recently with a certain level of stress and anxiety). To achieve this, we will employ K-nearest neighbors (KNN), Random Forest, and Multi-Layer Perceptron (MLP) classifiers, known for their effectiveness in classification tasks. By leveraging these models, we aim to develop accurate and robust classification models that can discern the presence or absence of stress and anxiety based on facial landmarks and pose coordinates. This research holds potential implications for identifying and understanding emotional states in individuals, paving the way for further advancements in emotion recognition and mental health assessment.

2.3.1 Ethics Approval and Privacy

This research project underwent an accelerated review process and received approval from the University of Hawaii Institutional Review Board (UH IRB) on December 20, 2022. The application met the criteria for expedited review as specified in 45 CFR 46.110, Category 6, 7, in accordance with 45 CFR 46.109.

Participants were assured that their identities would remain confidential and that participating in the survey carried no greater risk than regular internet use. The collected data was securely stored on Amazon Web Services (AWS), with additional security measures including multi-factor authentication (MFA) for authorized access.

2.3.2 Study Group

Stress and anxiety are prevalent mental health issues experienced by students worldwide, with many facing elevated levels of stress in their daily lives [100, 101]. Although considerable research accomplished on the context, there are still a massive number of students who remain underserved, or intentionally ignore this mental concern [102]. Moreover, studies that concentrate on identifying and addressing subjective health concerns often encounter challenges in obtaining accurate data from student populations. This particular demographic may exhibit a higher level of reluctance in seeking healthcare services due to various obstacles such as time constraints, social stigma, scheduling conflicts, and heavy academic workloads [103]. The primary objective of this study is to examine and identify anxiety levels among students, with a specific focus on their mental well-being. By doing so, valuable insights can be obtained to assist school counselors in understanding the causes of anxiety disorders and implementing effective evidence-based prevention and intervention strategies.

2.3.3 Recruitment, Demographics, and Incentives

The participants in our future larger scale study are going to be between 20 to 40 students (gender balances) aged 28 ± 10 years who have been diagnosed with a certain level of anxiety and stress in a recent clinical procedure. To be eligible for this study, it is required to have proficiency in English and possess a smartphone (either iOS or Android). Exclusion Criteria: Students who have not experienced a certain level of stress and anxiety during the last few years.

Through the form of course credit at the University of Hawai'i at Manoa, letters will be sent online to the prospective students. Face to Face interactions with other prospective students, and Internet ads are alternative options for recruitment. The participants will be interviewed shortly in the first stage for potential examinations and to check if they are a good fit for this project. In the case the outcome of the interview was positive, written consent will be collected from target users outlining the comprehensive research circumstances.

Upon the completion of the data collection stage, students at the ICS department in the University of Hawaii at Manoa who are going to take the Machine Learning course with Dr. Peter Washington will be awarded with course credits after positive participation. Other potential participants (students) will have the chance to win a $20 supermarket gift card for grocery shopping in a fair drawing because of positive participation in the research. The drawing will be held right after the data collection phase, and winners will be contacted through an email containing instructions on how to receive the gift card.

2.4 Chapter Summary

This chapter delves into the development of the STAND mobile application, designed specifically for the purpose of stress and anxiety monitoring. The application incorporates various innovative features and methodologies to facilitate accurate data collection and analysis, with the aim of advancing our understanding of mental states and fostering personalized interventions.

The chapter begins by highlighting the significance of remote mental health monitoring and the role of web and mobile applications in enabling scalable and accessible solutions. It emphasizes the need for personalized machine learning approaches to recognize mental states based on individual profiles, rather than relying solely on general-purpose solutions.

The development of the STAND application is introduced as a pivotal component of the research. The application leverages the concept of Ecological Momentary Assessment (EMA), providing users with daily notifications that redirect them to a user-friendly interface. This interface incorporates three essential components: a question regarding current stress and anxiety levels, a rating scale for quantifying responses, and the option to capture a selfie photograph. These data points are collected individually, allowing for nuanced analysis of stress and anxiety levels.

Furthermore, the STAND application includes additional features to enhance user engagement and data collection. Gamified elements, such as a star-rewarding system, incentivize positive and active participation, fostering intrinsic motivation. Passive sensor data collection, utilizing the accelerometer and gyroscope, enables continuous monitoring of physical behaviors that may indicate stress and anxiety levels. This unique feature ensures unobtrusive data collection and offers opportunities for future exploration and analysis.

3 Second Part: AI solution on Emotion Recognition

3.1 Overview

For collecting the data through STAND EMA mobile application, we encountered various challenges in collecting data, including scale and time limitations for concluding a master's thesis and also lack of required funding at earlier stages. To overcome these challenges, we decided to simulate the AI component of the thesis using a publicly available dataset known as the "Emognition Wearable Dataset 2020 [104]" This dataset is highly relevant to emotion recognition research and provides a rich collection of data for our analysis.

As some of the mental states such as stress and anxiety can also be regarded as representations of emotions, the Emognition dataset serves as a suitable resource for testing methods in emotion recognition. The dataset includes recordings from 43 participants who viewed short film clips designed to elicit nine distinct emotions: amusement, awe, enthusiasm, liking, surprise, anger, disgust, fear, and sadness. Additionally, the dataset contains physiological data such as EEG, BVP (2x), HR, EDA, SKT, ACC (3x), and GYRO (2x).

Given the nature and scope of our thesis work, we focused solely on the video recordings and their corresponding facial expressions, landmarks and AUs extracted from the consecutive frames (images), disregarding the physiological data. This allowed us to train the model specifically on the most relevant data for our future larger-scale research objective [99, 105]. The dataset provides detailed information on facial expressions for each individual frame of the video recordings, corresponding to each discrete emotion stimulus and participant. For instance, eye gaze, and landmarks coordinates, pose landmarks and estimations that are translation and rotation of the face in 3D space. They provide information about the position and orientation of the face relative to a reference point. Facial Action Units (FAUs) are also included. Facial Action Units (FAUs) are a set of distinct facial muscle movements or expressions that are universally recognized and can be objectively measured and categorized. They provide a standardized way to describe and analyze facial expressions in the field of facial recognition, emotion detection, and non-verbal communication. Each Facial Action Unit corresponds to the activation or contraction of specific facial muscles, leading to the formation of different facial expressions.

The comprehensive system of Facial Action Units allows researchers and professionals to break down and analyze the complexity of facial expressions, providing valuable insights into human emotions, intentions, and communication patterns.

3.2 Literature Review on Emotion Recognition

Our mental well-being is influenced by emotions, as they impact how we perceive things, affect our physical vitality, and can disrupt our cognitive processes. Emotions are unique inherent phenomena that encompass both the mind and the body. They regulate our physiological functions, impacting factors such as heart rate, respiration system, and body temperature. Additionally, they shape our psychological state, influencing our perception, beliefs, and mental imagery. By utilizing various outputs (either sequentially or simultaneously), emotions strive for manifestation to fulfill their need for expression. Depending on the context, physical state, available means of communication, and individual preference, emotions may be conveyed through a combination of facial expressions, vocalizations, gestures, and body movements [106, 107].

The ability of intelligent machines to express, recognize, and regulate users' emotions has become a vital component of emotional artificial intelligence. This capability enhances the overall user experience by facilitating effective interactions between machines and individuals [108]. Various methods and technologies are being used to identify users' emotions during their interactions with machines, whether through active or passive means. Emotion recognition techniques encompass analyzing text, interpreting facial expressions, assessing voice tones, utilizing biosensors, observing body movements, and recognizing gestures. These approaches can be used individually (single-modal) or in combination (multi-modal) to achieve precise emotion detection.

The rise of publicly available datasets and the capabilities of machine learning (ML) techniques have popularized automatic emotion recognition. Among the various methods employed, some studies have focused on utilizing handcrafted features and applying traditional ML algorithms like support vector machine (SVM), logistic regression, and random forest (RF) for pattern classification [109, 110]. While these techniques have shown promising results, they require significant effort and time to extract features, and there is a risk of losing valuable information specific to certain emotions due to the dependence on the researcher's domain knowledge.

In recent years, deep learning techniques have emerged as highly successful in video recognition tasks, and several studies have demonstrated significant improvements in emotion recognition performance compared to traditional approaches that utilize partial representation. These deep learning approaches excel in automatically learning effective hierarchical representations of emotions without the need for manual intervention [111]. Despite the proven performance of both conventional machine learning methods and deep learning methods, there are ongoing debates regarding their overall effectiveness [112]. The performance may vary depending on the specific task, and it remains unknown which approach achieves superior accuracy and results.

Extensive research has demonstrated the superior performance of deep learning methods in emotion recognition tasks, particularly when applied to large datasets [111-113]. These deep learning models excel at automatically extracting intricate patterns and features from vast amounts of data, enabling them to capture complex relationships and nuances within emotions. This advantage is primarily attributed to their ability to learn hierarchical representations without the need for explicit feature engineering. However, in our research, we are taking a personalized approach to develop a unique and personalized emotion recognition model for each individual. By leveraging existing data specific to each person, we aim to create tailored models that accurately capture and interpret their unique emotional expressions. This personalized machine learning technique allows us to build models that are specifically trained on individualized data, taking into account their specific emotional cues and characteristics.

By adopting this personalized approach, the comparison between conventional machine learning techniques and deep learning techniques becomes more equitable and realistic in our research. While deep learning methods have shown superior performance on large-scale datasets, our focus on personalization allows us to explore the effectiveness of these techniques in capturing individual nuances and characteristics of emotions. This personalized approach may provide valuable insights into the feasibility and effectiveness of deep learning methods in individualized emotion recognition, potentially uncovering new possibilities for personalized emotional AI systems.

Academic evidence supports the idea that personalization can enhance the performance of emotion recognition systems. Research studies have demonstrated the benefits of incorporating personalized data in developing accurate and robust emotion recognition models [16, 114, 115]. This highlights the potential of personalized machine learning techniques in unlocking more precise and context-aware emotion recognition capabilities compared to generic models.

As previously stated, academic investigations in the realm of deep learning have extensively examined the benefits of deep neural networks for handling vast amounts of data and extracting complex features. These studies have demonstrated that deep learning approaches have the potential to surpass traditional ML methods in dealing with extensive datasets. Nonetheless, the utilization of deep learning techniques for personalized emotion recognition necessitates additional exploration, and our research aims to make a contribution to this developing domain.

Leveraging the versatile Emognition dataset, by combining personalized machine learning techniques with the comparison of conventional ML techniques and deep learning techniques, our research aims to provide valuable insights into the efficacy of both approaches within the context of individualized emotion recognition. Through rigorous experimentation and analysis, we seek to advance the understanding of personalized emotion recognition systems and contribute to the development of more accurate and context-aware emotional AI technologies.

3.3 Prior Works and Background

Facial emotion recognition has become a subject of great interest for researchers across various fields. Particularly in Clinical Medicine, it is an active and evolving research topic that presents a significant challenge [116]. The current key challenge of facial emotion recognition stems from the inherent uncertainties and inaccuracies in mapping emotional states to detectable cues. Nonetheless, significant advancements have been made in this field, with numerous approaches and methods proposed for expression recognition [117, 118].

Numerous researchers have suggested the utilization of local binary pattern (LBP) to transform an image into a configuration of micro-patterns [119, 120]. In 2009, Shan et al. [121] conducted an extensive experiment to evaluate the effectiveness of LBP features in recognizing expressions. The results of the experiment confirmed that LBP features possess a certain level of efficiency. However, despite their usefulness, facial expression recognition methods based on

LBP also suffer from challenges including low accuracy in recognition and vulnerability to interference [122]. In other research, the use of histogram of oriented gradients (HoG) features has been applied to recognize facial expressions. In Dahmane et al.'s study, a combination of feature extractors, such as LBP, PCA, and HoG, was employed together with a SVM classifier to categorize static face images into six distinct emotions [123]. In a study conducted in 2012 [124] utilized Haar wavelet features along with multiscale analysis and statistical analysis to recognize facial expressions. However, the use of Haar-based facial expression recognition presents challenges, including a high rate of false recognition and incomplete extraction of facial expression information. Another method called scale-invariant feature transform (SIFT) was employed by [125] in 2011 to describe face pose and achieve expression recognition through the extraction of principal component information using singular value decomposition (SVD). Nevertheless, the utilization of the SIFT-based method for expression recognition faces obstacles such as limited computational efficiency and vulnerability to dimensionality problems.

An alternative strategy for recognizing facial expressions involves the application of deep learning techniques. In 2006, Hinton introduced the layer-by-layer training approach to tackle the complex task of training neural networks with multiple layers [126]. As a result, robust open-source learning frameworks such as Torch, Caffe, Deep Learn Toolbox, and Cxxnet were created, supported by substantial contributions from researchers and institutions. Deep learning has the capability to approximate high-dimensional data spaces, making it well-suited for learning intricate functions and extracting high-dimensional feature representations from images. While initially used for object image classification, deep learning has gradually found applications in face recognition as well [99, 105, 127].

In this research [128], conducted in 2014, a deep belief network (DBN) was utilized for feature learning and combined with boosting for classification. This approach yielded better outcomes compared to conventional feature-based methods when applied to the CK expression database. Similarly, in 2015, Lopes et al. employed convolutional neural networks (CNNs) for feature learning and augmented image data. Their findings showcased the advantages of convolutional feature learning in expression classification and highlighted the performance enhancements achieved with larger amounts of data [129]. Lecun et al. pioneered the early use of convolutional neural networks (CNNs) [130]. Compared to other models, CNN exhibits several advantages as it takes raw images as input instead of hand-coded features. Convolutional neural networks (CNNs) excel in learning optimal features for classification tasks. With their diverse layers, including convolutional and sub-sampling layers, CNNs have transformed computer vision and significantly improved various applications. Emotion recognition has also benefited from the use of CNNs in numerous studies. The study [129] devised a five-layer convolutional neural network (CNN) to categorize six specific emotions in the Cohn-Kanade (CK+) database. In the work [131], they presented a CNN architecture with two channels, with convolutional filters used in the upper channel and Gabor-like filters employed in the lower channel's initial layer. [132] employed convolutional modules within a distinctive CNN framework, designed to analyze the most impactful features for the next layer, minimizing repetition of learned features, and taking into account the similarities between filters within the same layer. In [133] They suggested long-term recurrent convolutional networks as a solution for visual recognition and description tasks.

In 2016, Zhang and colleagues presented a new method for recognizing facial expressions that are invariant to attitude. Their approach involved combining deep learning techniques, a principal component analysis network, and convolutional neural networks (CNN). Through extensive experiments on two publicly available databases, they demonstrated substantial enhancements in their method compared to traditional techniques used for expression recognition [134]. In 2017, [135] introduced in their study is an algorithm for extracting facial expressions using deep learning. They conducted an analysis of the existing research in this field and compared different methods. The findings revealed that deep learning techniques excel at extracting hierarchical features and leveraging them for image classification based on expressions. Consequently, these methods significantly improve recognition accuracy when compared to conventional approaches [136-138].

The DEAP database [139], compiled by researchers including Koelstra from various universities, serves as a valuable resource for studying human emotional states through multi-channel data. This publicly available database contains recordings of EEG signals and physiological signals (PPS) from 32 subjects. Researchers frequently utilize DEAP to explore and analyze the intricate aspects of human emotions. Tang et al. [140] and Yin et al. [141] are two additional studies that employed a multimodal approach to emotion recognition. Both studies utilized deep neural networks in conjunction with the DEAP dataset, which encompasses various modalities of data such as EEG signals and physiological signals. By leveraging these multimodal data sources, Tang et al. and Yin et al. aimed to enhance the accuracy and robustness of their emotion recognition systems. The eNTERFACE'05 dataset [142] is a widely used benchmark dataset in the field of facial expression analysis and emotion recognition. The dataset consists of synchronized video recordings of facial expressions along with corresponding emotion labels. It includes expressions of basic emotions such as happiness, sadness, anger, surprise, fear, and disgust. The dataset is valuable for developing and evaluating algorithms and models for facial expression analysis, as it provides a diverse range of facial expressions and emotions in different individuals and scenarios. Zhang et al. [143] and Nguyen et al. [144] are two notable research studies that focused on the task of utilizing the eNTERFACE'05 dataset for their respective investigations. In their work, Zhang et al. and Nguyen et al. recognized the potential of the eNTERFACE'05 dataset as a valuable resource for their research on a specific topic. By leveraging the diverse and comprehensive nature of the eNTERFACE'05 dataset, they were able to conduct in-depth analyses and draw meaningful conclusions. These studies contribute to the growing body of knowledge surrounding the eNTERFACE'05 dataset and its applicability in various research domains.

