EMOTION RECOGNITION USING PHYSIOLOGICAL SIGNALS

Piyusha Khode

A dissertation submitted to
The School of Computing Sciences of the University of East Anglia
in partial fulfilment of the requirements for the degree of
MASTER OF SCIENCE.
AUGUST, 2024

SUPERVISOR(S), MARKERS/CHECKER AND ORGANISER

The undersigned hereby certify that the markers have independently marked the dissertation entitled "**Emotion Recognition using Physiological Signals**" by **Piyusha Khode**, and the external examiner has checked the marking, in accordance with the marking criteria and the requirements for the degree of **Master of Science**.

Supervisor: ________________________
Dr Muhammad Awais

Markers: ________________________
Marker 1: Dr Muhammad Awais

____
Marker 2: Prof Beatriz De La Iglesia

External Examiner: ________________________
Checker/Moderator

Moderator: ________________________
Dr. Wenjia Wang

DISSERTATION INFORMATION AND STATEMENT

Dissertation Submission Date: **August, 2024**

Student:	**Piyusha Khode**
Title:	**Emotion Recognition using Physiological Signals**
School:	**Computing Sciences**
Course:	**Data Science**
Degree:	**MSc.**
Duration:	**2023–2024**
Organiser:	**Dr. Wenjia Wang**

STATEMENT:

Unless otherwise noted or referenced in the text, the work described in this dissertation is, to the best of my knowledge and belief, my own work. It has not been submitted, either in whole or in part for any degree at this or any other academic or professional institution.

Subject to confidentiality restriction if stated, permission is herewith granted to the University of East Anglia to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Signature of Student

Abstract

Recognising human emotions using physiological signals is becoming important in applications such as mental health monitoring and human-computer interaction. This dissertation describes a system that uses Blood Volume Pulse (BVP), Electrodermal Activity (EDA), and Heart Rate (HR) to accurately classify emotional states.

Beginning with an overview of the significance and objectives of emotion recognition, the dissertation looks into relevant research on physiological signal processing and machine learning. The literature review examines emotion theory, feature extraction, and algorithm selection, developing a solid foundation for the system's design.

The methodology section describes the technical aspects, such as data preprocessing, feature extraction and the use of machine learning models. It also discusses issues such as data imbalance and the methods used to address them. A flowchart depicts the system's design, while the tools and libraries used are discussed.

The system is assessed using a variety of models such as Random Forest, XGBoost, MLP and ANN. The results show that incorporating HR data improves emotion recognition accuracy, thus validating the chosen approaches.

Finally, the dissertation highlights the system's contribution to emotion recognition, particularly its possible impact on mental health monitoring and adaptive technologies. The study also identifies future directions, such as using more advanced models to address data imbalances and expanding the system's application.

Acknowledgements

I would like to thank my supervisor, Dr. Muhammad Awais, for his invaluable suggestions and constant support throughout this project. His guidance and encouragement was important to the successful completion of this dissertation.

I'm also extremely grateful to my family and friends for their constant encouragement and belief in me throughout this journey. This accomplishment would not have been possible without them.

Piyusha Khode

Norwich, UK.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background & Motivation

Human emotions are difficult, complex conditions that have an major impact on our lives and our physical and mental health. Positive emotions have been associated with better productivity and health, but negative emotions can cause a range of health problems, with long-term exposure even having the potential to worse disorders like depression. Du et al. (2023)Emotions comes on naturally, unlike moods, which are long lasting mental states. They are frequently accompanied by bodily shifts that can be seen in the skin, muscles, brain, and heart, among other body systems. Facial expressions, vocal inflections and other physiological reactions are examples of these physiological manifestations. The accurate and on time detection of human emotions is a difficult area in scientific study due to the complex relationship between our mental and physical processes in emotional experiences. Despite considerable efforts from many of the researchers, understanding of emotions remains limited, which defines the need for continued investigation in this field. It is possible to improve mental health measures, increase emotional intelligence and create more advanced emotion-recognition systems by expanding understanding of human emotions.Shu et al. (2018)

In today's increasingly technologically advanced and inter connected world, the ability to accurately identify and comprehend human emotions has become critical. This increasing relevance goes beyond theoretical or academic pursuits to include practical, real-world applications that have the potential to have a big influence on the lives of individuals and the way society operates. As a result, in an effort to increase the reliability and effectiveness of emotional recognition systems in wide range of circumstances scientists and business specialists are constantly developing and refining methods to precisely and consistently recognise emotions.Khare et al. (2024)There are many relevant area of research with several applications in various fields. Its ability to advance emotional computing, better healthcare outcomes especially mental health monitoring, enable more organic human-robot interactions, E-learning in the education system after Covid-19 Guo et al. (2024) and offer valuable insights for market research is what makes it so important. Also, Emotion recognition has been applied in other areas as well such as safe driving, social security and more. These applications highlight the flexibility and value of emotion recognition technologies in various industries.

Emotion recognition methods are broadly divided into two categories: those that use human physical signals such as facial expressions, speech and gestures and those that use internal physiological signals such as BVP, ECG, heart-rate, GSR and RSP. While physical signals are simple to collect and have been extensively studied, they can be unreliable because people can purposefully control them to hide their true emotions, particularly in social situations. whereas, physiological signals are connected to the central and autonomic nervous systems, which react involuntarily to emotional stimuli. These signals change in specific ways when people encounter certain situations, which is consistent with Cannon's theory of emotion.Cannon (1927)The main benefit of using physiological signals in emotion recognition is that they are largely independent of voluntary control, making them possibly more accurate indicators of genuine emotional states. The involuntary nature of physiological responses provides a more accurate approach for emotion detection, as individuals find it much more difficult to manipulate these internal signals than external physical expressions.Shu et al. (2018)

1.2 Aim and Objectives

The dissertation aims to create a system that will analyze physiological signals, such as Blood Volume Pulse (BVP) and Heart Rate (HR) to recognize and classify human emotions. The system will use a variety of resources and technologies, such as signal processing, feature extraction, machine learning, and data analysis. The system's primary goal is to improve the understanding of human emotions through physiological data, providing helpful insights for applications in mental health monitoring, human-computer interaction, E-learning and adaptive user interfaces. The ultimate goal is to contribute to the development of a more logical and responsive technology ecosystem that effectively responds to human emotional states. Following are the associated Objectives of the project.

1. **Conduct Comprehensive Research and Develop Signal Acquisition Methods:**

   Perform an extensive literature review on physiological signal processing, emotion recognition, and the use of BVP and HR in affective computing. Design and implement reliable methods for acquiring and preprocessing BVP and HR signals, ensuring data quality through noise and artifact removal.

2. **Feature Extraction and Machine Learning Model Training:**

   Identify and extract significant features from BVP and HR signals that are indicative of emotional states. Train and evaluate various machine learning models using these features to classify emotions, selecting the best-performing model based on accuracy, precision, recall, and F1-score.

3. **System Development, Analysis, and Real-World Validation:**

   Develop a user-friendly, non-invasive emotion recognition system capable of real-time analysis. Implement data analysis techniques to create insightful visualizations of emotion data. Validate the system in practical applications, such as mental health monitoring and human-computer interaction, ensuring robustness and gathering user feedback for improvement.

Chapter 2

Literature Review

2.1 Foundation of human emotion analysis

Hofmann et al. (2020)This paper discusses Ekman's theory, which defines several basic emotions such as anger, disgust, fear, happy sad and surprise. It states that these emotions are experienced by everyone and can be recognised through specific facial expressions. The study then expands on the discussion by introducing Plutchik's model, which builds on Ekman's research by proposing that basic emotions can combine to produce more complex emotional experiences, such as love, which is the result of a combination of happiness and trust. Plutchik also developed the concept of emotional oppositions, which regard emotions like anger and fear or happy and sad as opposites, as shown in figure 2.1 Wheels of Emotion.Together, these ideas provide a thorough understanding of how fundamental emotions can function independently and in combination to create the complex structure of human emotional experience.

Wang et al. (2020)This study done an emotion analysis using various models, with the Valence-Arousal model being the primary one under assessment. This approach examines emotions in two dimensions: valence, which defines whether an emotion is positive or negative and arousal which determines an emotion's strength or amount of energy. This method allows emotions to be placed on a range of states. Figure 2.2 shows high arousal and positive valence emotions such as satisfaction and excitement in the top right quadrant and low arousal and negative valence emotions such as sadness and boredom in the bottom left quadrant. Rather than categorising emotions, this model shows the changing characteristics of emotional experiences and provide a framework for understanding and classifying them.