While deep learning has made significant progress in facial expression recognition, there are still challenges that need to be addressed. These challenges include overfitting, gradient explosion, and parameter initialization issues in deep learning models. Furthermore, current algorithms face difficulties in effectively utilizing dynamic sequence information from expression images and lack robustness when applied in practical scenarios.

Vision Transformers (ViTs) have recently garnered significant attention as advanced deep learning models for computer vision tasks, specifically in the area of facial recognition. ViTs offer a modern and alternative approach to the commonly used convolutional neural networks in similar tasks. Inspired by the Transformer architecture originally designed for Natural Language Processing (NLP) tasks, ViTs operate by dividing input images into fixed-size patches, treating them as sequences of flattened vectors. These patches are then processed by a series of Transformer encoder layers, enabling the capture of both local and global information. Notably, ViTs have demonstrated impressive capabilities in learning meaningful representations directly from raw facial images, eliminating the need for manual feature extraction by researchers.

However, training ViTs for facial recognition typically requires a large labeled dataset of facial images and significant computational resources, and their performance might be limited when applied to smaller datasets. Additionally, fine-tuning and transfer learning approaches can be employed to adapt pre-trained ViT models to specific facial recognition tasks with smaller datasets [145].

3.4 Emognition Dataset [104]

The Emognition dataset includes data from 43 participants (21 females aged between 19 and 29) exposed to emotionally stimulating film clips representing nine distinct emotions (Table 1). Physiological signals and upper-body recordings were collected using wearable devices such as Muse 2, Empatica E4, and Samsung Galaxy Watch. The dataset offers advantages over previous datasets, including the integration of wearable devices into daily life, the use of discrete and dimensional emotional models, and a specific focus on distinguishing between positive emotions. The dataset provides valuable insights into the physiological responses associated with different emotional states, particularly within the realm of positive emotions.

The dataset allows for versatile analysis in emotion recognition, covering both physiological signals and facial expressions. The Emognition dataset can be utilized to address various research questions, including:

- Exploring multimodal approaches to emotion recognition.
- Comparing physiology-based emotion recognition with facial expression-based emotion recognition.
- Investigating emotion recognition using EEG signals versus BVP signals.
- Comparing emotion recognition using the Empatica E4 device with emotion recognition using the Samsung Watch (both devices provide BVP signals recorded in parallel).
- Classifying positive emotions versus negative emotions.
- Examining affect recognition in terms of low arousal versus high arousal and valence.
- Analyzing the relationship between discrete and dimensional models of emotions.

The exclusion criteria for participants in the study included significant health problems, drug use affecting cardiovascular function, previous cardiovascular disease diagnosis, hypertension, and a

BMI over 30. Measures were taken to minimize factors influencing cardiovascular function, such as avoiding alcohol, psychoactive drugs, caffeine, smoking, nonprescription medications, vigorous exercise, and eating before the study.

Table 1: stimulus clips used to elicit emotions [104]

Targeted emotion (Polish translation)	Source film	Scene	Duration [min.]	Ref.
Anger (złość)	American History X	A neo-nazi smashes a Black man's head on the curb killing him	02:00	22
Fear (strach)	The Blair Witch Project	The clip begins with suspense and ends with an intense burst	02:00	22
Surprise (zaskoczenie)	Capricorn One	Unexpectedly, men are bursting through the door	00:49	21
Sadness (smutek)	Champ	A boy cries at the death of his father	01:59	21
Disgust (obrzydzenie)	Trainspotting 2	A man suffering from violent diarrhea goes to an extremely dirty public restroom	01:08	22
Amusement (rozbawienie)	A Fish Called Wanda	Unexpectedly, the owners of the house get into the house and discover Archie dancing naked	02:00	21
Enthusiasm (entuzjazm)	London 2012	A montage of moments showing athletes' successful performance and their joyful reactions	01:59	20
Awe (zachwyt)	NEW York from PONE	A montage of architecture in a modern city	01:56	20
Liking (pragnienie)	Food	A presentation of desserts	01:51	20
Neutral (neutralny)	Blue	A woman goes up an escalator, carrying a box	02:01	22

The short film clips listed above were all from the databased with prior evidence of reliability and validity in eliciting targeted emotions [146-150].

Two types of self-assessment were used to check the manipulation of emotions: discrete and dimensional approaches. The discrete approach involved participants rating their levels of targeted emotions using single-item rating scales. The dimensional approach utilized the Self-Assessment Manikin (SAM) to assess valence, arousal, and motivation. Participants indicated their emotions on graphical scales. The main experiment included ten iterations with washout clips, emotional film clips, and self-assessments. The order of film clips was randomized using a Latin square design. After the experiment, participants provided additional information and received a voucher. The entire procedure lasted about 50 minutes (Figure 6).

Figure 6. The experiment procedure [104]

The upper-body recordings underwent processing using the OpenFace toolkit [151] (version 2.2.0, default parameters) and Quantum Sense software (Research Edition 2017, Quantum CX, Poland). The OpenFace library was employed to extract facial landmark points and action units' values, while Quantum Sense was used to recognize basic emotions and head pose.

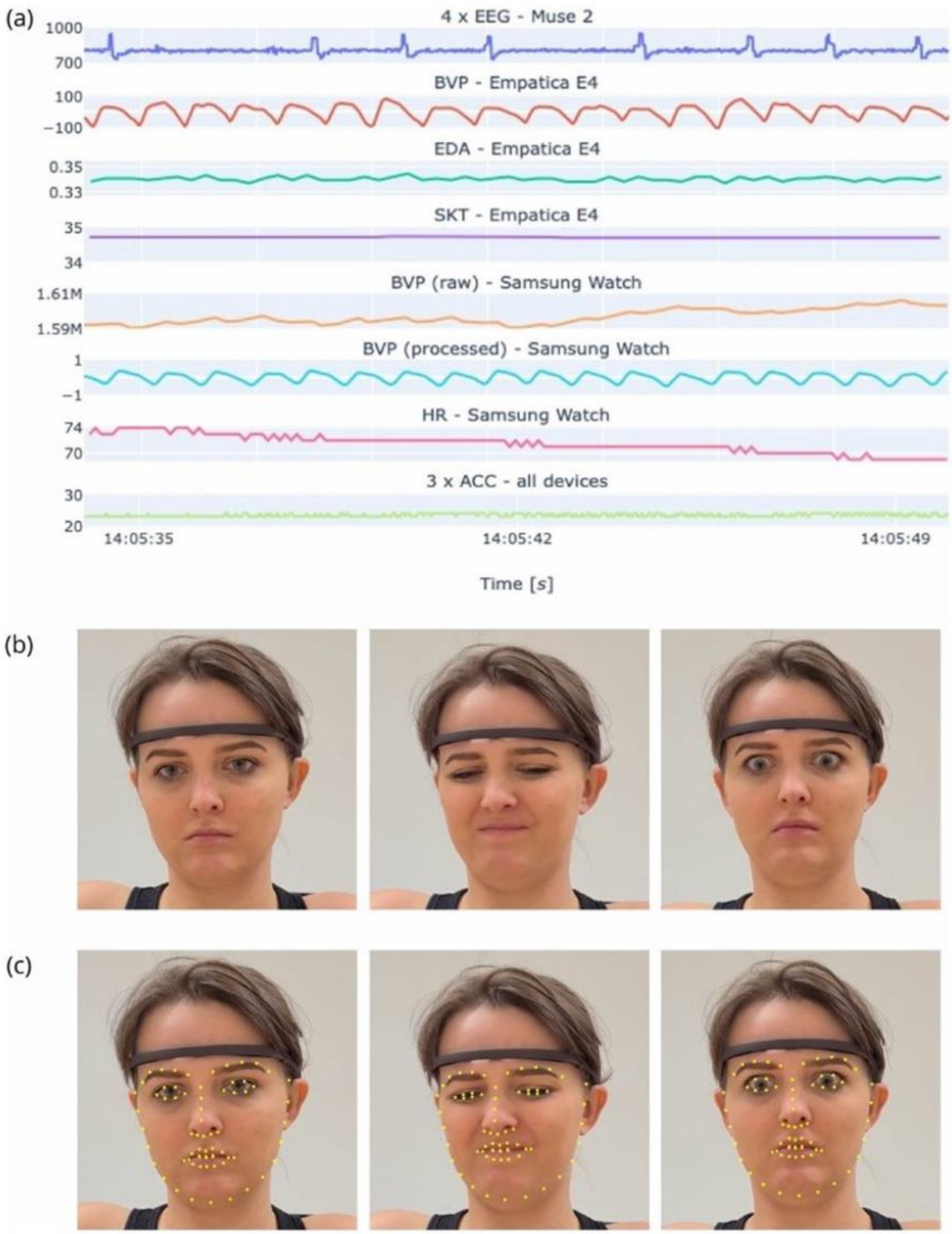

Figure 7. Various types of data in the Emognition dataset [104].

Figure 7 illustrates the diverse range of data included in the Emognition dataset: (a) Physiological signals captured by wearable devices include 4 x EEG signals from Muse 2, blood volume pulse (BVP), electrodermal activity (EDA), and skin temperature (SKT) measured by Empatica E4, as well as raw BVP, processed BVP, heart rate (HR), and accelerometer (ACC) data from Samsung Watch. (b) Upper-body recordings feature facial reactions to the stimuli, showcasing expressions of neutral, disgust, and surprise. (c) Facial landmarks, generated using the OpenFace library, contribute to emotion recognition by providing valuable facial feature information.

The collected data, which includes physiological signals, upper-body recordings, self-reports, and control questionnaires, is available at the Harvard Dataverse Repository [104]. The Emognition dataset consists of high-resolution MP4 files for upper-body recordings, occupying approximately 76GB of space. The remaining data is compressed into a study_data.zip package, with a size of 1GB (16GB after decompression). The data is organized by participant, with each participant having their own folder containing files from different experimental stages and devices used. Each participant's folder contains a total of 97 files, including signal files for film clips and baseline measurements, as well as a questionnaires.json file containing self-assessment responses and metadata.

Facial annotations are provided in two separate ZIP packages: OpenFace.zip and Quantum.zip. The OpenFace package contains facial landmark points and action units' values, stored in CSV format. On the other hand, the Quantum Sense package includes values for six basic emotions and head position, stored in JSON format. The files in both packages are assigned per video frame. The compressed size of the OpenFace.zip package is approximately 7.4GB, which expands to 25GB after decompression. In comparison, the Quantum.zip package is around 0.7GB when compressed and 4.7GB after decompression. For more technical details, including file naming conventions and the available variables in each file, researchers can refer to the README.txt file provided within the dataset.

3.5 Experimental Procedures

In this section, we will delve into three specific tasks centered around emotion recognition. These assignments have been thoroughly evaluated using a personalized machine learning approach, specifically applied to the Emognition Dataset. The process involves several key steps, namely data preparation, data exploration, model selection, and performance estimation for each model. By adopting a personalized machine learning approach, we aim to tailor the models and techniques to individual needs and characteristics, ultimately enhancing the accuracy and relevance of the emotion recognition system.

To initiate the analysis and classification process of the data, we have made a deliberate choice to utilize three widely acknowledged and extensively used models: K-Nearest Neighbors (KNN), Random Forest, and Multi-Layer Perceptron (MLP). These models have earned significant recognition within the field of classification tasks due to their proven effectiveness and reliability. By harnessing the inherent capabilities and strengths of these models, we can delve deeper into the dataset, uncovering valuable insights and achieving precise categorization of emotions. To better illustrate the workflow and procedure of the personalized machine learning approach, Figure 8 has been included. This visual representation showcases the sequential steps involved in the process, providing a clear overview of how data preparation, exploration, model selection, and performance estimation are interconnected in achieving successful emotion recognition.

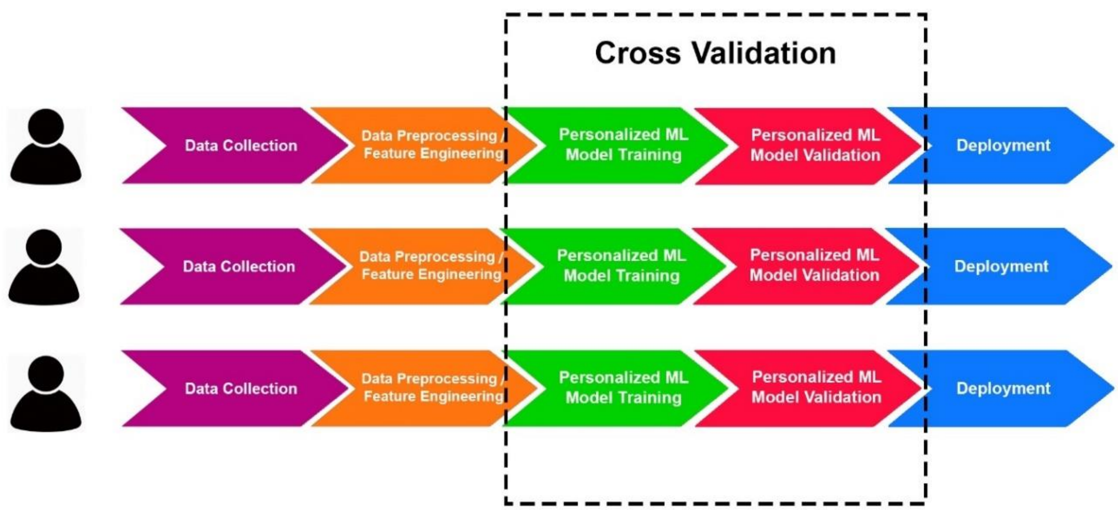

Figure 8. Workflow and Procedure of Personalized Machine Learning Approach for Emotion Recognition

Overall, through the utilization of a personalized machine learning approach and employing well-established models like KNN, Random Forest, and MLP, we aim to effectively analyze and classify emotions using the Emognition Dataset. By tailoring the models to individual needs and evaluating their performance, we strive to enhance the accuracy and applicability of emotion recognition systems in various real-world scenarios.

3.5.1 Preparations of the Data

Within the dataset, the data of each participant is organized into three separate groups to facilitate analysis and investigation. The first group, named "OpenFace," consists of CSV files that contain a comprehensive set of facial landmark points. These landmarks are obtained using the OpenFace tool, which processes video recordings frame by frame, capturing the subtle changes in facial expressions. As previously mentioned, the participants were exposed to a series of nine different emotion stimulus videos, and from these videos, data was extracted for six distinct categorical expressions: Anger, Disgust, Sadness, Neutral, Surprise, and Happiness. To simplify the analysis process and enable easy interpretation, we assign integer values to represent these emotions or classes. These values are based on a range of emotions, with the most negative emotion assigned the value of 0, and the most positive emotion assigned the value of 5 (as shown in Table 2). This mapping of emotions to integer values serves the purpose of streamlining future data analysis and facilitating further investigations in upcoming studies.

Considering the variability in emotional representations among individuals and our objective to personalize the models for each person, we had to adopt different approaches to consolidate the data for each emotion. During our analysis, we observed that the elicited emotional expressions varied from one individual to another.

Table 2. Emotion labels and their mapped numerical value

Emotion	Numerical Label
Anger	0
Disgust	1
Sadness	2
Neutral	3
Surprise	4
Happiness	5

For instance, some participants had no labels for specific emotion stimulus videos (e.g., no anger labels for anger video, no surprise label for surprise video etc.), indicating that watching those videos did not sufficiently stimulate the corresponding facial expression. This disparity in the number of labels suggests that the threshold and manner of eliciting each emotion differ from person to person. The following results (Table 3) illustrate the distribution of labels for each emotion across nine stimulus videos, along with a neutral video, for one of the participants.