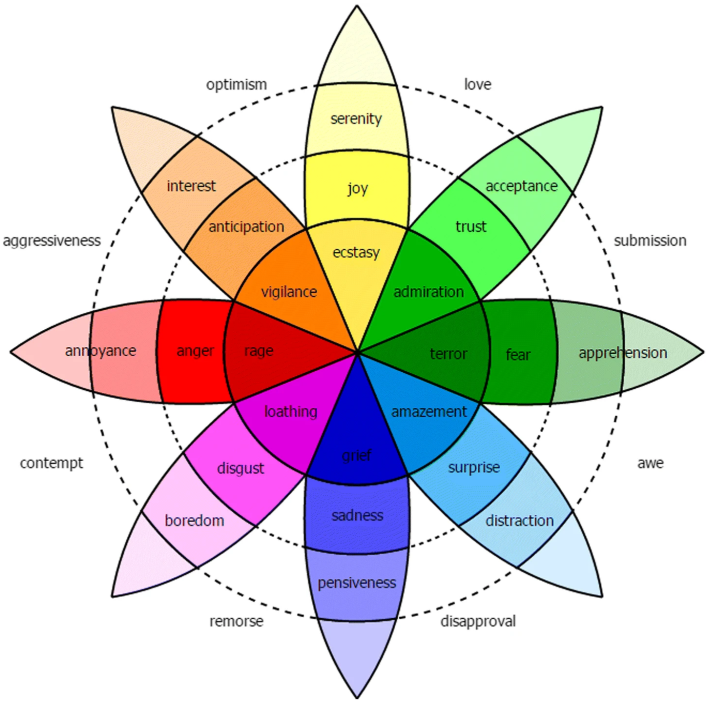

Figure 2.1: Plutchik's Wheel of Emotions

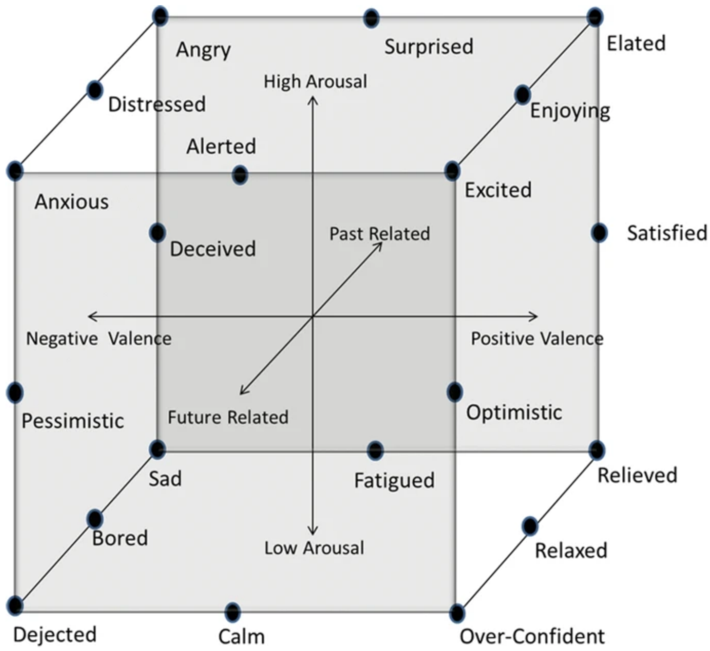

Figure 2.2: The specific emotion concepts inside the tri-dimensional model

Stark & Hoey (2021)This paper points out the complexities of emotions and how they can be understood, determined, and transformed into data for AI systems. It shows how emotions are multifaceted phenomena involving senses, behavioural reactions, psychological labelling, and physiological changes. This paper introduces two major challenges in emotion theory: the "problem of parts" (determining which elements are necessary for accurately describing and detecting emotions) and the "problem of plenty" (addressing how these elements interact to form an integrated emotional experience). Different conceptual models focus on different aspects of emotion, such as actual emotional states. These discussions are important for understanding the complexities involved in human emotion analysis, particularly in the context of AI where emotions must be accurately captured, measured and processed raising significant ethical and practical considerations.

2.2 Methods for Emotion Recognition Across Modalities

Emotion recognition across different modalities combines the inputs from the speech, text, facial expressions and biosensors to improve the accuracy of detecting human emotions. This multi-modal approach, powered by advances in AI and machine learning, enables more straightforward and empathetic interactions between humans and machines.Dzedzickis et al. (2020)

Singh et al. (2023) This paper developed an emotion recognition system using speech signals, integrating a self-attention-based deep learning model. The model is combination of 2D CNN with an LSTM network to capture spectral and rhythmic features. Mel Frequency Cepstral Coefficients were found to be the most effective for emotion recognition. A customized dataset merging RAVDESS, SAVEE and TESS was used to classify eight emotional states. The system achieved a 90 percent test accuracy better than existing models and showing its capability for automated mental health monitoring.

Gupta et al. (2023)The study presents a method for detecting facial expressions based on deep learning. Through facial expression classification during online sessions, the system generates an engagement index that predicts two states: "Engaged" and

"Disengaged." To find the best model emotion recognition, a number of deep learning models are utilised and compared, such as Inception-V3, VGG19, and ResNet-50.

Liu et al. (2023)This paper improves an emotion classification in text by using the Multi-label K-Nearest Neighbors classifier,combining data from adjacent sentences and the entire text.An accuracy and speed are increased through iterative corrections, particularly with short tweets. The effectiveness of these improvements in emotion classification is demonstrated by experiments that contrast the base MLkNN with its sample-based and label-based variations.

Le et al. (2023)This study proposes a multimodal emotion recognition in videos, combining video frames, audio and text subtitles for joint representation learning. The method improves multi-label classification performance by using multi-label representation. Analyses conducted on the IEMOCAP and CMU-MOSEI datasets reveal that the approach performs better than current baselines, proving that it's effective in recognising of emotions in videos.

2.3 Emotion Recognition Using Physiological Signals

Emotion recognition using physiological signals is gaining attention as they are linked with involuntary nervous system responses, which makes more reliable than physical signals like facial expressions or speech which can be consciously controlled and may not accurately reflect true emotions.Shu et al. (2018)

Fang et al. (2024)This study proposes an emotion recognition method using ECG signals with random convolutional kernels.The method maintains good accuracy while reducing training time and computing complexity by extracting the ECG features. It provides an effective method for emotion detection using ECG signals, outperforming earlier approaches in multi-class emotion recognition.

Veeranki et al. (2024) This study explores nonlinear signal processing for emotion recognition using EDA signals, addressing current limitations in capturing signal variations. In this study they combined IsaxEDA with an SVM classifier which delivered the best results when four approaches were compared using the CASE dataset, with an F1-score of 65percent. The results show that nonlinear EDA processing may improve the ability to recognise emotions and assist in the detection of mental health issues such as anxiety and depression.

Dobrokhvalov & Filatov (2024)This study explores emotion recognition using physiological signals like blood volume pulse (BVP), electrocardiogram (ECG), and electrodermal activity (EDA) with convolutional neural networks. A two-branch model processes heart data and EDA separately, then merges the results. The approach, tested on the various datasets which showed high accuracy and F1 scores and highlights the importance of bvp,eda and ecg signal.

Ayata et al. (2020) Taking advantage of improvements in wearable technology, this study developed a new emotion recognition using multimodal physiological data which is collected by a respiratory belt, photoplethysmography (PPG), and fingertip temperature (FTT) sensors. Using machine learning techniques such as logistic regression, random forest and SVM, arousal and valence levels were determined. By applying decision-level fusion of signals, the accuracy improved to 73.08% for arousal and 72.18% for valence, it shows its effectiveness in healthcare systems.

Chapter 3

Preparation and Design

3.1 MosCow Analysis

The MoSCoW Method, a prioritization technique commonly used in project management, that provides an organised way to rank requirements according to their urgency and significance.Kostev (2023) Here in my project MoSCoW analysis will help to categorize the requirements for developing an emotion recognition system using physiological signals such as Blood Volume Pulse (BVP), Electrodermal Activity (EDA) and Heart Rate (HR). The goal is to prioritize features based on their importance to the system's functionality.

3.1.1 Must Have

Follwing are the important components without which the system cannot function effectively.

1. **Data Collection and Processing** : The system must reliably capture physiological signals (BVP, EDA, HR) using appropriate sensors. Necessary to remove noise and artifacts to ensure clean data for analysis.

2. **Feature Extraction :** Extracting the relevant features from the physiological signals such as Mean, Arc-Length, Energy etc.

3. **Machine Learning Models:**Implementation of models like XGBoost, Random Forest or deep learning to classify emotions based on physiological signals.

4. **Model Validation:** Testing the models on datasets to ensure the accuracy.

5. **Performance Metrics:** Measuring the systems performance with respect to recall, accuracy and precision.

3.1.2 Should-Haves

Significant additions that will improve the system but are not necessary for its core functioning.

1. **Advanced Signal Processing:** Try to implement more advanced techniques to improve signal quality further.
2. **Personalization:** To increase accuracy, allow user-specific model modifications.
3. **Multi-modal system:** Combine physiological signals with other data sources or modalities such as facial expression, text to improve emotion detection accuracy.
4. **Emotion Reporting and Feedback:** Utilising a user interface, provide immediate visual feedback for recognising the emotions.

3.1.3 Could-Haves

characteristics that would be ideal to include if resources and time is enough.

1. **Future State Prediction:** Provide a function that uses history and current data to forecast future emotional states.
2. **Wearable Device Integration:** Enable connectivity for real-time data collecting with devices such as smartwatches.
3. **Cloud-Based Processing:** For scalable data processing and storage, use cloud computing.

3.1.4 Won't-Haves

These features are out of scope for the current project phase.

1. **Non-Physiological Data Sources:** The project will focus only on physiological signals, excluding behavioral data like social media activity.
2. **Clinical Trials:** The system will not undergo widespread clinical trials in this phase.

3. **Cross-Cultural Adaptation:** At this point, the system will not be modified to account for variations in expressing emotions.

3.2 Gathering and Preparation of data

3.2.1 Data Source

I looked through a number of online datasets in an attempt to find one that would be suitable for my dissertation work on physiological signal-based emotion recognition. I considered datasets such as DEAP, CASE which provides a wide range of physiological and emotional data, but ultimately after considering all my options, I selected the K-EMOCON Dataset available on Zenodo. Park et al. (2020)

I selected the K-EMOCON Dataset due to its ability to offer a vast collection of different physiological signal data, which is important for accurately researching emotion recognition. Here, 32 participants, ranging in age from 19 to 36 contributed data to the collection.

The participants took part in 16 paired up debate sessions, lasting roughly ten minutes each, where they debated issues of societal concern. The dataset collects physiological signals that are necessary for the study I intended to perform, such as Heart Rate (HR), Electrodermal Activity (EDA), and Blood Volume Pulse (BVP) and more.

The additional point to choose this data is that it contains emotional annotations for every participant which could be useful for extracting the labels for the emotion recognition system.

3.2.2 Data Collection

Methods and Tools

Data for the K-EMOCON Dataset was collected using:

- Empatica E4 Wristband: Collected Blood Volume Pulse (BVP) and Electrodermal Activity (EDA) data.
- Polar H7 Bluetooth Heart Rate Sensor: collected Heart Rate (HR) data from ECG signals.

Table 3.1: Summary of Data Collection and Annotation from K-EMOCON

Data Type	**Device**	**Sampling Rate**	**Arousal-Valence (1-5)**	**Categorical Emotions(1-4)**
BVP	E4 Wrist-band	64 Hz	Yes	Happy, Angry, Sad, Cheerful, Nervous
EDA	E4 Wrist-band	4 Hz	Yes	Happy, Angry, Sad, Cheerful, Nervous
HR (ECG-derived)	Polar H7 HR-Rate Sensor	Derived from ECG	Yes	Happy, Angry, Sad, Cheerful, Nervous

Environment Setup

To guarantee consistent physiological signals, data was gathered in a highly controlled laboratory environment.

Procedures:

baseline assessments and continuous records while debating. For every session, meta-data (start time, end time, and initial time) was captured.

Annotation Procedures

Participants self-annotated their emotions using two scales:

- **Arousal and Valence:** Rated on a scale from 1(low) to 5(high).
- **Categorical Emotions:** (Cheerful, Happy, Angry, Nervous, Sad) rated on a scale from 1(low) to 4(high).

3.2.3 Data Pre-Processing

BVP (Blood Volume Pulse) Data Preprocessing

The 64 Hz sampled BVP data needed to be carefully preprocessed in order to ensure quality. I first attempted median and simple high-pass filtering, but both techniques either distorted important signal details or left too much noise in the data. A fixed threshold for artifact removal also prooved too harsh resulted in the loss of important data.

Final Methods:

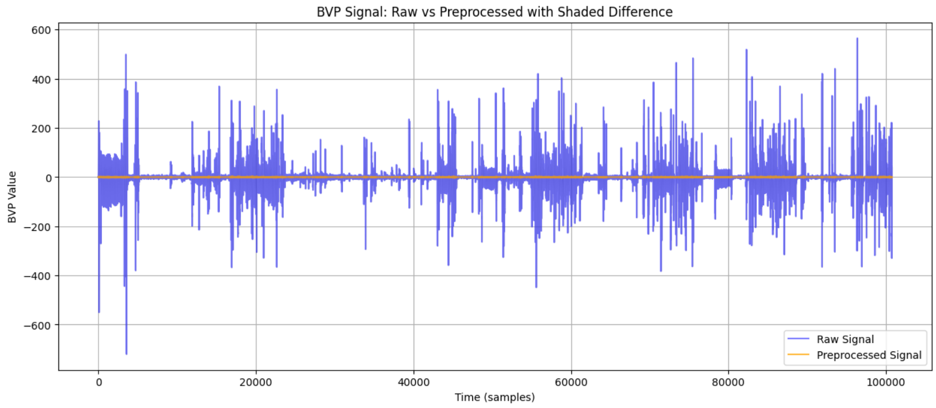

Figure 3.1: Raw vs. Preprocessed BVP signal

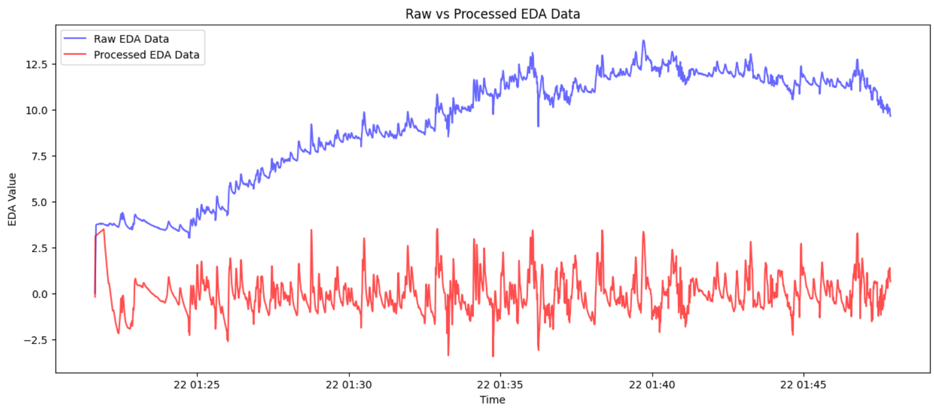

Figure 3.2: Raw vs processed EDA signal

- Bandpass Filtering: 0.5-8.0 Hz bandpass filter was used for successfully removing high-frequency noise and low-frequency drift.
- Z-Score Normalization: standardised the signal for taking individual variances into consideration and ensure cross-participant comparability.
- Artifact Removal: Replaced outliers which are more than three standard deviations from the mean, adapting better to signal variability.
- Savitzky-Golay Smoothing: used to reduce noise while maintaining significant signal properties.

EDA (Electrodermal Activity) Data Preprocessing

The EDA data which sampled at 4 Hz, was processed in order to get accurate measurement of emotional states. Initial approaches, such threshold-based artefact removal and basic filtering, were insufficient because they either eliminated important signal features or were unable to adjust to the data's variability.

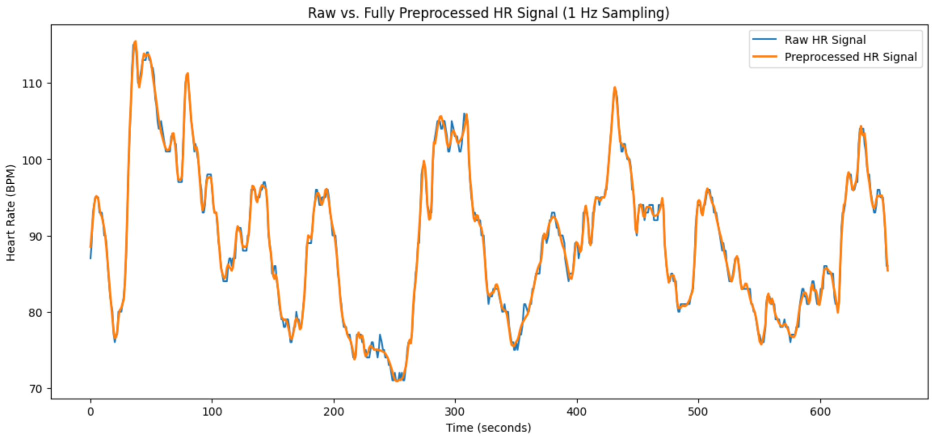

Figure 3.3: Raw vs. Processed

Final Methods:

- Low-Pass and High-Pass Filtering: Removed both high-frequency noise and low-frequency drift by applying a high-pass filter with a cutoff at 0.01 Hz and a low-pass filter with a cutoff at 1.0 Hz.
- Detrending: Removed the trends from the EDA signal in order to remove baseline shifts over time.
- Dynamic Artifact Removal: eliminated artifacts that were more than three standard deviations from the mean and the gaps left by the removed artifacts filled in via interpolation.
- Z-Score Normalization: Normalised the signal after removing artifacts to assure consistency of signal.

HR (Heart Rate) Data Preprocessing

The HR data was sampled at 1 Hz generated from ECG signals, it has already undergone considerable processing to ensure accuracy. To make sure the data was best suited for analysis, more preprocessing done to further improve it.

Methods Applied:

- **Low-Pass Filtering:** Butterworth Low-Pass Filter: A low-pass Butterworth filter with a cutoff frequency of 0.4 Hz was applied to the HR data.

- **Noise Reduction:** The low-pass filter used to the HR data to further smooth it out and eliminate any little fluctuations that might have been noise artifacts or leftover noise from the ECG signal.

Emotion Annotation Data Preprocessing

Every five seconds, participants annotated themselves. They rated their level of arousal and valence on a range of 1 to 5, as well as the emotions such as happy, angry, nervous, sad, and cheerful on a scale from 1 to 4. To simplify the data:

- **Arousal/Valence:** For each 5-second interval, a new column I created a new column to indicate whether arousal or valence had the higher rating, capturing the dominant emotional dimension.

- **Emotion Labeling:** Another column labeled ”emotion” also created, which is populated by the highest-rated emotion for each interval. If all emotions were rated as 1, the label ”neutral” emotion used.

By preprocessing and creating these emotion label, the emotion annotation data became easier to summarize for each 5-second interval.