Table 3. The population of each emotion labels in the in the 10 emotional stimulus videos experiments

	Amusement	Anger	AWE	Disgust	Enthusiasm	Fear	Liking	Neutral	Sadness	Surprise
Anger	0	0	0	0	0	3	0	0	0	0
Disgust	0	0	0	0	0	0	0	0	0	0
Sadness	398	213	135	227	87	111	128	1812	9	159
Neutral	6766	6752	6822	3339	7036	6993	6541	5468	7126	966
Surprise	0	37	0	0	0	116	1	0	0	0
Happiness	40	198	0	495	29	0	6	0	0	1832

As the outcomes indicate, for example for this participant, there is no labels for anger emotion in the anger video stimulus experiment. Similarly, there are no labels for surprise and disgust in their respective video stimulus experiment. In this case there was even no labels for disgust emotion, and only a few labels for anger emotion in the entire dataset, that makes it challenging to have a balanced dataset for training and classification. Because the emotions will not be evenly distributed in the dataset. However, this challenge is the main purpose for this study. This specific reason exactly shows how for specific tasks in recognition, we might need to personalize a separate model for each person. Because for example, in Digital Health, and Health Precision areas, a general model that has been trained on data for a population might not be effective for all of the individuals.

Therefore, to handle this situation our approach was to gather the emotion labels specific to one discrete emotion from all of the labels across all video stimulus video experiments and collect them in one DataFrame that is specific to one single emotion. Following results (Table 4) show the data accumulated for happiness emotion labels across all the frames of the anger video stimulus.

```
# For example happiness DataFrame of anger video
happiness_df['anger'].head()
```

Table 4. data accumulated for happiness emotion labels across all the frames of the anger video stimulus

frame	…	timestamp	…	gaze_0_x	gaze_0_y	gaze_0_z	gaze_1_x	gaze_1_y	…	AU45_c	class	class_int
3540	…	58.983	…	0.317077	0.428793	-0.845931	0.141895	0.430710	…	1.0	happiness	5
3541	…	59.000	…	0.340068	0.423029	-0.839881	0.141490	0.433410	…	1.0	happiness	5
3542	…	59.017	…	0.335010	0.425157	-0.840839	0.130892	0.441724	…	1.0	happiness	5
3543	…	59.033	…	0.327101	0.427093	-0.842969	0.077166	0.430728	…	1.0	happiness	5
3544	…	59.050	…	0.353747	0.411227	-0.840092	0.090929	0.419869		1.0	happiness	5

The table 3 is illustrating the demographic of the emotion labels in each video stimulus experiment for this specific person. This way at the end, we will gather all emotion labels specific to one discrete emotion across all video stimulus DataFrames to one single DataFrame which will enable us to create a final DataFrame consisting of all emotion labels together to train the model by (Figure 9).

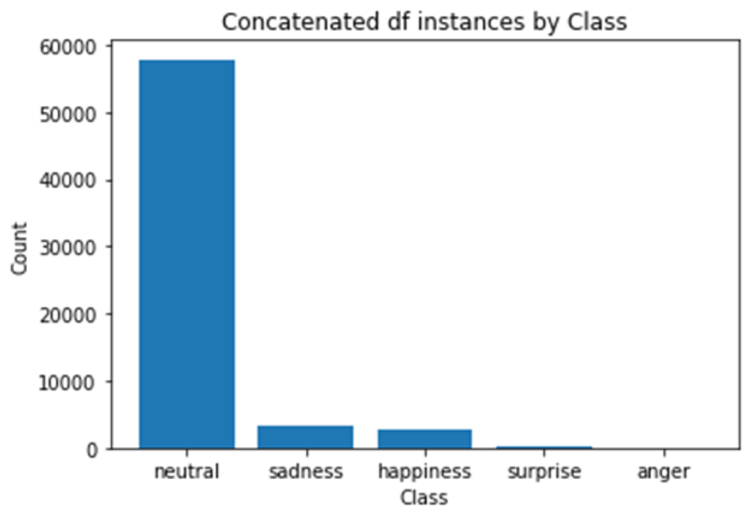

Figure 9. Plot of the concatenated DataFrame consisting the population of each 6 discrete emotions

In order to mitigate the impact of imbalanced label distribution within the dataset, a careful selection process was implemented to ensure accurate and unbiased classification results. This approach aimed to address the variations observed in the number of labels assigned to each emotion. By focusing on emotions with a substantial number of labels, we aimed to minimize the potential for misleading conclusions and biased outcomes in the final classification results. For instance, when examining this specific participant's data, it was evident that the labels for Sadness, Neutral, and Happiness were well-represented. To ensure robust analysis, a subset of 1600 instances was meticulously chosen for each of these three emotions, and subsequent classification tasks were exclusively conducted on this reduced and more balanced subset.

This approach allows for a more rigorous investigation of the emotions with sufficient representation in the dataset, thereby mitigating the potential influence of label imbalance on the classification results. By selectively focusing on emotions that have a substantial number of instances, we can enhance the reliability and generalizability of the classification outcomes.

Table 5 depict the merged DataFrames obtained by combining the chosen 3 emotion labels from the entire dataset for an individual.

Table 5. DataFrame consisting of all emotion labels together

frame	…	timestamp	…	gaze_0_x	gaze_0_y	gaze_0_z	gaze_1_x	gaze_1_y	…	AU45_c	class	class_int
3649	…	60.800	…	0.072371	0.310261	-0.947893	-0.062951	0.279758	…	0.0	happiness	5
4801	…	80.000	…	0.097013	0.332150	-0.938224	-0.092203	0.413495	…	1.0	sadness	2
107	…	1.767	…	0.029820	0.158033	-0.986984	-0.058043	0.154024	…	0.0	neutral	3
5312	…	88.517	…	-0.049779	0.244642	-0.968335	-0.132930	0.235061	…	0.0	sadness	2
4354	…	72.550	…	0.111518	0.247558	-0.962434	0.063423	0.207213		0.0	neutral	3
2470	…	41.150	…	-0.020723	0.237491	-0.971169	-0.111848	0.252432		0.0	happiness	5
1005	…	16.733	…	0.004767	0.212191	-0.977217	-0.097346	0.222793		0.0	sadness	2
2454	…	40.883	…	-0.021982	0.237019	-0.971256	-0.111571	0.245028		0.0	happiness	5
6214	…	103.550	…	-0.039225	0.219062	-0.974922	-0.130964	0.231310		0.0	sadness	2
527	…	8.767	…	-0.027562	0.258719	-0.965559	-0.152739	0.270964		0.0	neutral	3

3.5.2 Feature Selection

To facilitate feature selection for stress and anxiety detection in the initial experiment, we adopted a proactive approach. We identified certain features, namely eye gaze, eye landmarks, pose landmarks, and Action Unit (AU) values, which displayed notable effectiveness. Consequently, we focused on extracting these relevant features using their corresponding indices from the final DataFrame (Table 6). As previously explained in the dataset introduction, eye gaze and eye landmarks consisted of multiple values in each dimensional direction (x, y, z). To streamline the dataset and enhance its representativeness for the desired feature set, we employed a feature aggregation technique. This involved calculating the mean values of the relevant features within each dimensional space direction and consolidating them into a single feature. Code snippets provided bellow are for understanding the feature selection and aggregation processes conducted for the subsequent training and classification tasks.

```
# Define the column ranges
column_ranges = [
    (5, 298),
    (index, 715)
]

# Extract the desired columns
selected_columns = []
for start, end in column_ranges:
    selected_columns.extend(shuffled_df.iloc[:, start:end+1].columns)

# Create a new DataFrame with the selected columns
```

```
new_df = shuffled_df[selected_columns]

new_df.sample(10)
```

By exploring the dataset further and delving into its intricacies, we aim to extract additional relevant and evident features.

```
subsets = [
    ['gaze_0_x', 'gaze_1_x'],
    ['gaze_0_y', 'gaze_1_y'],
    ['gaze_0_z', 'gaze_1_z'],
    ['eye_lmk_x_' + str(i) for i in range(56)],
    ['eye_lmk_y_' + str(i) for i in range(56)],
    ['eye_lmk_X_' + str(i) for i in range(56)],
    ['eye_lmk_Y_' + str(i) for i in range(56)],
    ['eye_lmk_Z_' + str(i) for i in range(56)]
]

# Create a new DataFrame to store the aggregated features
aggregated_df = pd.DataFrame()

# Define custom column names for the mean columns
mean_column_names = ['gaze_x', 'gaze_y', 'gaze_z', 'eye_lmk_x', 'eye_lmk_y',
'eye_lmk_X', 'eye_lmk_Y', 'eye_lmk_Z']

# Perform mean aggregation for each subset
for subset, column_name in zip(subsets, mean_column_names):
    subset_mean = new_df[subset].mean(axis=1)   # Calculate the mean along the
row axis
    aggregated_df[column_name] = subset_mean

# Drop the original features used for aggregation
new_df = new_df.drop(columns=[col for subset in subsets for col in subset])

# Concatenate the aggregated features with the unchanged features
final_df = pd.concat([aggregated_df, new_df], axis=1)

final_df.sample(10)
```

Table 6. feature aggregation for the selected features to reduce the size of the dataset

gaze_x	gaze_y	gaze_z	eye_lmk_x	eye_lmk_y	eye_lmk_X	eye_lmk_Y	eye_lmk_Z	gaze_angle_x	gaze_angle_y	…	AU45_c	class	class_int
-0.083	0.218	-0.971	473.39	937.61	-36.25	-11.72	750.07	-0.086	0.221	…	0.0	sadness	2
-0.098	0.252	-0.961	505.44	934.88	-19.55	-13.31	760.58	-0.102	0.256	…	0.0	sadness	2
-0.058	0.216	-0.973	475.20	937.35	-35.39	-11.91	753.45	-0.060	0.218	…	0.0	neutral	3
0.019	0.295	-0.953	482.17	942.61	-32.64	-9.52	784.36	0.020	0.301	…	1.0	sadness	2
0.012	0.231	-0.968	466.46	912.81	-38.48	-24.93	750.00	0.013	0.235	…	1.0	sadness	2
-0.009	0.252	-0.965	455.77	940.16	-45.54	-10.34	748.34	-0.010	0.256	…	0.0	sadness	2
-0.044	0.252	-0.965	458.62	940.03	-43.29	-10.33	742.09	-0.046	0.255	…	0.0	happiness	5
-0.040	0.273	-0.956	550.21	941.82	5.08	-10.40	821.19	-0.042	0.279	…	1.0	happiness	5

-0.056	0.240	-0.968	468.73	946.08	-38.69	-7.17	745.94	-0.059	0.244	…	0.0	happiness	5
-0.044	0.252	-0.965	462.45	940.12	-41.99	-10.49	755.89	-0.046	0.255	…	0.0	happiness	5

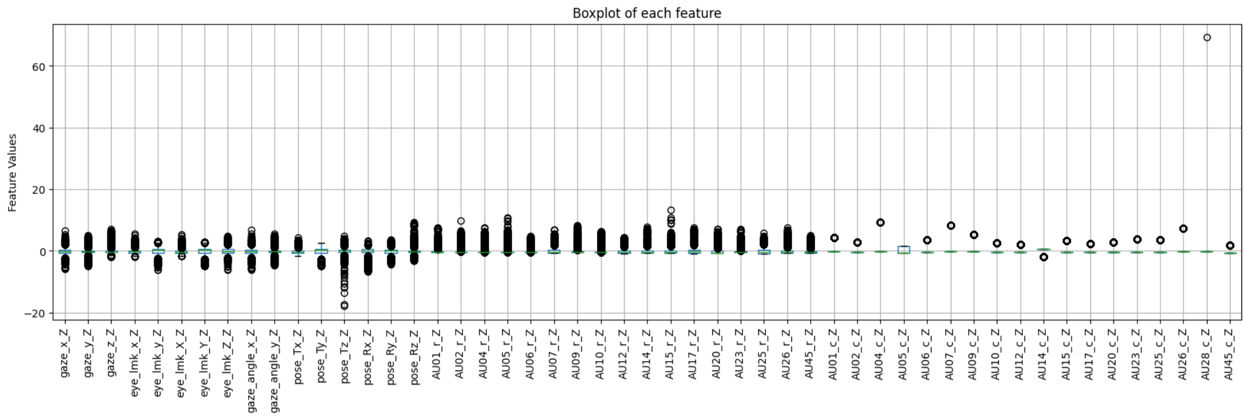

Figure 10. Boxplot of each feature

Z-score standardization: To normalize the data, we implemented Z-score standardization, which is a method for standardizing scores on a common scale by dividing the deviation of a score by the standard deviation of the dataset. This process generates a standardized score, which represents the number of standard deviations a particular data point deviates from the mean (as depicted in formula 1).

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Figure 10 is representing boxplots which are a graphical representation of the distribution of a dataset divided by each feature. It displays key statistical measures such as the median, quartiles, and any outliers or extreme values. Figure 12, is another graphical representation of the distribution of a dataset (histogram). It consists of a series of adjacent rectangles, where the width of each rectangle represents a class interval and range of values for each emotion, and the height represents the frequency or count of observations falling within that emotion interval. Histograms provide a visual summary of the data's frequency distribution, allowing us to identify patterns, central tendencies, and skewness. As illustrated in the histograms, for each emotion there are significant overlaps for detecting emotions that is exactly the challenging part of the emotion recognition tasks. It is also usually different form person to person and that is the reason that makes the use of a personalized machine learning approach more inevitable to use to gain more accuracy for each person. However, for some feature such as average values of pose and eye landmarks, we can distinguish more separation (less overlaps) between the values of each emotion representations (refer to the figure 16 illustrating feature importance obtained in RF classification).

In order to gain a comprehensive and nuanced understanding of the intricate relationships between the various features and the target variable within our dataset, we employed an insightful and effective visualization technique known as a Pairplot. This visual exploration tool enabled us to delve into the pairwise interactions and dependencies among all the features, unveiling valuable insights into their collective behavior and how they influence the target variable. By creating scatterplots for every possible combination of features and plotting the target variable against each feature, the Pairplot facilitated the identification of intricate patterns, correlations, and trends that might have otherwise gone unnoticed. This comprehensive analysis provided us with a holistic view of the dataset's distribution and structure, empowering us to uncover hidden connections and make informed decisions for subsequent analysis and modeling. The Pairplot visualization unveiled complex data relationships, enabling deeper insights and more accurate modeling.

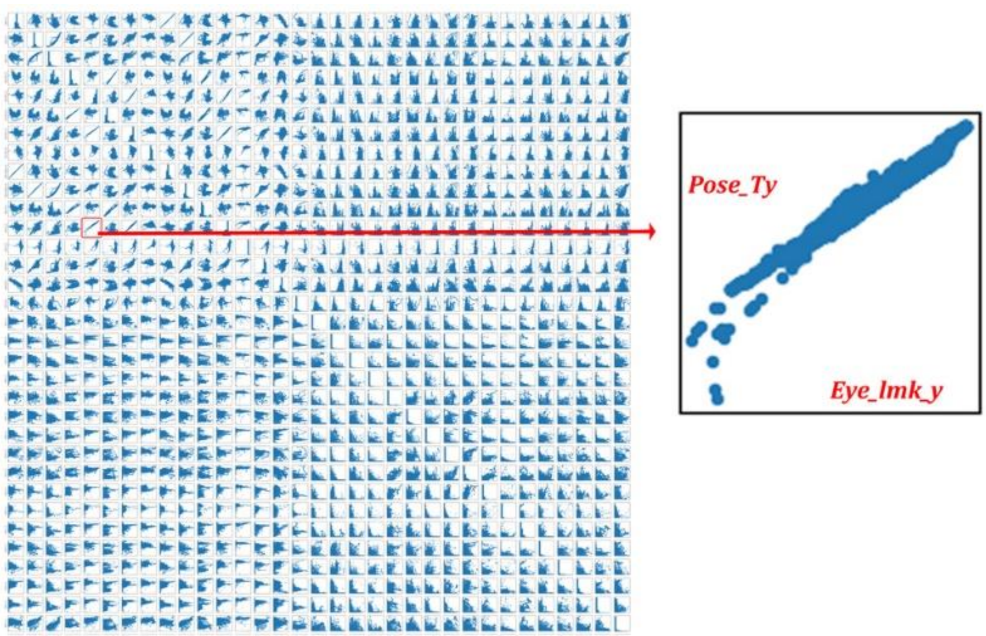

Figure 11. Features pair plot (each feature against each feature and also the label against each feature)

Based on the Pairplot shown in Figure 11, it is evident that there are noticeable linear and nonlinear associations between various pairs of variables, such as the average y-axis coordination values of pose and eye landmarks. In the following sections, we will delve deeper into the findings and implications derived from the Pairplot analysis, highlighting significant feature interactions and their impact on the classification of emotions.