3.3 Methods, Classifiers and Technology Stack

3.3.1 Feature extraction

Feature extraction involves expressing the most important patterns in a binary, categorical or continuous form from the signal data.I extracted features of BVP, EDA and Heart rate (ECG) for each participant.Khan et al. (2023)

- **BVP (Blood Volume Pulse) Feature Extraction**

I extracted a variety of features from the BVP signal in order to record the time- and frequency-domain characteristics that are necessary for identifying emotions.

Time-Domain Features

1. Mean : represents the BVP signal's average value, which shows the total amount of cardiovascular activity.

2. Standard Deviation : evaluates the signal's variability by taking into account variations from the mean.

3. Arc-Length :measures the total variance across time to capture the complexity of the signal.

4. Skewness : measures the signal's asymmetry and displays a bias towards either higher or lower values.

5. Kurtosis : determines the signal's distribution extremity, which shows the existence of outliers.

Frequency-Domain Features

1. High-Frequency Power Spectral Density (HF-PSD): indicates the strength in the high-frequency range that is associated with the impact of respiration on heart rate.

2. Sample Entropy (SampEn): Evaluates the signal's complexity or unpredictable nature.

Table 3.2: Feature Extraction Formulas for BVP Signal

Feature	**Formula**
Mean (μ)	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
Standard Deviation (σ)	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$
Arc-Length	$\text{Arc-Length} = \sum_{i=1}^{N-1} \sqrt{1 + (x_{i+1} - x_i)^2}$
Skewness (γ_1)	$\gamma_1 = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{x_i-\mu}{\sigma}\right)^3$
Kurtosis (γ_2)	$\gamma_2 = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{x_i-\mu}{\sigma}\right)^4 - 3$
High-Freq Power Spectral Density (HF_PSD)	$\text{HF\_PSD} = \int_{f_1}^{f_2} P(f)\, df$
Sample Entropy (SampEn)	$\text{SampEn}(m, r, N) = -\ln\left(\frac{A}{B}\right)$

- **EDA (Electrodermal Activity) Feature Extraction:**

For the EDA signal, I extracted features in three categories: trough-to-peak features, decomposition-based features, and statistical features.Lutin et al. (2021)

Trough-to-Peak Features

1. Number of SCRs: Total reactions in skin conductance.

2. Summed Magnitude of SCRs: total of the emotional reactions' intensity.

3. Summed Area Under SCR: Cumulative intensity, considering both magnitude and duration.

Decomposition-Based Features

1. Mean Tonic Level: initial value of skin conductance.

2. Standard Deviation of Tonic Level: Baseline arousal variability.

3. Standard Deviation of Phasic Activity: Differentiation in emotions of response.

4. Number of Responses: number of unique arousal events.

Statistical Features

1. Skewness: Asymmetry in the distribution of the EDA signal.

2. Kurtosis: distributional extremity or tails of the signal.

Table 3.3: Feature Extraction Formulas for EDA Signal

Feature	**Formula**
Number of SCRs	Count of skin conductance responses
Summed Magnitude of SCRs	$\sum_{i=1}^{N}$ Magnitude(SCR_i)
Summed Area Under SCR	$\sum_{i=1}^{N}$ Area under SCR_i
Mean Tonic Level	$\mu_{\text{tonic}} = \frac{1}{N}\sum_{i=1}^{N} T_i$
Standard Deviation of Tonic Level	$\sigma_{\text{tonic}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - \mu_{\text{tonic}})^2}$
Standard Deviation of Phasic Activity	$\sigma_{\text{phasic}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - \mu_{\text{phasic}})^2}$
Number of Responses	Count of distinct responses in the phasic activity
Skewness	$\gamma_1 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^3$
Kurtosis	$\gamma_2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^4 - 3$

- **Heart Rate (HR) Feature Extraction:**

For the HR signal derived from ECG, I extracted key heart rate features:

1. **AVNN:** Mean heart rate is the average of all normal-to-normal (NN) intervals.

2. **SDNN:** An indication of total HR is the standard deviation of NN intervals.

Table 3.4: Feature Extraction Formulas for Heart Rate (HR) Signal

Feature	**Formula**
Average NN Interval	$\text{AVNN} = \frac{1}{N}\sum_{i=1}^{N} NN_i$
Standard Deviation of NN Intervals	$\text{SDNN} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(NN_i - \mu_{NN})^2}$
Root Mean Square of Successive Differences	$\text{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(NN_{i+1} - NN_i)^2}$
Standard Deviation of Successive Differences	$\text{SDSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(NN_{i+1} - NN_i)^2}$
Coefficient of Variation	$\text{CV} = \frac{\sigma_{NN}}{\mu_{NN}}$

3. **RMSSD:** short-term heart rate is reflected by the root mean square of the subsequent differences between NN intervals.

4. **SDSD:** The standard deviation shows the variation in heart rate with successive differences in NN intervals.

5. **CV:** coefficient of variation, which uses the mean heart rate to normalise Heart rate.

3.3.2 Data Synchronization and Dataset Creation

The next step involved combining the above features with the emotion annotations to produce a complete dataset, after the preprocessing of the physiological signals (BVP, EDA, and HR) and the extraction of important features. In order to ensure accurate alignment of the physiological data and emotion annotations, a careful synchronisation method was required.

Merging Process

- **Timestamps for Features:** At the time of data collection, each physiological feature (BVP, EDA, and HR) was first captured with a precise timestamp.

- **Emotion Annotation Timing:** A seconds column in the emotion annotation file provided the times at which participants reported their emotional states at 5-second intervals (e.g., 5, 10, 15 seconds).

- **Synchronization Process:** I converted the physiological features timestamps into a datetime format and matched them with the 5-second intervals in the emotion annotations to synchronise the data.

Physiological features were aggregated to match these 5-second intervals,by calculating relevant metrics for each interval, ensuring the data accurately represented the participant's emotional state.

- **Data Alignment:** The start and end times of each participant's session was given , I used that to further improve the synchronisation.

  The resulting dataset consists merged data of all participants which consists rows representing 5-second intervals, each containing synchronized physiological features and self emotion annotation columns with participant ID.

3.3.3 Classification models

The next stage was to choose appropriate machine learning models for emotion recognition after generating a synchronised dataset that matches physiological features with emotion labels. I reviewed several kinds of research papers on the recognition of emotions from physiological signals. Based on this investigation, I found and selected the following models for my project due to their strong performance in similar contexts:

Random Forest (RF)

With the use of the Bagging approach, Random Forest (RF), an ensemble learning methodology, builds several decision trees. At every split, a random subset of features is selected to identify the ideal split, with each tree being trained on a bootstrap sample. Robust against overfitting and effectively managing high-dimensional data, random forests (RF) aggregate the output of all trees by majority voting for classification or averaging for regression. Cai et al. (2023)

XGBoost (Extreme Gradient Boosting)

A advanced gradient boosting technique called XGBoost sequentially constructs decision trees, each of which fixes the mistakes of the one before it. In order to avoid overfitting, regularisation (L1 and L2) is included, and a second-order Taylor expansion is used to more efficiently optimise the loss function. For large datasets, XGBoost is particularly efficient and scalable since it enables parallel processing and manages sparse data efficiently.Lebaka et al. (2023)

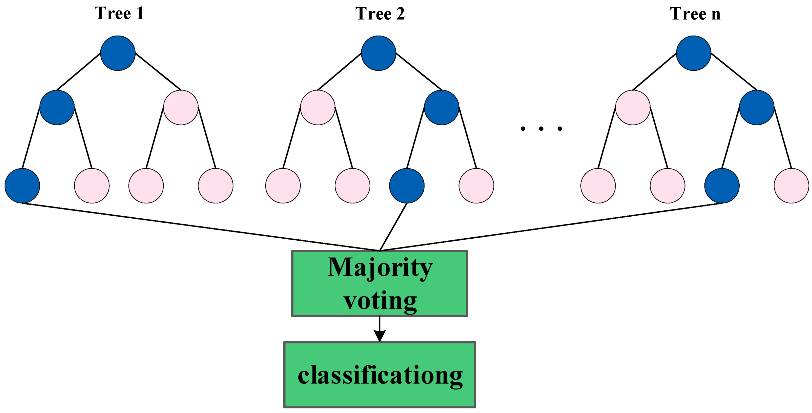

Figure 3.4: Structure of RF Cai et al. (2023)

Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) is a computational model that is based on the structure of the human brain, with layers of interconnected neurones. Each neurone gets input, applies weights, adds bias, and transmits the outcome with an activation function like ReLU or sigmoid to generate nonlinearity.Aung et al. (2022) :

$$z = \sum_{i=1}^{n} w_i \cdot x_i + b$$

where $x_1, x_2, \ldots, x_n$ are inputs, $w_1, w_2, \ldots, w_n$ are weights, and b is the bias. The output is then:

$$a = \sigma(z)$$

In forward propagation, the input data passes through each layer, and the network produces an output prediction. To improve the model, backpropagation adjusts the weights by minimizing the loss function through gradient descent:

$$\Delta w = -\eta \frac{\partial L}{\partial w}$$

Chapter 4

Technical Implementation and Evaluation

4.1 Design of methodology

Building on the preprocessing of physiological signals (BVP, EDA, HR) and emotion annotation detailed in section 3.2, this section focuses on the practical implementation of the emotion recognition system. Here the discussion on the feature extraction methods, feature selection process and the training of various machine learning models, along with relevant technical aspects.

4.1.1 Feature Extraction Methods

Feature extraction was an important step in this study, that require the use of various analytical methods to extract meaningful patterns from physiological signals. Below, I describe the methods used to extract relevant features, which were explained in 3.3.1.

1. **Statistical Analysis:** This method was used to obtain basic time-domain features from the signals, such as mean and standard deviation. These features summarise the signal's overall behaviour, including its primary tendency and variability.

2. **Welch's Method:** To estimate the signals' Power Spectral Density (PSD), Welch's method was used. This method is essential for analysing frequency-domain features, particularly in non-stationary physiological data, as it captures how power is distributed in different frequency bands.

3. **STFT (Short-Time Fourier Transform)**: STFT was used to determine how

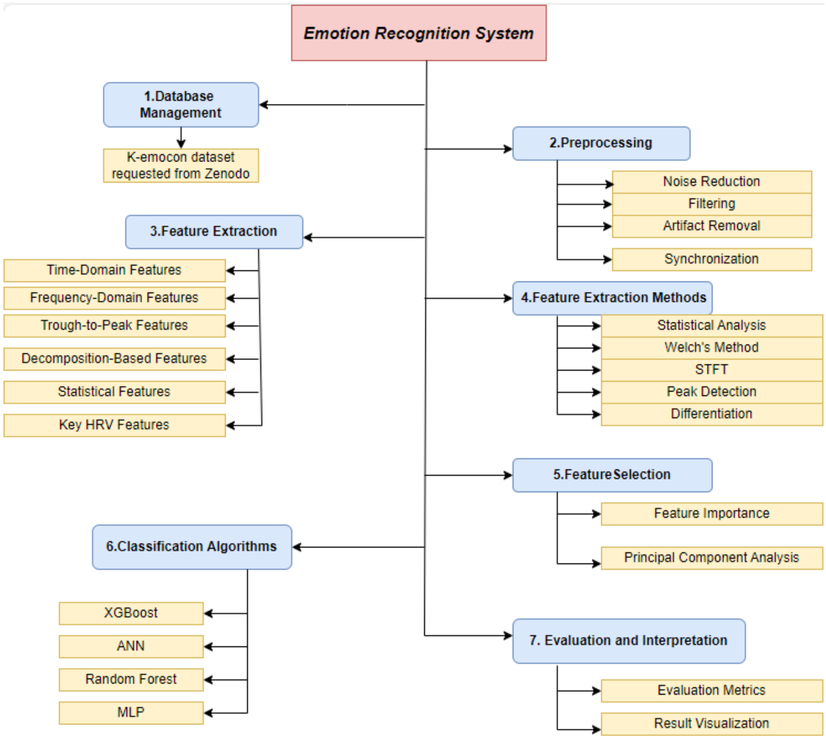

Figure 4.1: Flowchart

the frequency content of the signals changes over time. This method is important for identifying time-varying frequency components, which provides more understanding into how physiological responses develop in response to emotional stimuli.

4. **Peak Detection**: Peak detection algorithms were used to find significant peaks in the EDA signal that correlated to Skin Conductance Responses (SCRs). This method allowed for the accurate estimation of independent physiological responses related to emotional arousal.

5. **Differentiation**: Differentiation was used to identify the phasic components of the EDA signal, which changes rapidly in response to stimuli. This technique allowed for the extraction of features that detect fast, emotion-driven changes in physiological state.

Creation of Lag Features

Given the continuous nature of physiological signals, I created lag features by shifting the original BVP signal features by various time intervals. This method was important in capturing the delayed effects of stimuli on physiological responses allowing the model to recognise patterns that appear over time. Specifically, I developed lag features such as arc-length-lag, mean-lag, and standard-dev-lag. These features represent the mean, standard deviation and arc length of the BVP signal with a one-time-step delay, allowing for the analysis of temporal dependencies in the data.

4.1.2 Feature Selection

As mentioned in the paper Htun et al. (2023)The method for choosing features intends at minimising irrelevant variables and thus enhance the performance of machine learning models. If we only use an appropriate amount of features as input for an ML model, the data may be insufficient to make predictions. A large number of features adds operating time and impacts generalisation effectiveness due to lack of dimensionality. As a result, selecting just the important features makes sure the model receives accurate and helpful data for prediction, which boosts its overall performance. In this project, I used the correlation matrix to figure out how various input features are linked as shown in figures 4.2 4.3. Correlation helps to determine whether one feature can predict another. When a correlation between features is close to 1 or -1, it suggests that they have a close relationship and can be predicted from each other. To improve the accuracy of the model, it is better to eliminate one of these closely related features. This also allows us to avoid the lack of dimensionality when selecting features. Here are the pairs of highly related features and their results:

Variance Inflation Factor (VIF)

I used a VIF analysis to figure out multicollinearity for the features. The VIF plot showed that the const feature has a high VIF, which would be expected for the term of intercept and did not cause concern. All other features have VIF values greatly less than the threshold of 10, suggesting low multicollinearity as hown in the figure 4.4. This analysis shows that the features I chose provide unique data that is suitable for use in the model.

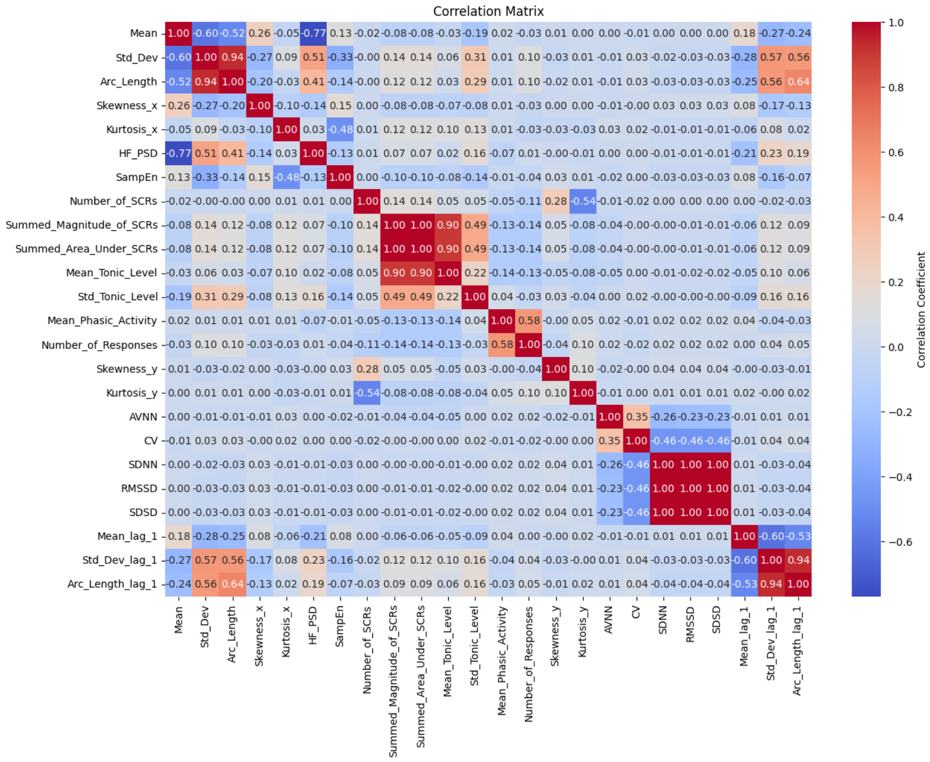

Figure 4.2: highly correlated features

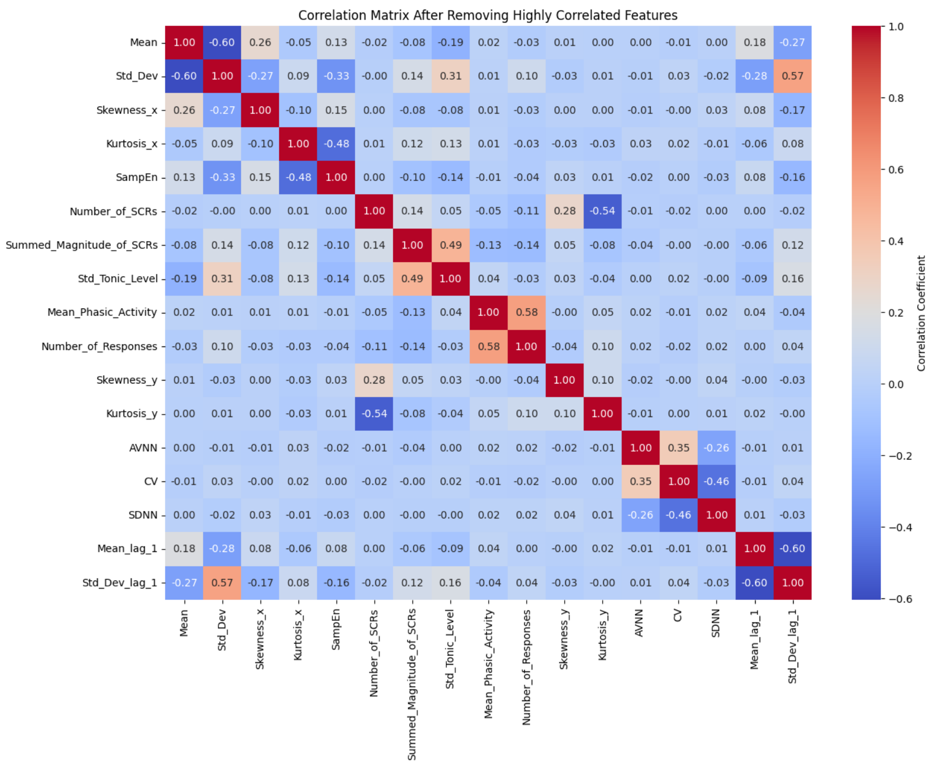

Figure 4.3: Removed highly correlated features

4.1.3 Data Process and Feature Scaling

Categorical encoding

Machine learning methods work with numerical values data, not the data that is text-based. As an outcome, in order for methods to recognise categorical features, they must be encoded as numerical values. The two most common encoding methods are label encoding and one-hot encoding. In label encoding, every category gets an independent numerical value. However, in one-hot encoding, every class gets a binary vector that includes 1s and 0s. One disadvantage of one-hot encoding is that it includes new features to the a database, which can result in the curse of dimensions. This indicates to increased data complexity as result of high dimensions, which can have a negative effect on performance. Label encoding allocates each category a particular value and as a result, a machine learning algorithm may believe that one category is better over another based on its numerical value. In this project, I adopted label encoding for the features. Label encoding replaces the desired categories such as happy, angry, nervous, cheerful, and neutral with numbers 0, 1, 2, 3, and 4, and target column a-v arousal and valence with 0 and 1, respectively .