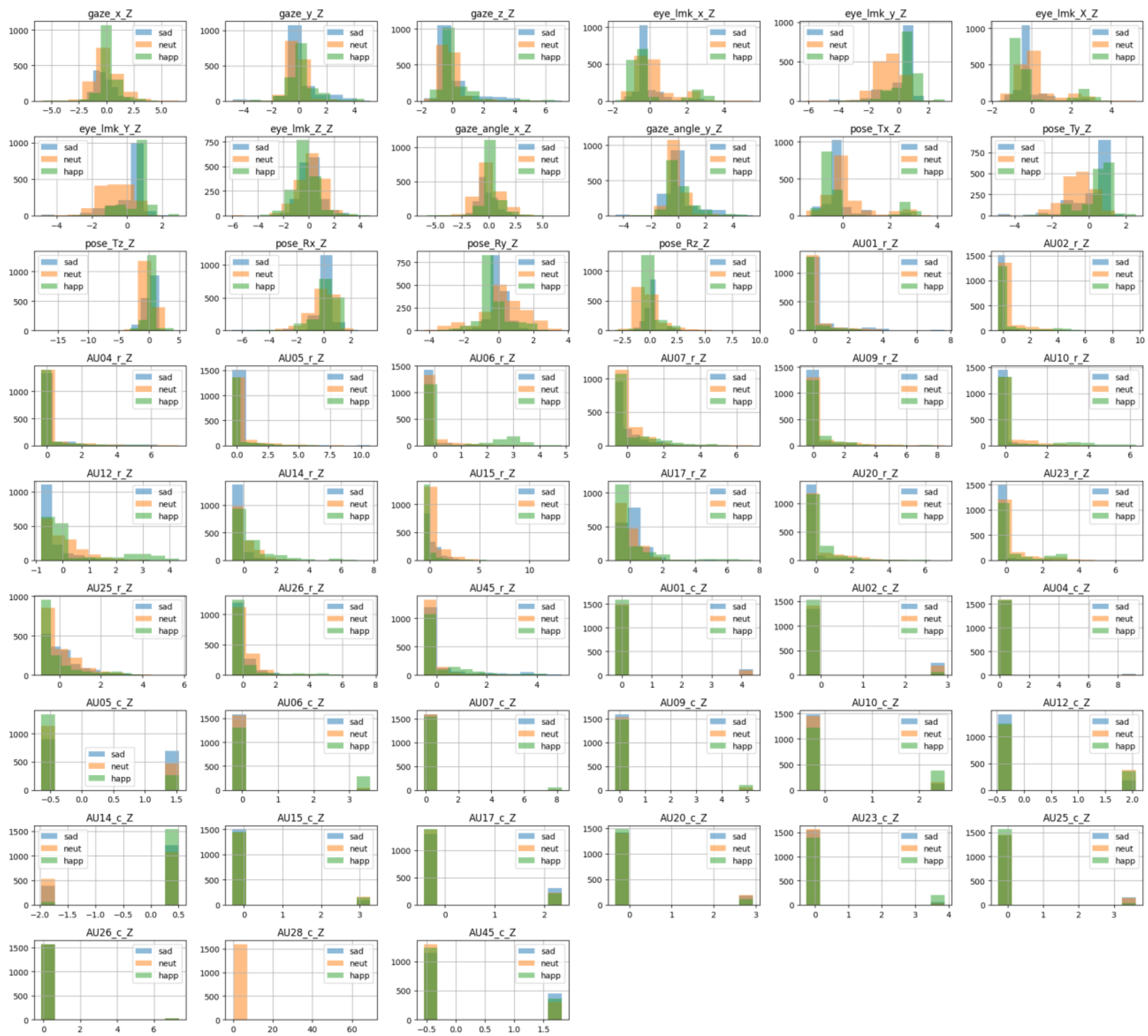

Figure 12. histograms of each emotion classification in each feature

Performing Principal Component Analysis (PCA):

Principal component analysis (PCA) is a widely used method for examining extensive datasets that consist of numerous dimensions or features per observation. Figure 13 is a representation of PCA performed on the dataset used in this thesis for further analysis. It aids in enhancing the comprehensibility of data by retaining the most crucial information while reducing its dimensionality. By employing PCA, researchers can effectively visualize complex multidimensional data, such as those encountered in the field of emotion recognition. In the context of our previous discussion on emotion recognition, PCA could be employed to explore patterns and relationships among various emotional features, potentially revealing valuable insights into the nature and representation of emotions in the dataset.

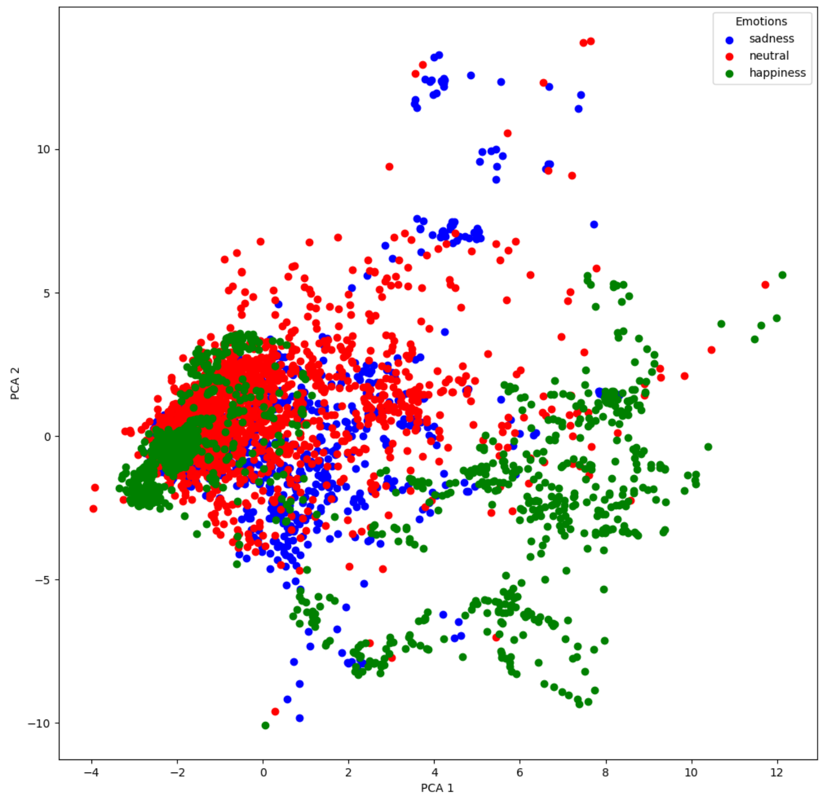

Figure 13. Principal Component Analysis representing the variance between the data by two PCA components.

3.5.3 Model Selection

In this study, we employed three different machine learning models: K-Nearest Neighbors (KNN), Random Forest, and Multilayer Perceptron (MLP) for the task of emotion classification. Each of these models offers unique advantages that make them suitable for this particular classification task.

By utilizing these three models, we aim to explore and compare their performance in classifying emotions. Each model has its own strengths and capabilities that can contribute to the accurate classification of emotions in our dataset. The subsequent sections of this study will delve into the implementation, evaluation, and analysis of these models, shedding light on their effectiveness in capturing and classifying emotions.

3.5.3.1 KNN Classifier

The K-Nearest Neighbors (KNN) algorithm is a straightforward yet effective method for classifying new instances by comparing their similarities to labeled instances in the training set. By assigning a new instance to the class that is most common among its closest neighbors, KNN can capture intricate nonlinear relationships present in the data. This property makes KNN a valuable option for emotion classification tasks, as emotions frequently display complex and nonlinear patterns. Figure 14 illustrates the performance accuracy values for different values of k in the range of k_selection (1 to 30).

In our endeavor to optimize the performance of the K-Nearest Neighbors (KNN) algorithm for emotion classification, we employed a thorough grid search methodology. This involved systematically testing a range of k values and utilizing repeated k-fold cross-validation to assess their impact on classification accuracy. Through this rigorous evaluation, we discovered that setting k to 1 yielded outstanding results, achieving an impressive classification accuracy of 93% (0.928) for our emotion classification task. This finding underscores the effectiveness of the KNN algorithm in accurately categorizing emotions within our dataset.

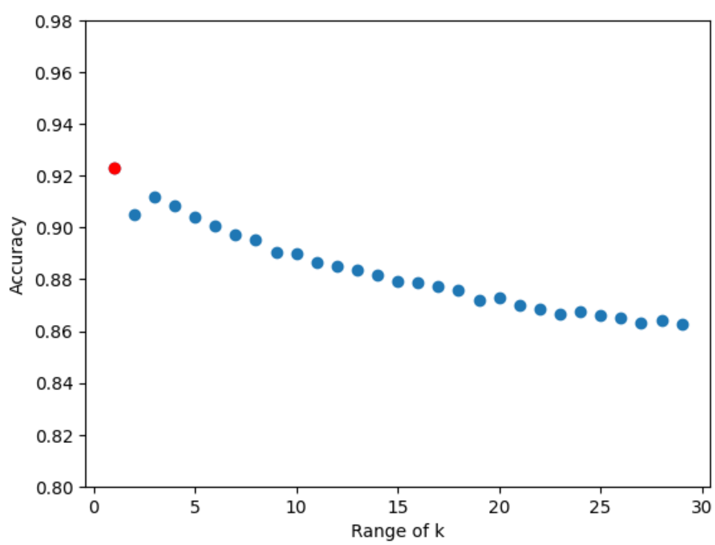

Figure 14. Performance accuracy of the model in the range of k between 0 to 30.

3.5.3.2 Random Forest Classifier

The Random Forest algorithm is a type of ensemble learning that leverages multiple decision trees to make accurate predictions. It generates a diverse set of decision trees by using random subsets of the data and features. The final prediction is made by aggregating the predictions of individual trees. Random Forest offers several advantages, such as handling high-dimensional data effectively, handling both categorical and numerical features, and providing a measure of feature importance. These characteristics make Random Forest a suitable choice for emotion classification tasks, where the dataset may consist of a large number of features and involve a combination of different data types. The implementation involves utilizing Random Forest classifier for emotion classification. A grid search is performed using repeated k-fold cross-validation to determine the optimal hyperparameters for the Random Forest algorithm. The hyperparameters being tuned are 'max_depth' and 'max_features'. Various values are tested for each hyperparameter: 'max_depth' is tested with values of 5, 10, 15, 20, and None, while 'max_features' is tested with 'sqrt', 'log2', 0.25, 0.5, 0.75, and None. The grid search selects the best combination of hyperparameters, resulting in a 'max_depth' of 20 and 'max_features' set to 'log2'. With this configuration, the model achieves an impressive accuracy of approximately 95% (0.946) for emotion classification. This demonstrates the effectiveness of the Random Forest algorithm in capturing the intricate relationships between features and emotions. Additionally, Figure 15 illustrates the performance accuracy across the range of tested 'max_depth' values, providing visual evidence of the impact of varying this hyperparameter on the classification accuracy.

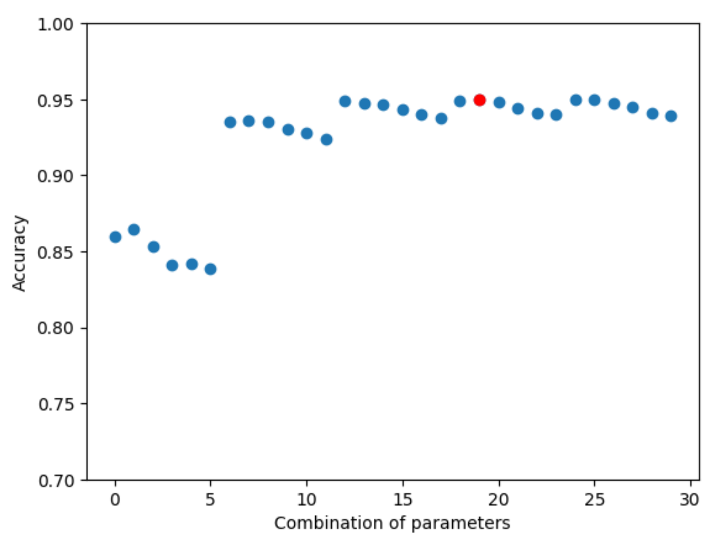

Figure 15. Performance accuracy of the Random Forest model

In addition to the model training, we conducted an analysis of feature importance using the Random Forest algorithm with the best parameters. This analysis allows us to gain insights into the relevance and contribution of each feature in our dataset towards the classification of emotions. By examining the feature importance plot, we can identify the key features that significantly influence the prediction of emotions. Figure 16 showcases the prominent importance of eye landmarks and pose as influential features in our analysis, underscoring their significant role in accurately interpreting the data.

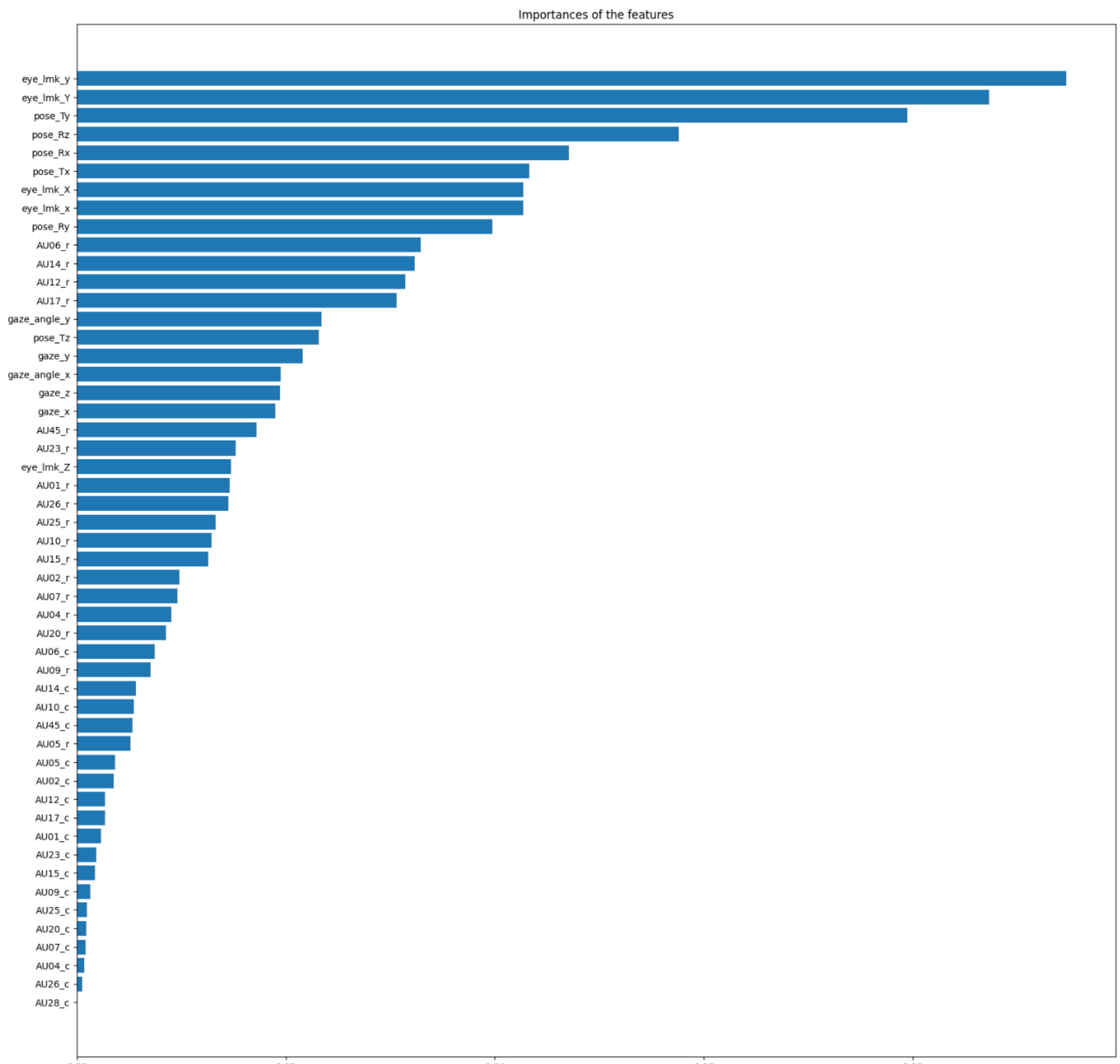

Figure 16. feature importance plot using random forest

This information is valuable for understanding the underlying factors and characteristics that drive the classification of emotions. Through this analysis, we aim to uncover the most influential features, providing valuable insights into the relationships between these features and the corresponding emotional states. The subsequent sections of this study will further discuss the findings and implications derived from the feature importance analysis, shedding light on the crucial features that contribute to accurate emotion classification.

3.5.3.3 MLP Classifier

Multilayer Perceptron (MLP) is a type of artificial neural network with multiple layers of interconnected nodes, or neurons. MLP is known for its ability to learn complex patterns and relationships in data. It is particularly effective in capturing nonlinear dependencies and performing well on classification tasks with high-dimensional data. MLP can be trained using backpropagation, which adjusts the weights of the connections between neurons to minimize the prediction error. For emotion classification, MLP offers the advantage of being able to capture intricate and nuanced patterns in the data, potentially leading to accurate and fine-grained emotion predictions.

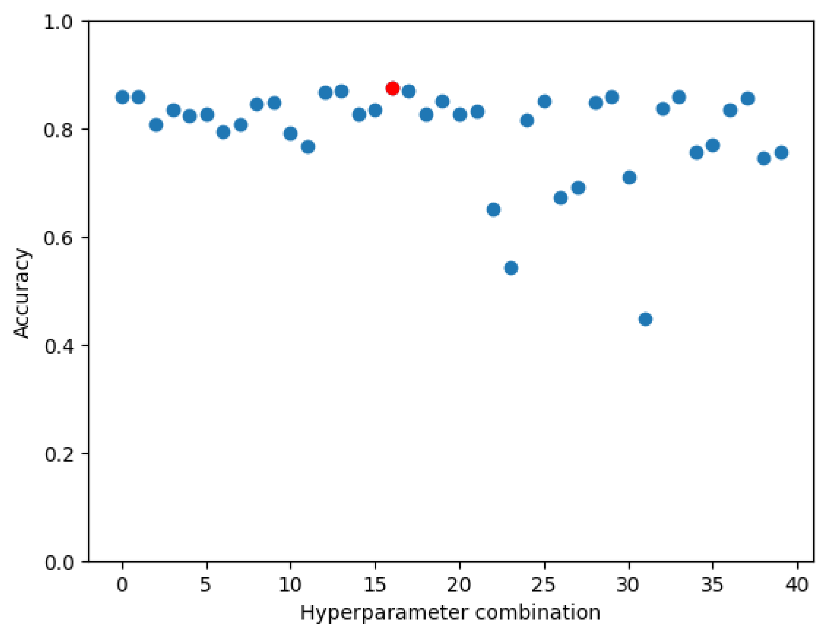

Figure 17. Performance accuracy of MLP in different range of hyperparameters.

The MLP classifier is configured with specific parameters, including 'max_iter' set to 10000 and 'early_stopping' enabled for efficient training.

A grid search is performed using repeated k-fold cross-validation to find the optimal combination of hyperparameters for the MLP classifier. The hyperparameters being tuned are 'hidden_layer_sizes', 'activation', 'solver', and 'validation_fraction'. The 'hidden_layer_sizes' parameter is tested with values ranging from 3 to 8, 'activation' is tested with 'relu' and 'logistic', 'solver' is tested with 'adam' and 'sgd', and 'validation_fraction' is tested with 0.1 and 0.5.

The grid search selects the best combination of hyperparameters, resulting in an optimal configuration of 'activation' as 'relu', 'hidden_layer_sizes' as (7,), 'solver' as 'adam', and 'validation_fraction' as 0.1. This combination yields the best performance of 87.3% for emotion classification using the MLP classifier.

3.5.4 Performance Estimation

In order to estimate the performance of our models, we employed a robust evaluation procedure. The code snippet provided demonstrates the process applied for each model, with specific variations in the hyperparameters considered for each one.

```
mlp_clfr = MLPClassifier(max_iter=10000, early_stopping=True,
random_state=20)

hid_layers_s = [(i,) for i in range(3,8)]
val_frac = [0.1, 0.5]
opt_select = ['adam', 'sgd']
activ_f = ['relu', 'logistic']

parameter_space = {
    'hidden_layer_sizes': hid_layers_s,
    'activation': activ_f,
    'solver': opt_select,
    'validation_fraction': val_frac
}

cons = RepeatedKFold(n_splits=5, n_repeats=3, random_state=5)

# search and select the best hyperparameter combs
em_gs = GridSearchCV(mlp_clfr, parameter_space, cv=cons)
em_gs.fit(X, y)
print('Combination of best hyperparameters:')
em_gs.best_params_
```

The first step involves setting up the necessary components, including the range of hyperparameters to explore and the cross-validation strategies. For the K-Nearest Neighbors (KNN) classifier, we iterated over different values of the number of neighbors (k) using a grid search approach. The outer loop of 10-fold cross-validation splits the data into training and testing sets, while the inner loop of 5-fold repeated cross-validation is used to search for the best k value within each training set.

Following that, the grid search technique is applied to the training data, allowing us to train the KNN classifier using different combinations of hyperparameters and assessing their performance. Through this process, we aim to identify the set of hyperparameters that yield the best model performance. Subsequently, the optimized model is utilized to make predictions on the test set, and the predicted labels are combined with the true labels to evaluate the overall effectiveness of the model in capturing the underlying patterns in the data.

Additionally, retaining the best combination of hyperparameters for each iteration is a critical aspect as it enables us to pinpoint the optimal settings and gain deeper insights into the behavior of the model. Analyzing the performance of the KNN model using the provided confusion matrix offers valuable insights and allows us to interpret the results with a comprehensive understanding of its classification capabilities.

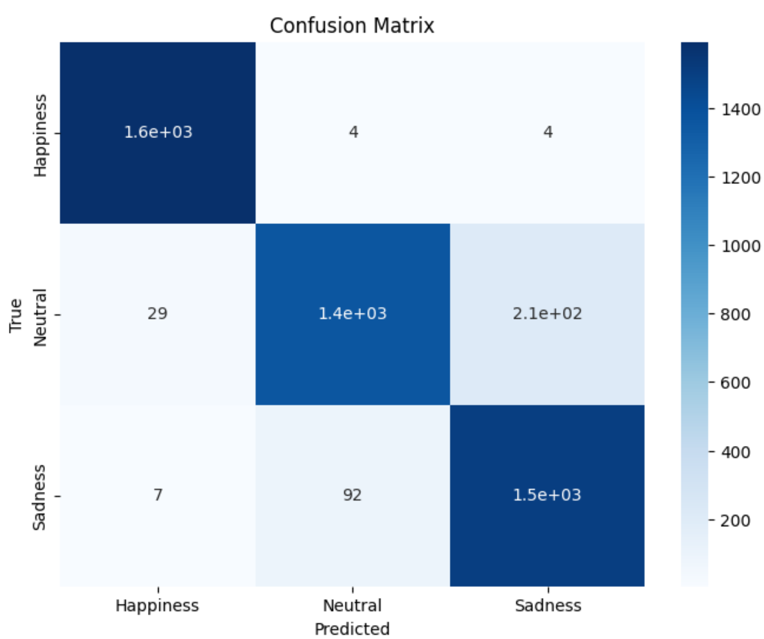

Figure 18. Confusion matrix for KNN classifier

- Happiness: The classifier correctly predicted 1592 instances as happiness. However, it misclassified 4 instance as neutral, 4 instances as sadness.

- Neutral: The classifier correctly predicted 1360 instances as neutral. However, it misclassified 211 instances as sadness, and 29 instances as happiness.
- Sadness: The classifier correctly predicted 1501 instances as sadness. However, it misclassified 92 instances as neutral and 7 instances as happiness.

The overall accuracy of the KNN classifier is calculated to be 0.946, indicating a high level of accuracy in classifying emotions. The accuracy represents the proportion of correctly classified instances out of the total number of instances.

It's important to note that these results are specific to the given dataset and may vary when applied to different datasets or with different classification algorithms.

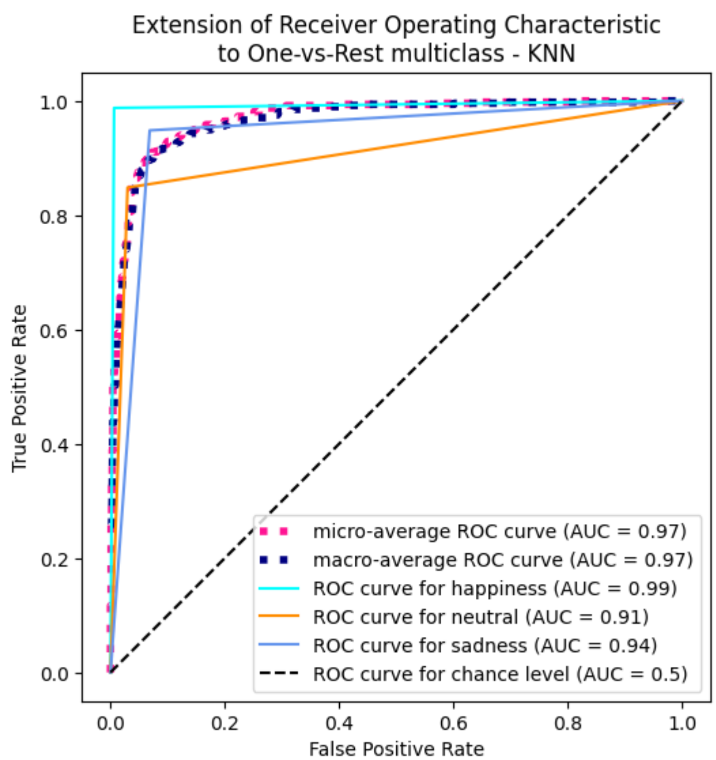

Figure 19. Emotion Recognition ROC Curves and AUCs for KNN Classification

The ROC curve (Figure 19) visualizes the trade-off between the true positive rate (sensitivity) and the false positive rate as the classification threshold is varied. By plotting the ROC curve for each emotion class, we can assess how well the models distinguishes between positive instances (correctly classified emotions) and negative instances (incorrectly classified emotions) for different levels of sensitivity and specificity.

The AUC value represents the area under the ROC curve and provides a single scalar measure of the model's performance. AUC ranges from 0 to 1, with higher values indicating better discrimination capability. In this case, the AUC values for each emotion class are as follows:

- For the "happiness" class, the AUC is 0.99, indicating high accuracy in identifying happiness.
- For the "Neutral" class, the AUC is 0.91, suggesting good performance in distinguishing neutral emotions.
- For the "sadness" class, the AUC is 0.94, indicating satisfactory discrimination ability for surprise.

Additionally, the micro-average ROC curve is calculated by aggregating the true positive rate and false positive rate across all classes. The micro-average AUC is 0.95, representing the overall performance of the KNN model in emotion recognition and classification. The macro-average ROC is the average of individual ROC curves calculated for each class separately. It gives equal weight to each class, regardless of class imbalance, providing a balanced evaluation of the classifier's performance across all classes. The macro-average AUC is 0.95 as well for this model. Overall, these results indicate that the KNN model shows strong discriminatory power for differentiating between various emotions, achieving high AUC values for most classes and a strong overall performance.

The procedure outlined above is also applied to the Random Forest and Multilayer Perceptron (MLP) models, with slight modifications in the hyperparameters tested. For the Random Forest model, we explore different values of the maximum depth and maximum features to determine the best configuration. On the other hand, for the MLP model, we consider a range of hidden layer sizes, activation functions, optimization algorithms, and validation fractions.

By executing this evaluation procedure for each model, we can assess their respective performances and identify the best hyperparameters for achieving optimal results. The code measures the execution time to provide insights into the computational requirements of the process. The resulted confusion matrix from Random Forest model performance estimation, represents the performance of the Random Forest classifier for emotion classification. It displays the number of instances predicted for each emotion category (rows) compared to their true labels (columns).

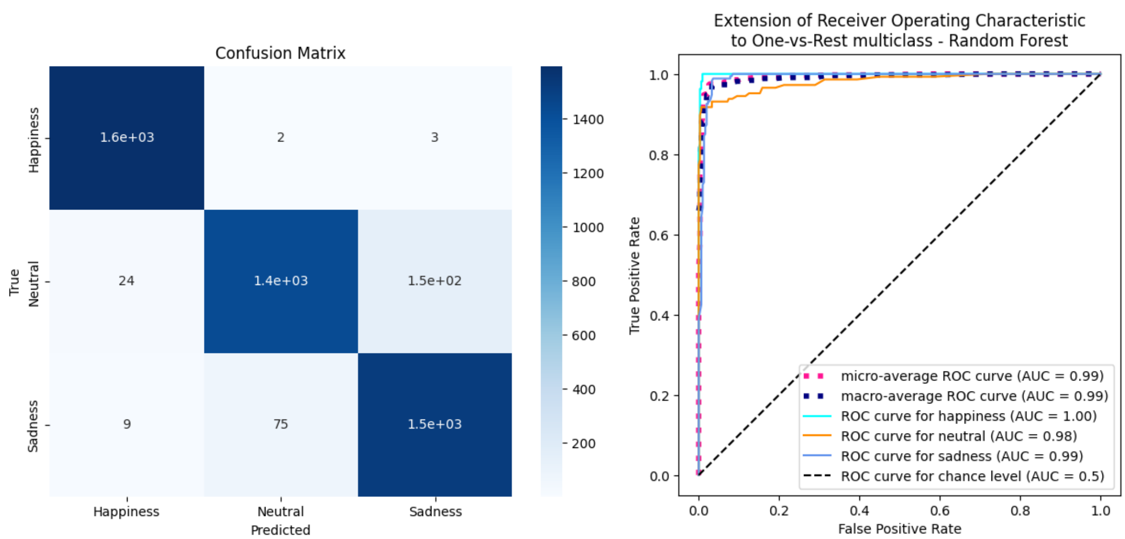

Figure 20. Confusion matrix, and ROC Curves and AUCs for Random Forest Classification

- Happiness: The classifier correctly predicted 1595 instances as happiness. It misclassified 2 instance as neutral and 3 instances as sadness.
- Neutral: The classifier correctly predicted 1430 instances as neutral. It misclassified 24 instances as happiness, 146 instances as sadness.
- Sadness: The classifier correctly predicted 1516 instances as sadness. It misclassified 75 instances as neutral, 9 instances as happiness.

The overall accuracy of the Random Forest classifier is calculated to be 0.946, indicating a high level of accuracy in classifying emotions. The accuracy represents the proportion of correctly classified instances out of the total number of instances.

Furthermore, the combinations of best parameters for the Random Forest model are specified as (max_depth, max_features) along with the number of times each combination was selected as the optimal choice during the evaluation process. The provided counts indicate that the combination (None, 'sqrt') was selected 2 times, (None, 'log2') was selected 4 times, and (20, 'sqrt'), and (20, 'log2') were selected 2 times each. These results demonstrate the effectiveness of the Random Forest model in accurately classifying emotions, and the different parameter combinations explored to optimize its performance. The provided Figure 21 are the confusion matrix and performance estimation (ROC, AUC) of the MLP (Multi-Layer Perceptron) classifier for emotion classification.