Data Splitting

To evaluate the models' performance, I divided the dataset into training and testing sets using standard format. To assure that the class proportions in the target variables remained consistent across both sets.

Class Distribution and Balancing

Given the observed imbalance in the target classes, I applied specific techniques to balance the class distribution, ensuring that the models were trained on a representative dataset:

- **Upsampling:** As shown in the figure.4.6 For arousal-valence data, I used upsampling to address class imbalance in models. This method involves producing more copies of samples from the minority classes until they equal the number of samples in the majority classes. Upsampling makes sure that all classes have an equal representation in the training set, which allows the model to learn from a

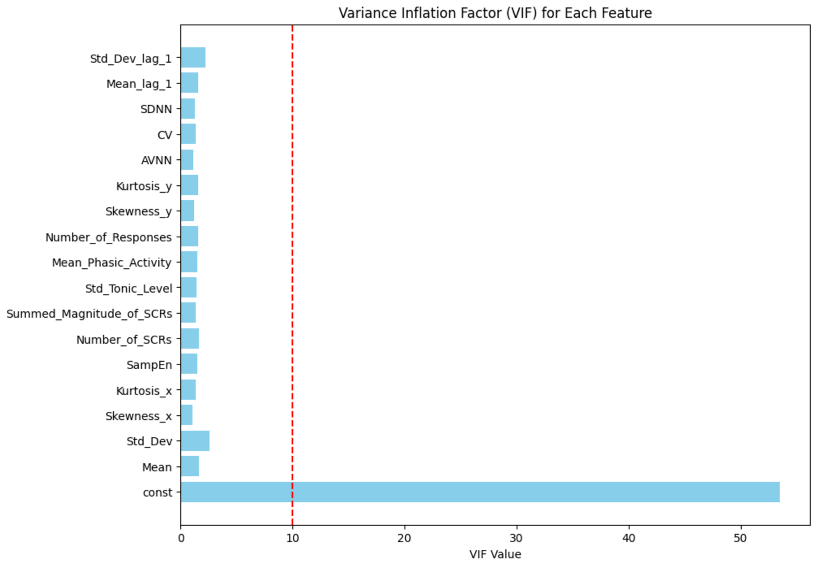

Figure 4.4: Variance Inflation Factor

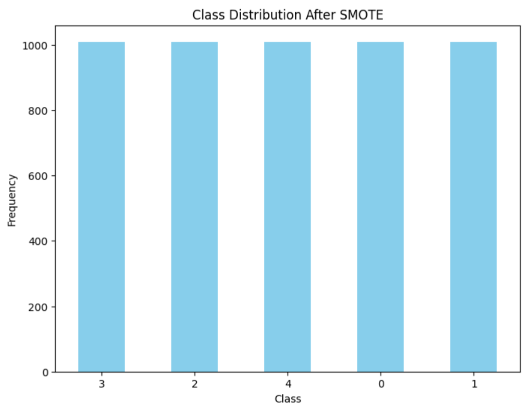

Figure 4.5: Emotion Class after SMOTE

Figure 4.6: Arousal-Valence after distribution

balanced dataset.

- **SMOTE (Synthetic Minority Over-sampling Technique):**

  As illustrated in the figure.4.5 For emotion class, I used SMOTE. This approach addresses class imbalance by generating synthetic samples for minority classes using interpolation from existing samples. SMOTE produces new, diverse data points, producing a more representative and unique dataset than simple duplication.

Feature Scaling

To standardise the features and make sure that they are on a similar scale, I used standard scaling, which normalises the features to a mean of zero and a standard deviation of one. This step was particularly necessary for models like ANN and MLP, where optimising algorithms are dependent on the scale of input features, resulting in better model performance during training.

4.1.4 Model Training and Hyperparameter Tuning

After completing the processing and feature extraction stages, I moved forward with training the selected machine learning models—Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), and Artificial Neural Network (ANN)—as discussed in section 3.3.3.

Training Process

The training process had been organised to investigate various combinations of physiological features and target columns. Training the models separately for both of these targets enabled an evaluation of their ability to handle both categorical and continuous data.

Feature Combinations:

1. BVP + EDA: The models were first trained with features obtained from Blood Volume Pulse (BVP) and Electrodermal Activity (EDA) signals.

2. BVP + EDA + HR: To improve the accuracy of predictions, Heart Rate (HR)

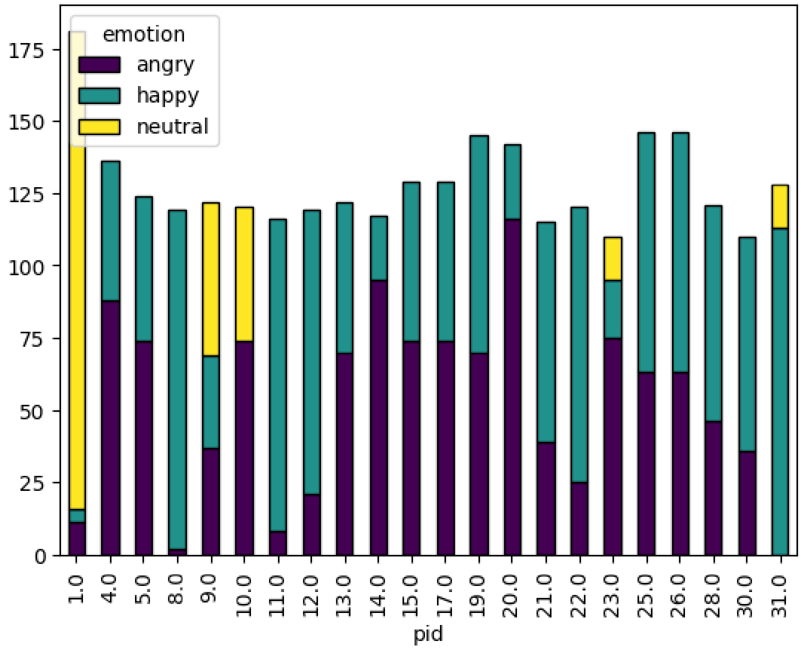

Figure 4.7: Class imbalance for loso

data was added to the feature set. The feature of this physiological signal significantly improved emotion recognition accuracy and precision.

Training for Target Columns:

1. Emotion Target Column: The first phase of training involved using the emotion target column, where the goal was to classify different emotional states.

2. Arousal-Valence Target Column: Following the emotion classification, the models were retrained using the arousal-valence target column.

Challenges with LOSO Validation:

I tried to evaluate the models using the Leave-One-Subject-Out (LOSO) validation method, which is commonly used to make sure cross-subject generalisation. However, due to major data imbalance, the models struggled during validation, resulting in poor performance shown in the figure 4.7. As a result, LOSO was considered unsuitable, and other validation methods were prioritised for more reliable results.

Hyperparameter Tuning and cross validation

Hyperparameter tuning is essential for improving machine learning models. GridSearchCV determines all predefined hyperparameter combinations systematically to ensure that the best set is used for optimal performance. RandomizedSearchCV, which I used in this project, is a faster alternative that randomly samples a fixed number of combinations, thus being more effective for wide hyperparameter spaces. Agrawal & Agrawal (2021). Cross-validation, implemented via the 'cross-val-score' function in scikit-learn, was used to train and evaluate the Random Forest classifier, providing accuracy scores for each fold to ensure robust model performance.

4.2 Technical Aspects

This section provides an overview of the tools and resources utilized throughout the dissertation, covering all stages of the research process, including data collection, pre-processing, model building, and evaluation.

Programming Language

Python had been the most commonly used programming language due to its extensive libraries and frameworks designed for data science and machine learning. Its flexibility made it perfect for handling complex data, extracting features, and developing models.

Integrated Development Environment (IDE)

Google Colab used as the IDE, providing a cloud-based platform for large-scale computing. Its effortless compatibility with Python libraries, as well as its collaborative features, made it the ideal tool for developing and executing machine learning models.

Data Source

The physiological data was sourced from the K-EMOCON dataset on Zenodo, which provided recordings of BVP, EDA, and HR signals and self emotion annotations from each participant, important for training and evaluating the emotion recognition models.

Data Storage

Data was stored in CSV format because it is simple and compatible, allowing for efficient storage and retrieval during the modelling phase.

Data Visualization

Data visualisation was done using Matplotlib and Seaborn, which allowed for the creation of detailed and visually appealing charts that were necessary for exploratory data analysis and results presentation.

Machine Learning Modeling

Scikit-learn was the main library for building and implementing machine learning models, including Random Forest, XGBoost, MLP and ANN providing a robust toolkit for model training and evaluation.

Hyperparameter Tuning

RandomizedSearchCV and GridSearchCV are used to tune hyperparameters, resulting in optimal model performance through wide parameter exploration.

Model Evaluation

Model evaluation utilized Scikit-learn metrics, including accuracy, precision, recall, and F1-score, to assess the effectiveness of the models in predicting emotional states based on physiological signals.

4.3 Evaluation and results

4.3.1 Results

The performance evaluation of the machine learning models revealed different variations across various feature sets and target columns. Initially, models trained on features derived from Blood Volume Pulse (BVP) and Electrodermal Activity (EDA) alone demonstrated moderate accuracy levels. The Random Forest model, for instance, achieved an accuracy of 49% in emotion classification and 59% in arousal-valence (A-V) prediction. Similarly, the XGBoost model showed 47% accuracy for emotions and 62% for A-V, while the Multi-Layer Perceptron (MLP) model lagged behind with 35% accuracy for emotions but performed better in A-V prediction with 68% accuracy.

Significant improvements were observed when Heart Rate (HR) data was incorporated into the feature set as we can see in the 4.8 and also hyperparameter tuning was applied. The Random Forest model's emotion classification accuracy rose to 66%,

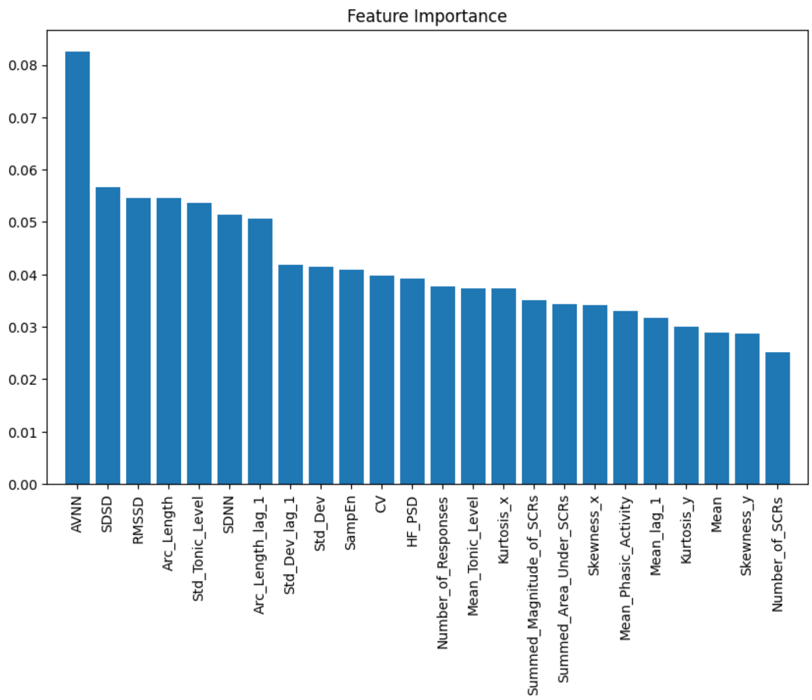

Figure 4.8: Feature importance

and its A-V prediction accuracy improved to 74%. XGBoost also showed enhanced performance, reaching 64% accuracy for emotions and 73% for A-V. The ANN model, evaluated only on the A-V target, achieved an accuracy of 69%, highlighting the positive impact of including HR data.

These results highlight the importance of feature selection and model optimisation in improving the accuracy of emotion and A-V state predictions from physiological signals.

Feature Set	**Model**	**Emotion Accuracy (%)**	**A-V Accuracy (%)**
BVP + EDA	Random Forest	49	59
	XGBoost	47	62
	MLP	35	68
BVP + EDA + HR	Random Forest	66	74
	XGBoost	64	73
	ANN	-	69

Table 4.1: Evaluation Results for Different Feature Sets and Models

4.3.2 Evaluation of Results

Performance of key metrics

In this project, I employed various evaluation metrics to assess the performance of the machine learning models on the emotion recognition task. The key metrics used include Precision, Recall, and F1-Score as we can see in figures 4.9, 4.10.below are the key metrics for best performed model.

- Precision: Precision is defined as the proportion of true positive predictions to all positive predictions made by the model. It indicates the model's ability to correctly recognise instances of relevance among all predictions. The Random Forest model performed well in predicting the "happy" and "valence" classes, indicating that it is correct.
- Recall : Recall is defined as the proportion of true positive predictions to all instances in the actual class. It evaluates the capacity of the model to detect all relevant instances. In results, the model showed powerful recall for the "happy" and "valence" classes, indicating its ability to correctly identify these classes even in the presence of class imbalances.
- F1-Score : The F1-Score is the inverse mean of Precision and Recall, representing a balance between the two metrics. It is particularly helpful when working with unbalanced datasets. The Random Forest model had the highest F1-Scores in the "happy" and "valence" classes, as it was the most reliable model in study for these particular target columns.

Confusion Matrix Analysis

1. **BVP + EDA with Emotion Target:** The confusion matrix for the best-performing model using BVP and EDA shows strong classification of "Happy" with 289 correct predictions, but struggles with "Angry," "Cheerful," and other emotions due to class imbalance. The MLP model faced significant challenges, leading to misclassifications, particularly in less frequent classes.
2. **BVP + EDA with Arousal-Valence (A-V) Target:** For the A-V target, the model showed a bias towards "Valence" with 370 correct predictions, but struggled with "Arousal." The MLP model, despite 68% accuracy, had difficulty with "Arousal" classification even after balancing techniques.
3. **BVP + EDA + HR with Emotion Target:** Including HR data significantly improved model performance, reducing misclassifications between emotions like

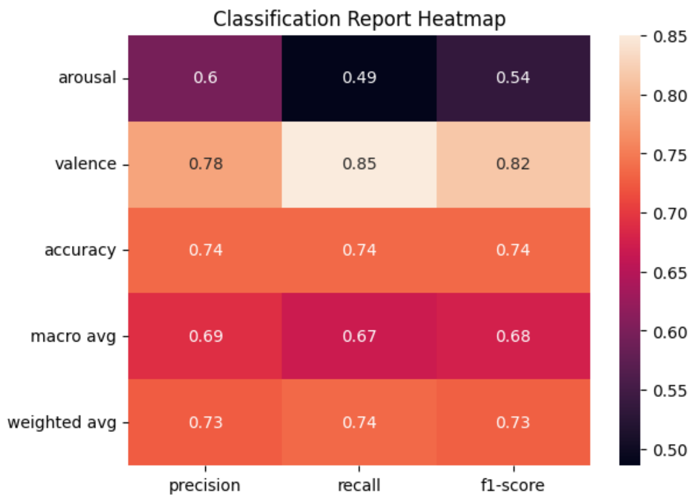

Figure 4.9: Arousal-Valence Heatmap

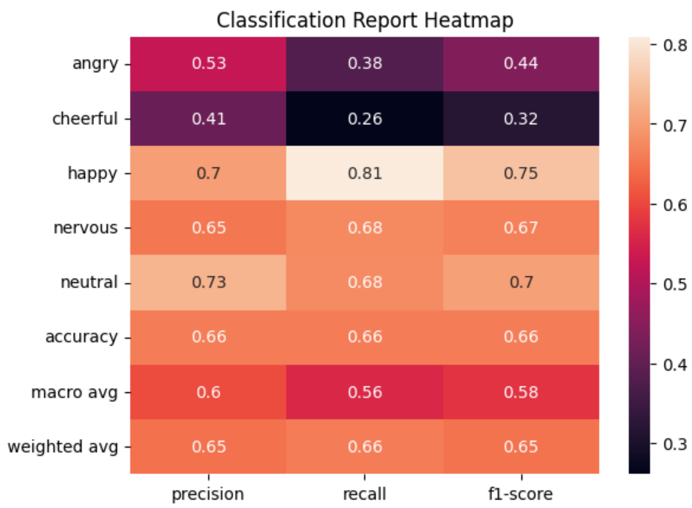

Figure 4.10: Emotion Heatmap

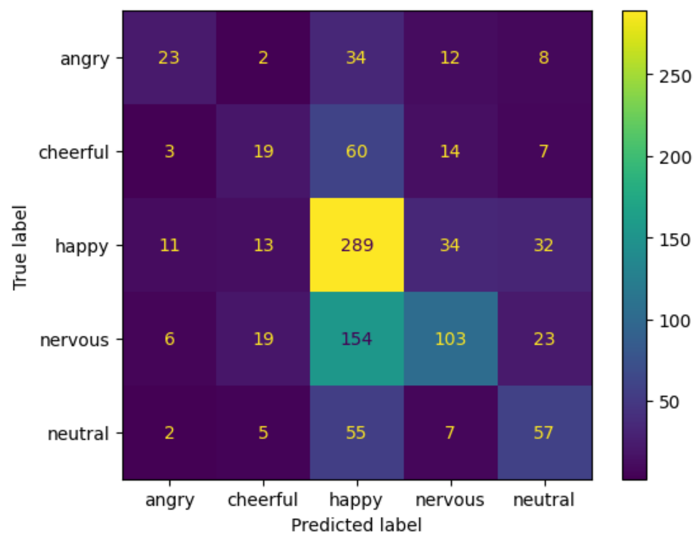

Figure 4.11: Random Forest with Emotion(BVP+EDA)