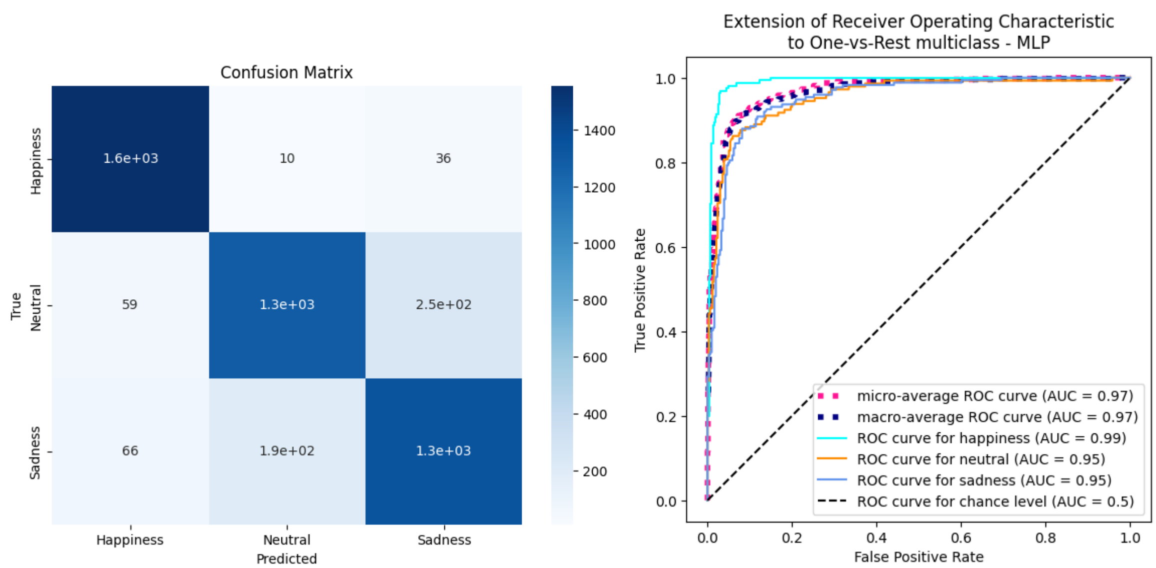

Figure 21. Confusion matrix, and ROC Curves and AUCs for MLP Classification

- Happiness: The classifier correctly predicted 1554 instances as happiness. It misclassified 10 instances as neutral, 36 instances as sadness.
- Neutral: The classifier correctly predicted 1294 instances as neutral. It misclassified 18 247 instances as sadness, and 59 instances as happiness.
- Sadness: The classifier correctly predicted 1341 instances as sadness. It misclassified 66 instances as happiness, 193 instances as neutral.

The overall accuracy of the MLP classifier is calculated to be 0.873, indicating a relatively high level of accuracy in classifying emotions. The accuracy represents the proportion of correctly classified instances out of the total number of instances.

Additionally, the combinations of best parameters for the MLP model are specified as (activation function, hidden_layer_sizes, solver-optimizer, validation_fraction) along with the number of times each combination was selected as the optimal choice during the evaluation process. The provided counts indicate that the combination ('relu', (7,), 'adam', 0.1) was selected 6 times and ('logistic', (7,), 'adam', 0.5) was selected 4 times.

These results demonstrate the performance of the MLP classifier in accurately classifying emotions, and the different parameter combinations explored to optimize its performance.

The personalized machine learning approach, employing KNN, RF, and MLP models, showcased promising performance in the task of emotion recognition compared to previous research that utilized general-purpose models. The personalized approach demonstrated superior results by leveraging the specific characteristics and nuances of the dataset to enhance classification accuracy. Through the utilization of personalized models, tailored to the task at hand, the classification performance was significantly improved, surpassing the performance of general-purpose models typically used in previous studies. This emphasizes the effectiveness of the personalized approach in capturing the unique patterns and features associated with emotions, leading to enhanced performance and more accurate emotion recognition. Table 7 provides a comprehensive comparison of the performance results between these models and the personalized approach adopted for this thesis.

Table 7. Overall performance accuracy of the personalized models using in this thesis and previous general models on the same task (Emotion Recognition)

Author	Data type	Dataset	Accuracy
Kolestra [139]	EEG + PPS	DEAP	61.5%
Tang [140]	EEG + PPS	DEAP	83.5%
Yin [141]	EEG + PPS	DEAP	83.5%
Nguyen [144]	Speech + Facial Images	eNTER FACE’05	90.85%
Zhang [143]	Speech + Facial Images	eNTER FACE’05	85.97%
Thesis - KNN	Pose/Facial Landmarks	Emognition	92.80%
Thesis - RF	Pose/Facial Landmarks	Emognition	94.60%
Thesis - MLP	Pose/Facial Landmarks	Emognition	87.30%

The overall results from the provided table indicate the performance of various approaches in different datasets for emotion recognition. Kolestra achieved an accuracy of 61.5% using EEG and PPS data from the DEAP dataset. Tang and Yin achieved higher accuracies of 83.5% using the same dataset and data types. Nguyen achieved a notable accuracy of 90.85% using speech and facial images from the eNTERFACE'05 dataset, while Zhang achieved an accuracy of 85.97% using the same dataset and data types. The results from the thesis work using Pose/Facial Landmarks data in the Emognition dataset demonstrate high accuracies with KNN achieving

92.80%, RF achieving 94.60%, and MLP achieving 87.30%. These results showcase the effectiveness of mentioned approaches in capturing and recognizing emotions in datasets, highlighting the potential of personalized machine learning models in achieving high accuracy in emotion recognition tasks.

We conducted the experiment using different dataset sizes, including a large dataset for the reported results. When working with smaller datasets, we observed a performance decrease of up to 5 percent for each model. Additionally, it is important to investigate the possibility of hyperparameter tuning leaking information into the validation set, which warrants further examination in future studies. Furthermore, it is crucial to test the optimized model on future datasets to further validate its effectiveness.

4 Conclusion

Previous studies have established the effectiveness of mobile applications focused on mental health, highlighting their ability to engage users even without incorporating gamification features. We anticipate that the STAND app will further amplify user engagement. To achieve this, it would be valuable to explore the integration of personalized and visually appealing design elements. In our research methodology, we introduced an additional step by analyzing video frames capturing facial expressions. This approach enabled us to examine and compare the performance of three popular machine learning models for personalized emotion classification. Such analytical techniques can also prove beneficial in stress and anxiety recognition. The integration of personalized machine learning methods within the realm of digital health presents promising opportunities for the advancement and refinement of models related to mental health diagnosis, management, and treatment.

5 Discussion and Future Works

The primary objective of our research is to empirically demonstrate the efficacy of advanced technologies, specifically ecological momentary assessment and personalized machine learning, in classifying stress and anxiety levels among young adults. By incorporating these technologies into digital mental health interventions, we aim to offer an alternative to conventional care practices. While our study employs innovative methods to quantify anxiety and explore potential correlations between facial cues and stress, there are additional technologies that warrant investigation and integration in future endeavors to enhance recognition mechanisms.

One such technology is crowdsourcing, which involves human intervention for evaluating and validating subjective information, leading to potential improvements. Advancements in this domain include algorithms for efficient data management, handling complexity, and debugging and verifying AI models.

Another avenue for enhancing the effectiveness of mental health interventions lies in leveraging modern smartphones and their built-in sensors. Given the widespread usage and integration of smartphones into people's daily lives, they represent a cutting-edge tool in the field of digital health. Various aspects, such as physical activities tracked by acceleration sensors, location data derived from Wi-Fi and GPS information, and social interactions and activities, can serve as indicators of an individual's mental health state. These aspects can be passively monitored and recorded using portable devices like smartphones.

Our plan involves developing the versatile STAND app, which can be tailored for research purposes and collecting computer vision data on facial expressions to build personalized machine learning models. These models can support digital health interventions focused on mental states and developmental disorders. This work can be integrated with existing research on automatic emotion recognition across diverse contexts, including mental illness diagnosis, understanding human social and physiological interactions, and developing interactive systems such as sociable robotics.

For example, accurate classification and response to human emotions, such as stress and anxiety, play a crucial role in identifying certain types of developmental disorders like Autism Spectrum Disorder (ASD). Children with ASD often exhibit emotions differently and face challenges in displaying appropriate facial expressions. By creating an optimal model for human-computer interactions, we anticipate enhancing social communication. Initial efforts have explored the use of digital and wearable devices to facilitate therapy sessions within the home environment. The proposed personalized affective AI can be integrated into these interventions, offering improved digital and mobile therapies tailored to the specific needs of children with ASD and other developmental delays.

References

[1] S. Chan, J. Torous, L. Hinton, and P. Yellowlees, "Towards a Framework for Evaluating Mobile Mental Health Apps," *Telemed J E Health,* vol. 21, no. 12, pp. 1038-41, Dec 2015, doi: 10.1089/tmj.2015.0002.

[2] C. Hollis *et al.*, "Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review," *J Child Psychol Psychiatry,* vol. 58, no. 4, pp. 474-503, Apr 2017, doi: 10.1111/jcpp.12663.

[3] E. S. Epel *et al.*, "Accelerated telomere shortening in response to life stress," *Proc Natl Acad Sci U S A,* vol. 101, no. 49, pp. 17312-5, Dec 7 2004, doi: 10.1073/pnas.0407162101.

[4] R. S. Lazarus, *Psychological stress and the coping process* (McGraw;Hill series in psychology). New York: McGraw-Hill (in English ; Undetermined), 1966.

[5] J. Tomaka, J. Blascovich, R. M. Kelsey, and C. L. Leitten, "Subjective, physiological, and behavioral effects of threat and challenge appraisal," *Journal of personality and social psychology,* vol. 65, no. 2, p. 248, 1993.

[6] L. Welty, A. Harrison, K. Abram, N. Olson, D. Aaby, and K. McCoy, "Substance abuse and mental health services administration.(2017). Key substance use and mental health indicators in the United States: results from the 2016 national survey on drug use and health (HHS Publication No. SMA 17-5044, NSDUH Series H-52). Rockville, MD: Center for Behavioral Health Statistics and Quality," *Substance Abuse and Mental Health Services Administration. Retrieved. College of Health Sciences,* vol. 106, no. 5, p. 128, 2019.

[7] D. Chisholm *et al.*, "Scaling-up treatment of depression and anxiety: a global return on investment analysis," *The Lancet Psychiatry,* vol. 3, no. 5, pp. 415-424, 2016.

[8] S. Cho, I. Ensari, C. Weng, M. G. Kahn, and K. Natarajan, "Factors Affecting the Quality of Person-Generated Wearable Device Data and Associated Challenges: Rapid Systematic Review," *JMIR Mhealth Uhealth,* vol. 9, no. 3, p. e20738, Mar 19 2021, doi: 10.2196/20738.

[9] T. L. C. A. Health, "Growing up in a digital world: benefits and risks," vol. 2, ed, 2018, p. 79.

[10] M. Aitken, B. Clancy, and D. Nass, "The growing value of digital health-evidence and impact on human health and the healthcare system. USA: IQVIA Institute for Human Data Science 2017," ed, 2021.

[11] R. Mojtabai *et al.*, "Barriers to mental health treatment: results from the National Comorbidity Survey Replication," *Psychological medicine,* vol. 41, no. 8, pp. 1751-1761, 2011.

[12] D. D. Ebert *et al.*, "Internet-and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health," *European Psychologist,* 2018.

[13] I. Husain and D. Spence, "Can healthy people benefit from health apps?," *BmJ,* vol. 350, 2015.

[14] S. Greene, H. Thapliyal, and A. Caban-Holt, "A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health," *IEEE Consumer Electronics Magazine,* vol. 5, no. 4, pp. 44-56, 2016.

[15] A. Halaris, "A primary care focus on the diagnosis and treatment of major depressive disorder in adults," *Journal of Psychiatric Practice®,* vol. 17, no. 5, pp. 340-350, 2011.

[16] R. V. Shah *et al.*, "Personalized machine learning of depressed mood using wearables," *Translational psychiatry,* vol. 11, no. 1, pp. 1-18, 2021.
[17] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work?--a literature review of empirical studies on gamification," in *2014 47th Hawaii international conference on system sciences*, 2014: Ieee, pp. 3025-3034.
[18] I. T. Dykens, A. Wetzel, S. L. Dorton, and E. Batchelor, "Towards a unified model of gamification and motivation," in *International Conference on Human-Computer Interaction*, 2021: Springer, pp. 53-70.
[19] A. Hoffmann, C. A. Christmann, and G. Bleser, "Gamification in stress management apps: a critical app review," *JMIR serious games,* vol. 5, no. 2, p. e7216, 2017.
[20] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Computer methods and programs in biomedicine,* vol. 108, no. 3, pp. 1287-1301, 2012.
[21] N. Sharma and T. Gedeon, "Modeling observer stress for typical real environments," *Expert Systems with Applications,* vol. 41, no. 5, pp. 2231-2238, 2014.
[22] D. F. Dinges *et al.*, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviation, space, and environmental medicine,* vol. 76, no. 6, pp. B172-B182, 2005.
[23] L. Hoffman, *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge, 2015.
[24] M. Csikszentmihalyi, *Handbook of research methods for studying daily life*. Guilford Press, 2011.
[25] S. Shiffman, A. Stone, and M. Hufford, "Ecological momentary assessment. Annu. Rev. Clin. Psychol," *Ecological Momentary Assessment. Annu Rev Clin Psychol,* vol. 4, no. 1, 2008.
[26] M. Balsamo, F. Cataldi, L. Carlucci, C. Padulo, and B. Fairfield, "Assessment of late-life depression via self-report measures: a review," *Clinical interventions in aging,* vol. 13, p. 2021, 2018.
[27] W. H. Organization. Mental Health of Older Adults [Online] Available: https://www.who.int/en/news-room/fact-sheets/detail/mental-health-of-older-adults
[28] S. T. Doherty, C. J. Lemieux, and C. Canally, "Tracking human activity and well-being in natural environments using wearable sensors and experience sampling," *Social science & medicine,* vol. 106, pp. 83-92, 2014.
[29] E. Andreou *et al.*, "Perceived stress scale: reliability and validity study in Greece," *International journal of environmental research and public health,* vol. 8, no. 8, pp. 3287-3298, 2011.
[30] G. J. Meyer *et al.*, "Psychological testing and psychological assessment: A review of evidence and issues," *American psychologist,* vol. 56, no. 2, p. 128, 2001.
[31] L. H. Rogler, R. G. Malgady, and W. W. Tryon, "Evaluation of mental health: Issues of memory in the Diagnostic Interview Schedule," *The Journal of nervous and mental disease,* vol. 180, no. 4, pp. 215-222, 1992.
[32] T. F. Van de Mortel, "Faking it: social desirability response bias in self-report research," *Australian Journal of Advanced Nursing, The,* vol. 25, no. 4, pp. 40-48, 2008.
[33] O. P. John and R. W. Robins, "Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism," *Journal of personality and social psychology,* vol. 66, no. 1, p. 206, 1994.

[34] E. Russell, G. Koren, M. Rieder, and S. Van Uum, "Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions," *Psychoneuroendocrinology,* vol. 37, no. 5, pp. 589-601, 2012.

[35] D. H. Hellhammer, S. Wüst, and B. M. Kudielka, "Salivary cortisol as a biomarker in stress research," *Psychoneuroendocrinology,* vol. 34, no. 2, pp. 163-171, 2009.

[36] G. Giannakakis, D. Grigoriadis, and M. Tsiknakis, "Detection of stress/anxiety state from EEG features during video watching," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015: IEEE, pp. 6034-6037.

[37] R. S. Lewis, N. Y. Weekes, and T. H. Wang, "The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health," *Biological psychology,* vol. 75, no. 3, pp. 239-247, 2007.

[38] K. Nidal and A. S. Malik, *EEG/ERP analysis: methods and applications*. Crc Press, 2014.

[39] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on information technology in biomedicine,* vol. 14, no. 2, pp. 410-417, 2009.

[40] A. de Santos Sierra, C. S. Ávila, J. G. Casanova, and G. B. Del Pozo, "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Transactions on Industrial Electronics,* vol. 58, no. 10, pp. 4857-4865, 2011.

[41] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[42] Y. Masaoka and I. Homma, "The effect of anticipatory anxiety on breathing and metabolism in humans," *Respiration physiology,* vol. 128, no. 2, pp. 171-177, 2001.

[43] K. B. Nilsen, T. Sand, L. J. Stovner, R. B. Leistad, and R. H. Westgaard, "Autonomic and muscular responses and recovery to one-hour laboratory mental stress in healthy subjects," *BMC Musculoskeletal Disorders,* vol. 8, no. 1, pp. 1-12, 2007.

[44] I. R. Murray, C. Baber, and A. South, "Towards a definition and working model of stress and its effects on speech," *Speech Communication,* vol. 20, no. 1-2, pp. 3-12, 1996.

[45] P. Wittels, B. Johannes, R. Enne, K. Kirsch, and H.-C. Gunga, "Voice monitoring to measure emotional load during short-term stress," *European journal of applied physiology,* vol. 87, no. 3, pp. 278-282, 2002.

[46] C. Darwin and P. Prodger, *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.

[47] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, "A decision theoretic model for stress recognition and user assistance," in *AAAI*, 2005, vol. 2005, pp. 529-534.

[48] D. Metaxas, S. Venkataraman, and C. Vogler, "Image-based stress recognition using a model-based dynamic face tracking system," in *International Conference on Computational Science*, 2004: Springer, pp. 813-821.

[49] U. Hadar, T. J. Steiner, E. C. Grant, and F. Clifford Rose, "Head movement correlates of juncture and stress at sentence level," *Language and speech,* vol. 26, no. 2, pp. 117-129, 1983.

[50] A. Adams, M. Mahmoud, T. Baltrušaitis, and P. Robinson, "Decoupling facial expressions and head motions in complex emotions," in *2015 International conference on affective computing and intelligent interaction (ACII)*, 2015: IEEE, pp. 274-280.

[51] S. Leal and A. Vrij, "Blinking during and after lying," *Journal of Nonverbal Behavior,* vol. 32, no. 4, pp. 187-194, 2008.

[52] J. Mackintosh, R. Kumar, and T. Kitamura, "Blink rate in psychiatric illness," *The British Journal of Psychiatry,* vol. 143, no. 1, pp. 55-57, 1983.

[53] R. C. Chan and E. Y. Chen, "Blink rate does matter: a study of blink rate, sustained attention, and neurological signs in schizophrenia," *The Journal of nervous and mental disease,* vol. 192, no. 11, pp. 781-783, 2004.

[54] G. W. Ousler III, K. W. Hagberg, M. Schindelar, D. Welch, and M. B. Abelson, "The ocular protection index," *Cornea,* vol. 27, no. 5, pp. 509-513, 2008.

[55] J. A. Taylor, "The relationship of anxiety to the conditioned eyelid response," *Journal of Experimental psychology,* vol. 41, no. 2, p. 81, 1951.

[56] M. Honma, "Hyper-volume of eye-contact perception and social anxiety traits," *Consciousness and cognition,* vol. 22, no. 1, pp. 167-173, 2013.

[57] H. Simpson and F. Molloy, "Effects of audience anxiety on pupil size," *Psychophysiology,* vol. 8, no. 4, pp. 491-496, 1971.

[58] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive physiological monitoring for automated stress detection in human-computer interaction," in *International Workshop on Human-Computer Interaction*, 2007: Springer, pp. 29-38.

[59] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *International journal of human-computer studies,* vol. 59, no. 1-2, pp. 185-198, 2003.

[60] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science,* vol. 349, no. 6245, pp. 255-260, 2015.

[61] D. D. Miller and E. W. Brown, "Artificial intelligence in medical practice: the question to the answer?," *The American journal of medicine,* vol. 131, no. 2, pp. 129-133, 2018.

[62] H. Suominen, L. Zhou, L. Hanlen, and G. Ferraro, "Benchmarking clinical speech recognition and information extraction: new data, methods, and evaluations," *JMIR medical informatics,* vol. 3, no. 2, p. e4321, 2015.

[63] D. I. R. Boyle and N. Rafael, "BioGrid Australia and GRHANITE™: Privacy-protecing subject matching," in *Health Informatics: The Transformative Power of Innovation*: IOS Press, 2011, pp. 24-34.

[64] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care—addressing ethical challenges," *The New England journal of medicine,* vol. 378, no. 11, p. 981, 2018.

[65] D. Graff and C. Cieri, "English gigaword, linguistic data consortium," ed, 2003.

[66] C. Wang, H. Zhao, and H. Zhang, "Chinese college students have higher anxiety in new semester of online learning during COVID-19: a machine learning approach," *Frontiers in psychology,* vol. 11, p. 587413, 2020.

[67] T. Richter, B. Fishbain, A. Markus, G. Richter-Levin, and H. Okon-Singer, "Using machine learning-based analysis for behavioral differentiation between anxiety and depression," *Scientific reports,* vol. 10, no. 1, pp. 1-12, 2020.

[68] A. Priya, S. Garg, and N. P. Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," *Procedia Computer Science,* vol. 167, pp. 1258-1267, 2020.

[69] S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, "Predicting women with depressive symptoms postpartum with machine learning methods," *Scientific reports,* vol. 11, no. 1, pp. 1-15, 2021.

[70] D. McDuff, R. Kaliouby, T. Senechal, M. Amr, J. Cohn, and R. Picard, "Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions

collected," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2013, pp. 881-888.

[71] Q. Tariq, J. Daniels, J. N. Schwartz, P. Washington, H. Kalantarian, and D. P. Wall, "Mobile detection of autism through machine learning on home video: A development and prospective validation study," *PLoS medicine,* vol. 15, no. 11, p. e1002705, 2018.

[72] M. Sultana, M. Al-Jefri, and J. Lee, "Using machine learning and smartphone and smartwatch data to detect emotional states and transitions: exploratory study," *JMIR mHealth and uHealth,* vol. 8, no. 9, p. e17818, 2020.

[73] A. Hart, D. Reis, E. Prestele, and N. C. Jacobson, "Using Smartphone Sensor Paradata and Personalized Machine Learning Models to Infer Participants' Well-being: Ecological Momentary Assessment," *Journal of Medical Internet Research,* vol. 24, no. 4, p. e34015, 2022.

[74] G. M. Harari, S. R. Müller, M. S. Aung, and P. J. Rentfrow, "Smartphone sensing methods for studying behavior in everyday life," *Current opinion in behavioral sciences,* vol. 18, pp. 83-90, 2017.

[75] R. Mafrur, I. Nugraha, and D. Choi, "Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose," *Human-centric Computing and Information Sciences,* vol. 5, no. 1, pp. 1-18, 2015.

[76] L. Pei *et al.*, "Human behavior cognition using smartphone sensors," *Sensors,* vol. 13, no. 2, pp. 1402-1424, 2013.

[77] J. Asselbergs, J. Ruwaard, M. Ejdys, N. Schrader, M. Sijbrandij, and H. Riper, "Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study," *Journal of medical Internet research,* vol. 18, no. 3, p. e5505, 2016.

[78] A. E. Whitton *et al.*, "Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress," *JMIR mental health,* vol. 2, no. 1, p. e3573, 2015.

[79] C. E. Branson, P. Clemmey, and P. Mukherjee, "Text message reminders to improve outpatient therapy attendance among adolescents: a pilot study," *Psychological services,* vol. 10, no. 3, p. 298, 2013.

[80] G. Clarke *et al.*, "Overcoming Depression on the Internet (ODIN)(2): a randomized trial of a self-help depression skills program with reminders," *Journal of medical Internet research,* vol. 7, no. 2, p. e412, 2005.

[81] S. M. Kelders, R. N. Kok, H. C. Ossebaard, and J. E. Van Gemert-Pijnen, "Persuasive system design does matter: a systematic review of adherence to web-based interventions," *Journal of medical Internet research,* vol. 14, no. 6, p. e2104, 2012.

[82] R. Neff and J. Fry, "Periodic prompts and reminders in health promotion and health behavior interventions: systematic review," *Journal of medical Internet research,* vol. 11, no. 2, p. e1138, 2009.

[83] L. Maenhout, C. Peuters, G. Cardon, G. Crombez, A. DeSmet, and S. Compernolle, "Nonusage attrition of adolescents in an mHealth promotion intervention and the role of socioeconomic status: secondary analysis of a 2-arm cluster-controlled trial," *JMIR mHealth and uHealth,* vol. 10, no. 5, p. e36404, 2022.

[84] D. King, F. Greaves, C. Exeter, and A. Darzi, "'Gamification': Influencing health behaviours with games," vol. 106, ed: SAGE Publications Sage UK: London, England, 2013, pp. 76-78.

[85] C. Lister, J. H. West, B. Cannon, T. Sax, and D. Brodegard, "Just a fad? Gamification in health and fitness apps," *JMIR serious games,* vol. 2, no. 2, p. e3413, 2014.

[86] M. Schmidt-Kraepelin, P. A. Toussaint, S. Thiebes, J. Hamari, and A. Sunyaev, "Archetypes of gamification: analysis of mHealth apps," *JMIR mHealth and uHealth,* vol. 8, no. 10, p. e19280, 2020.

[87] F. F.-H. Nah, Q. Zeng, V. R. Telaprolu, A. P. Ayyappa, and B. Eschenbrenner, "Gamification of education: a review of literature," in *International conference on hci in business*, 2014: Springer, pp. 401-409.

[88] C. Dichev and D. Dicheva, "Gamifying education: what is known, what is believed and what remains uncertain: a critical review," *International journal of educational technology in higher education,* vol. 14, no. 1, pp. 1-36, 2017.

[89] H. Kalantarian, P. Washington, J. Schwartz, J. Daniels, N. Haber, and D. Wall, "A gamified mobile system for crowdsourcing video for autism research," in *2018 IEEE international conference on healthcare informatics (ICHI)*, 2018: IEEE, pp. 350-352.

[90] H. Kalantarian, K. Jedoui, P. Washington, and D. P. Wall, "A mobile game for automatic emotion-labeling of images," *IEEE transactions on games,* vol. 12, no. 2, pp. 213-218, 2018.

[91] R. Farzan, J. M. DiMicco, D. R. Millen, C. Dugan, W. Geyer, and E. A. Brownholtz, "Results from deploying a participation incentive mechanism within the enterprise," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2008, pp. 563-572.

[92] C. Cheong, F. Cheong, and J. Filippou, "Quick quiz: A gamified approach for enhancing learning," 2013.

[93] W. LI, T. GROSSMAN, and G. FITZMAURICE, "Gamified tutorial system for first time AutoCAD users," in *UIST*, 2012, vol. 12, pp. 7-10.

[94] T. D.-L. Guin, R. Baker, J. Mechling, and E. Ruyle, "Myths and realities of respondent engagement in online surveys," *International Journal of Market Research,* vol. 54, no. 5, pp. 613-633, 2012.

[95] A. L. Smith and L. Baker, "Getting a clue: creating student detectives and dragon slayers in your library," *Reference Services Review,* 2011.

[96] J. Lin, D. Ebert, D. Lehr, M. Berking, and H. Baumeister, "Internet based cognitive behavioral interventions: state of the art and implementation possibilities in rehabilitation," *Die Rehabilitation,* vol. 52, no. 3, pp. 155-163, 2013.

[97] M. Cunningham, V. Wuthrich, R. Rapee, H. Lyneham, C. Schniering, and J. Hudson, "The Cool Teens CD-ROM for anxiety disorders in adolescents," *European child & adolescent psychiatry,* vol. 18, no. 2, pp. 125-129, 2009.

[98] A. L. Calear and H. Christensen, "Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents," *Medical Journal of Australia,* vol. 192, pp. S12-S14, 2010.

[99] A. Kargarandehkordi and P. Washington, "Computer Vision Estimation of Stress and Anxiety Using a Gamified Mobile-based Ecological Momentary Assessment and Deep Learning: Research Protocol," *medRxiv,* p. 2023.04. 28.23289168, 2023.

[100] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication," *Archives of general psychiatry,* vol. 62, no. 6, pp. 593-602, 2005.

[101] A. J. Bravo, M. C. Villarosa-Hurlocker, and M. R. Pearson, "College student mental health: An evaluation of the DSM–5 self-rated Level 1 cross-cutting symptom measure," *Psychological Assessment,* vol. 30, no. 10, p. 1382, 2018.
[102] R. P. Gallagher, "National survey of college counseling centers 2014," 2015.
[103] A. Gulliver, K. M. Griffiths, and H. Christensen, "Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review," *BMC psychiatry,* vol. 10, no. 1, pp. 1-9, 2010.
[104] S. Saganowski *et al.*, "Emognition dataset: emotion recognition with self-reports, facial expressions, and physiology using wearables," *Scientific data,* vol. 9, no. 1, p. 158, 2022.
[105] Y. Qian *et al.*, "Computer Vision Estimation of Emotion Reaction Intensity in the Wild," *arXiv preprint arXiv:2303.10741,* 2023.
[106] M. Clynes, *Sentics: The touch of emotions*. Anchor Press, 1977.
[107] A. Heraz and M. Clynes, "Recognition of emotions conveyed by touch through force-sensitive screens: observational study of humans and machine learning techniques," *JMIR Mental Health,* vol. 5, no. 3, p. e10104, 2018.
[108] P. Kendall, "THE HEART OF THE MACHINE: Our Future in a World of Artificial Intelligence," *Perspectives on Science and Christian Faith,* vol. 70, no. 2, pp. 141-144, 2018.
[109] S. Dutta, B. K. Mishra, A. Mitra, and A. Chakraborty, "An analysis of emotion recognition based on GSR signal," *ECS Transactions,* vol. 107, no. 1, p. 12535, 2022.
[110] S. Chen, X. Li, Q. Jin, S. Zhang, and Y. Qin, "Video emotion recognition in the wild based on fusion of multimodal features," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 494-500.
[111] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research: a scoping review," *Translational Psychiatry,* vol. 10, no. 1, p. 116, 2020.
[112] J. Wagner, D. Schiller, A. Seiderer, and E. André, "Deep learning in paralinguistic recognition tasks: Are hand-crafted features still relevant?," 2018.
[113] A. Y. Kim, E. H. Jang, S.-H. Lee, K.-Y. Choi, J. G. Park, and H.-C. Shin, "Automatic Depression Detection Using Smartphone-Based Text-Dependent Speech Signals: Deep Convolutional Neural Network Approach," *Journal of Medical Internet Research,* vol. 25, p. e34474, 2023.
[114] A. Ripoli *et al.*, "Personalized machine learning approach to predict candidemia in medical wards," *Infection,* vol. 48, pp. 749-759, 2020.
[115] A.-W. de Leeuw, S. van der Zwaard, R. van Baar, and A. Knobbe, "Personalized machine learning approach to injury monitoring in elite volleyball players," *European journal of sport science,* vol. 22, no. 4, pp. 511-520, 2022.
[116] T. D. Lalitharatne *et al.*, "Facial expression rendering in medical training simulators: Current status and future directions," *IEEE Access,* vol. 8, pp. 215874-215891, 2020.
[117] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition,* vol. 77, pp. 354-377, 2018.
[118] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE transactions on affective computing,* vol. 13, no. 3, pp. 1195-1215, 2020.
[119] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 28, no. 12, pp. 2037-2041, 2006.

[120] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," *Multimedia Tools and Applications,* vol. 76, pp. 7803-7821, 2017.

[121] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and vision Computing,* vol. 27, no. 6, pp. 803-816, 2009.

[122] F. An and Z. Liu, "Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM," *The visual computer,* vol. 36, no. 3, pp. 483-498, 2020.

[123] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2011: IEEE, pp. 884-888.

[124] M. Satiyan, M. Hariharan, and R. Nagarajan, "Recognition of facial expression using Haar wavelet transform," *Journal of Electrical and Electronics Systems Research (JEESR),* vol. 3, pp. 89-96, 2010.

[125] H. Soyel and H. Demirel, "Improved SIFT matching for pose robust facial expression recognition," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2011: IEEE, pp. 585-590.

[126] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science,* vol. 313, no. 5786, pp. 504-507, 2006.

[127] A. Banerjee, O. C. Mutlu, A. Kline, S. Surabhi, P. Washington, and D. P. Wall, "Training and profiling a pediatric facial expression classifier for children on mobile devices: machine learning study," *JMIR formative research,* vol. 7, p. e39917, 2023.

[128] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1805-1812.

[129] A. T. Lopes, E. De Aguiar, and T. Oliveira-Santos, "A facial expression recognition system using convolutional networks," in *2015 28th SIBGRAPI conference on graphics, patterns and images*, 2015: IEEE, pp. 273-280.

[130] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[131] D. Hamester, P. Barros, and S. Wermter, "Face expression recognition with a 2-channel convolutional neural network," in *2015 international joint conference on neural networks (IJCNN)*, 2015: IEEE, pp. 1-8.

[132] S. Xie and H. Hu, "Facial expression recognition with FRR-CNN," *Electronics Letters,* vol. 53, no. 4, pp. 235-237, 2017.

[133] J. Donahue *et al.*, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625-2634.

[134] F. Zhang, Y. Yu, Q. Mao, J. Gou, and Y. Zhan, "Pose-robust feature learning for facial expression recognition," *Frontiers of Computer Science,* vol. 10, pp. 832-844, 2016.

[135] T. Zhang, "Facial expression recognition based on deep learning: a survey," in *Advances in Intelligent Systems and Interactive Applications: Proceedings of the 2nd International Conference on Intelligent and Interactive Systems and Applications (IISA2017)*, 2018: Springer, pp. 345-352.

[136] K. Zhang, Y. Huang, Y. Du, and L. Wang, "Facial expression recognition based on deep evolutional spatial-temporal networks," *IEEE Transactions on Image Processing,* vol. 26, no. 9, pp. 4193-4203, 2017.
[137] X. Zhao *et al.*, "Peak-piloted deep network for facial expression recognition," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, 2016: Springer, pp. 425-442.
[138] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "Facewarehouse: A 3d facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics,* vol. 20, no. 3, pp. 413-425, 2013.
[139] S. Koelstra *et al.*, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing,* vol. 3, no. 1, pp. 18-31, 2011.
[140] H. Tang, W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using deep neural networks," in *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part IV 24*, 2017: Springer, pp. 811-819.
[141] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Computer methods and programs in biomedicine,* vol. 140, pp. 93-110, 2017.
[142] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The eNTERFACE'05 audio-visual emotion database," in *22nd international conference on data engineering workshops (ICDEW'06)*, 2006: IEEE, pp. 8-8.
[143] S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian, "Learning affective features with a hybrid deep model for audio–visual emotion recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 28, no. 10, pp. 3030-3043, 2017.
[144] D. Nguyen, K. Nguyen, S. Sridharan, D. Dean, and C. Fookes, "Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition," *Computer vision and image understanding,* vol. 174, pp. 33-42, 2018.
[145] K. Han *et al.*, "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence,* vol. 45, no. 1, pp. 87-110, 2022.
[146] J. Hewig, D. Hagemann, J. Seifert, M. Gollwitzer, E. Naumann, and D. Bartussek, "A revised film set for the induction of basic emotions," *Cognition and emotion,* vol. 19, no. 7, p. 1095, 2005.
[147] L. D. Kaczmarek, M. Behnke, J. Enko, M. Kosakowski, P. Guzik, and B. M. Hughes, "Splitting the affective atom: Divergence of valence and approach-avoidance motivation during a dynamic emotional experience," *Current Psychology,* vol. 40, pp. 3272-3283, 2021.
[148] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cognition & emotion,* vol. 9, no. 1, pp. 87-108, 1995.
[149] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition and emotion,* vol. 24, no. 7, pp. 1153-1172, 2010.
[150] E. Reynaud, M. El-Khoury-Malhame, O. Blin, and S. Khalfa, "Voluntary emotion suppression modifies psychophysiological responses to films," *Journal of Psychophysiology,* 2012.

[151] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, 2018: IEEE, pp. 59-66.

Appendices

Ethical approval of the application (IRB)

UNIVERSITY of HAWAI'I® SYSTEM

Office of Research Compliance
Human Studies Program

DATE:	December 20, 2022
TO:	Washington, Peter, PhD, Information and Computer Sciences, University of Hawaii at Manoa
FROM:	Rivera, Victoria, Dir, Ofc of Rsch Compliance, Social & Behavioral
PROTOCOL TITLE:	Logging of Mental Health Status using a Computer Vision Smartphone Application
FUNDING SOURCE:	None
PROTOCOL NUMBER:	2022-00937
APPROVAL DATE:	December 20, 2022

NOTICE OF APPROVAL FOR HUMAN RESEARCH

Under an expedited review procedure, the research project identified above was approved on December 20, 2022 by the University of Hawaii Institutional Review Board (UH IRB). The application qualified for expedited review under 45 CFR 46.110 and 21 CFR 56.110, Category 6, 7. Per 45 CFR 46.109, **a Continuing Review is not required, however you may be requested to submit a progress report.**

This memorandum is your record of the IRB approval of this study. Please maintain it with your study records.

The Human Studies Program approval must be maintained for the entire term of your project. Please see guidance at Final Revisions to the Common Rule on the regulatory requirements for ongoing review and/or monitoring of research approved under an expedited review category.

If, during the course of your project, you intend to make changes to this study, you must obtain approval from the Human Studies Program prior to implementing any changes. You can submit your proposed changes via the UH eProtocol application. If an Unanticipated Problem occurs during the course of the study, you must notify the Human Studies Program within 24 hours of knowledge of the problem. A formal report must be submitted to the Human Studies Program within 10 days.The definition of "Unanticipated Problem" may be found at: HSP Policies & Guidance Quicklink . The report form may be submitted via the eProtocol application.

You are required to maintain complete records pertaining to the use of humans as participants in your research. This includes all information or materials conveyed to and received from participants as well as signed consent forms, data, analyses, and results. These records must be maintained for at least three years following project completion or termination, and they are subject to inspection and review by the Human Studies Program and other authorized agencies.

Study Closure: Please notify this office when your project is complete. Upon notification, we will close our files pertaining to your project. Please contact this office if you have any questions or require assistance. We appreciate your cooperation, and wish you success with your research.

2425 Campus Road, Sinclair 10
Honolulu, HI 96822
Telephone: (808) 956-5007 • Fax: (808) 956-9150
An Equal Opportunity/Affirmative Action Institution

Association for the Accreditation of Human Research Protection Programs, Inc. Full Accreditation

Consent Forms

Consent Form

We are asking you to participate in a research study titled **"A Gamified Mobile-Based EMA on Stress and Anxiety Detection Using Personalized ML and Computer Vision"**. We will describe this study to you and answer any of your questions. This study is being led by Peter Yiğitcan Washington, Information and Computer Sciences at the University of Hawaiʻi at Mānoa and Ali Kargarandehkordi, Information and Communication Technology Department at the University of Turku, Finland.

What the study is about

The purpose of this research is to screen and detect the stress/anxiety level of participants through a gamified mobile-based EMA (Ecological Momentary Assessment) using a personalized machine learning technique and computer vision. Stress and, particularly, anxiety are among subjective matters that require methodological and practical experts to be analyzed. Therefore, measuring such states through objective means can be quite demanding, and most of the time, not accurate. Generally, the data required for detecting such states is collected by providing various questionnaires designed by psychological experts to the users, which, in itself could be quite a boring procedure and in most cases, this monotonous method has a negative impact on the final result. Expanding usage of AI and machine learning techniques in recent research, throw light on complexities and mechanisms behind detecting these subjective experiences.

What we will ask you to do

For an approximate period of 4 to 6 weeks, we will ask you to fill in a very short questionnaire - one question per notification, five times a day. Alongside each question, you will be asked to take/upload a selfie photo. The collected daily photos are going to be analyzed in the second phase for potential facial expressions relevant to stress and anxiety e.g., eyes' pupil size variation. We do not expect doing the mentioned tasks to take more than 5 minutes a day. By participating in the research, you can win a gift card in a drawing held at the end of the data collection phase. If more funding is received, gift cards will be awarded to all participants.

Risks and discomforts

We do not anticipate any risks from participating in this research.

Benefits

As this research only focuses on the detection of anxiety, there will not be any sort of treatment involved in the research. There are no direct benefits (e.g., from therapeutic or intervention research) to the participants, but information from this study may benefit people indirectly now or in the future. This research is going to provide useful data for other researchers to conduct similar methods in their research, which will result in detecting subjective health issues more efficiently while raising social awareness about such mental concerns.

Compensation for participation

By participating in the research, you can win a gift card in a drawing held at the end of the data collection phase. Gift cards will be a $20 supermarket voucher (tax-free) for grocery shopping. Winners will be contacted through email on how to receive the gift card. If more funding is received, gift cards will be awarded to all participants.

Audio/Video Recording
As you will be asked to take selfie photographs within the questionnaire for the analysis of possible facial expressions and activities beyond research analysis (publications, presentations), be aware of the following important notations:

- We ask you to grant us the right to make, use and publish the data and images in whole or in part in text formats and media forms now known (such as film, slides, and digital audio) or developed in the future. The dataset will be stored on cloud storage provided by Amazon Web Services (AWS) with extra security measures and might be published online for future use in the analysis of potential facial expressions in other AI studies and research. This includes the right to edit or duplicate any images/recordings;
- You do not have rights to inspect or approve the finished product or printed/published matter that uses the images/recordings or versions of the images/recordings
- You will not receive any financial compensation for commercial and/or non-commercial (as appropriate) uses of the images/recordings.

Please sign below if you agree to the statements above. You may still participate in this study if you are not willing to have the interview recorded.

☐ I do not agree with any of the statements above.
☐ I agree with all the conditions mentioned above.

Signed: ______________________________
Date: ______________________________

Privacy/Confidentiality/Data Security
All personal information will be kept confidential to the extent allowed by law. Several public agencies with responsibility for research oversight, including the UH Human Studies Program, have authority to review research records.

The research involves no more than minimal risk to participants; We anticipate that your participation in this survey presents no greater risk than everyday use of the Internet. We ensure you that your identities will not be disclosed to anyone outside of the research team. The data collected during the data collection phase will be stored on cloud storage provided by Amazon Web Services (AWS) with extra security measures. As a safe action multi-factor authentication (MFA) will be activated on the AWS account root user and any users with interactive access to AWS Identity and Access Management (IAM).

Please note that email communication is neither private nor secure. Though we are taking precautions to protect your privacy, you should be aware that information sent through email could be read by a third party.

Sharing De-identified Data Collected in this Research
De-identified data from this study may be shared with the research community at large to advance science and health. We will remove or code any personal information that could identify you before files are shared with other researchers to ensure that, by current scientific standards and known methods, no one will be able to identify you from the information we share.

Future use of Identifiable Data Collected in this Research
Future research projects may use your photos, user data, and identifiable data without requesting additional consent. Please know any future research will only publish deidentified data to the public.

Taking part is voluntary
Your involvement is voluntary, you may refuse to participate before the study begins, discontinue at any time, or skip any questions/procedures that may make you feel uncomfortable, with no penalty, and no effect on the compensation earned before withdrawing, or your academic standing, record, or relationship with the university or other organization or service that may be involved with the research.

Follow up studies
We may contact you again to request your participation in a follow-up study. As always, your participation will be voluntary and we will ask for your explicit consent to participate in any of the follow-up studies.

May we contact you again to request your participation in a follow-up study? Yes☐ No☐

If you have questions

The main researcher conducting this study is Ali Kargarandehkordi, a graduate student at the University of Turku. You may contact him by email: alkarg@utu.fi or Phone: +358414986790 (Also available on WhatsApp or Telegram if you are an international participant e.g., living in Hawai'i). For questions about your rights as a research participant, contact the University of Hawaii Human Studies Program by phone at 808.956.5007 or by email at uhirb@hawaii.edu.

Mental Health Resources and Guidelines:

You can always have access to mental health resources provided by the University of Hawai'i provided below. In addition, The Adult Mental Health Division Crisis Line of Hawaii provides a team of trained and experienced professionals to help individuals in times of a mental health crisis. The Crisis Line of Hawaii helps you 24 hours a day, 7 days a week. This information is also accessible within the application, implemented on the home screen tab.

National Crisis Helpline for Suicide Prevention:

- http://manoa.hawaii.edu/c3od2a/crisis-helplines/
- Crisis Helpline on Oahu: call at 832-3100
- Crisis Helpline on the Neighbor Islands: Call toll-free at 1-800-753-6879
- The National Suicide Prevention Lifeline can be reached at 1-800-273-8255 (TALK). Veterans Crisis Line is 1-800-273-8255 (Press 1)

University of Hawai'i Mental Health Resources:

- https://research.hawaii.edu/orc/wp-content/uploads/sites/7/2021/12/mental-health-and-suicide-crisis.pdf

You can also have access to the resources provided by the State of Hawai'i, Department of Health at https://health.hawaii.gov/emsipsb/injury-prevention/suicide-prevention/

All participants will be given a copy of this form.

Statement of Consent

I have read the above information, and have received answers to any questions I asked. I consent to take part in the study with the title of "A Gamified Mobile Based EMA on Stress and Anxiety Detection Using Personalized ML and Computer Vision".

Your Signature____ Date_____________

Your Name (printed)____

Signature of person obtaining consent___________________________________ Date________

Printed name of person obtaining consent_____________________________________

This consent form will be kept by the researcher for five years beyond the end of the study.

Recruitment Ad. (Flyer)

A Gamified Mobile-Based EMA on Stress and Anxiety Detection Using Personalized ML and Computer Vision

Our research team would like to welcome you to participate in our research to test the accuracy of a machine-learning predictive model to detect stress/anxiety levels through gamification. Our team is working at the University of Hawai'i at Mānoa (Digital Health Lab - PI: Dr. Peter Washington), and the University of Turku (Information and Communication Technology Department).

We aim to screen and detect the stress/anxiety level of target users through a gamified smartphone-based EMA (Ecological Momentary Assessment) app, using a personalized machine learning technique and computer vision. An application has been developed to send daily notifications to the users, containing a short questionnaire asking about anxiety-related symptoms. A selfie-taking capability has also been implemented within the questionnaires to collect daily photographs of our participants. Selfies are being used in order to analyze the correlation of facial parameters with the perceived amount of stress/anxiety, particularly investigating eyes' closure and pupil size variation which is a promising discriminative indicator of such mental states. The analysis is mainly focused on a limited number of voluntary students in which specific daily health conditions are going to be tracked down, particularly ones relevant to stress and anxiety.

In order to collect the required data, we would like to recruit individuals (students) who have experienced a certain level of stress and anxiety during the last few years. The research period is to be four weeks. Before recruitment, you might be asked for a possible online interview with one of our research team members to evaluate if you fit perfectly into the research requirements. The interview will last between 30-60 minutes. At the end of the data collection phase, you can win a supermarket gift card (tax-free) for grocery shopping in a fair drawing for their positive participation in the research. The drawing will be held right after the data collection phase, and winners will be contacted through an email containing instructions on how to receive the gift card.

Criteria for participation:

- *Students aged 28 ± 10*
- *Having access to a smartphone (iOS, or Android)*
- *Students who have experienced a certain level of stress and anxiety during the last few years*

*If you fulfill the above criteria and are interested, please contact **Ali Kargarandehkordi** through the contact details below. Upon receiving your email, SMS, or call, we will set up a follow-up online meeting to evaluate your participation and ensure that you are a good fit for the research. In case you were eligible to participate in the research, you will receive an email containing the installation link for our application in addition to the required instructions on how to install and use the application properly. For calls inside Finland, there will not be any extra charge. For international applicants, including students living in Hawai'i interested in participating in the research, WhatsApp or Telegram calls are also possible through the same number.*

alkarg@utu.fi

+358414986790

UH IRB Approval of the Flyer: December 20, 2022