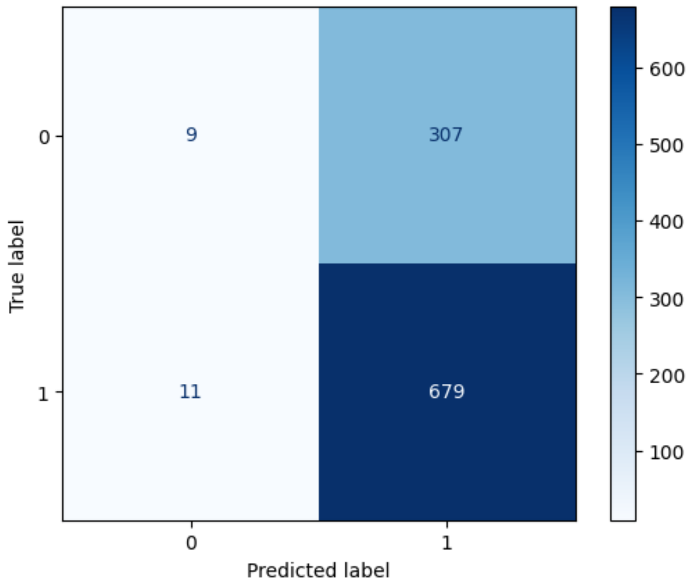

Figure 4.12: MLP with A-V

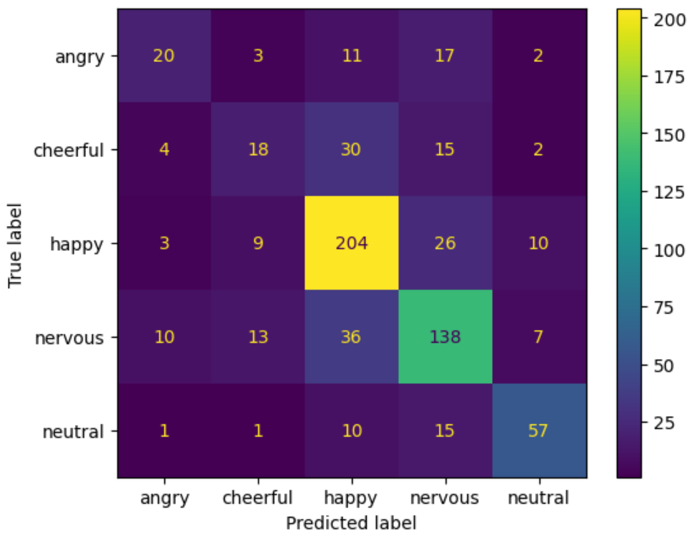

Figure 4.13: Random Forest with Emotion(BVP+EDA+HR)

"Nervous" and "Cheerful," and leading to a more balanced classification overall.

4. **BVP + EDA + HR with Arousal-Valence (A-V) Target:** The model achieved the most balanced confusion matrix, performing well in both "Arousal" and "Valence." While XGBoost and ANN also performed well, they were slightly less effective, demonstrating the value of HR in improving predictions.

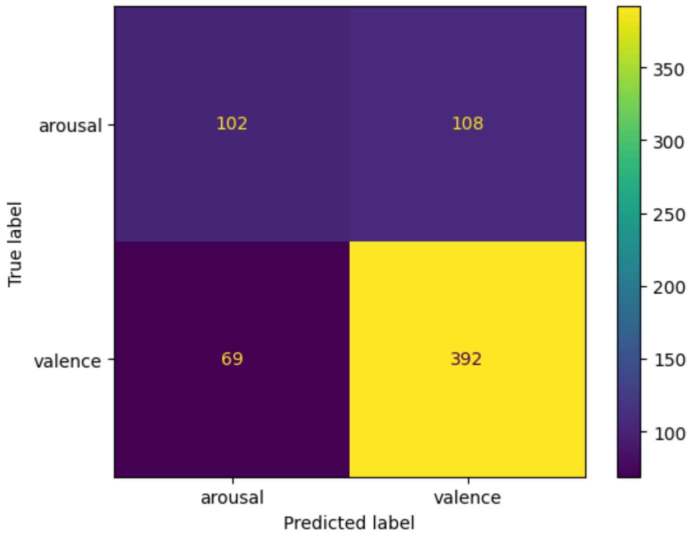

Figure 4.14: Random Forest with A-V

Chapter 5

Conclusion and Future Scope

5.1 Conclusion

This dissertation investigated the potential of using physiological signals—specifically Blood Volume Pulse (BVP), Electrodermal Activity (EDA), and Heart Rate (HR)—for emotion recognition. The study aimed to develop various machine learning models that are capable of accurately classifying human emotional states based on physiological data, which have been recognised to be important in applications that range from mental health monitoring to human-computer interaction.

Significance of Multi-Modal Physiological Data

The study demonstrated that combining multiple physiological signals improves the accuracy and reliability of emotion recognition systems. While BVP and EDA data only offered moderate accuracy, adding HR data significantly improved model performance. This finding highlights the importance of a multi-modal approach, in which the combination of various physiological indicators provides an improved and subtle understanding of an individual's emotional state.

Machine Learning Model Performance

The application of various machine learning models, including Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), and Artificial Neural Network (ANN) revealed their respective strengths and limitations in the context of emotion recognition. Traditional models such as RF and XGBoost performed well as baselines, especially in scenarios with relatively balanced data. Deep learning models such as MLP and ANN demonstrated greater potential for capturing the complex and nonlinear relationships inherent in physiological data despite being more prone to challenges such as class imbalance.

The study also highlighted the importance of hyperparameter tuning in optimising model performance. The models' capacity to accurately classify emotions was greatly improved after systematically exploring different parameter settings. This procedure was important in maximising the predictive power of these models and ensuring that they performed effectively across different circumstances.

Challenges with Class Imbalance

One of the most difficult challenges encountered during the study was class imbalance, which had a significant impact on the models' performance, particularly in accurately classifying minority emotional states. This problem got worse during the Leave-One-Subject-Out (LOSO) validation, a strict testing method in which each subject's data was used as a test set and the model was trained on the remaining data. LOSO validation is especially difficult in the context of imbalanced data because it can result in scenarios in which the model has limited exposure to particular classes during training, reducing its ability to generalise across subjects.

The use of techniques such as SMOTE (Synthetic Minority Over-sampling Technique), Upsampling assisted in solving some of these issues by balancing training data. However, LOSO validation revealed that, even with these techniques, the models struggled to predict less frequent emotional states. This highlighted the need for advanced methods to addressing class imbalance, particularly when dealing with personalised models that must generalise across different individuals.

Final thoughts

Finally, this dissertation has advanced the study of emotion recognition from physiological signals by exploring truly into the impact of various features, models, and training methods on classification accuracy. The findings highlight the importance of taking a multi-modal approach to capturing the complexities of human emotions as well as the need for ongoing research to address the challenges of class imbalance and real-time implementation. As the field develops, the findings from this study will provide a solid foundation for developing more accurate and ethically responsible emotion recognition systems.

5.2 Future Scope

The future scope of this research is wide, with major opportunities for further development and application in a variety of fields. As the field of emotion recognition from physiological signals advances, there are many areas for future research that could improve the current system's capabilities and expand its scope of application.

Advanced Machine Learning and Deep Learning Techniques

While machine learning models such as Random Forest and XGBoost were used in this study, there is lots of scope to investigate more advanced deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer models. These models can automatically learn and extract features from raw physiological data, possibly resulting in higher accuracy and robustness in emotion classification.

Incorporation of Additional Physiological Signals

This study focused on BVP, EDA, and HR signals; however, other physiological signals could be combined to provide a deeper understanding of emotional states. As discovered in the project, including a heart rate signal improved accuracy significantly. Similarly, respiration rate (RR) and body temperature may provide more context for emotion recognition. The addition of these signals may improve the system's accuracy and reliability.

Real-Time Emotion Recognition Systems

One of the most promising applications of this research is real-time emotion recognition systems, particularly those built into wearable devices or mobile applications. Creating lightweight, efficient models that can run in real time on devices with limited resources is an important challenge that must be addressed.

Addressing Data Imbalance and Enhancing Generalization

Data imbalance created a major challenge in this study, particularly with the Leave-One-Subject-Out (LOSO) validation method. Future research could look into more advanced methods for dealing with imbalanced datasets, such as advanced resampling methods, synthetic data generation using techniques like Generative Adversarial Networks (GANs), or Support vector machine (SVM) or its variants, such as V-SVM, which can handle class imbalance appropriately.

References

Agrawal, T. & Agrawal, T. (2021), 'Hyperparameter optimization using scikit-learn', *Hyperparameter optimization in machine learning: make your machine learning and deep learning models more efficient* pp. 31–51.

Aung, S. T., Hassan, M., Brady, M., Mannan, Z. I., Azam, S., Karim, A., Zaman, S. & Wongsawat, Y. (2022), 'Entropy-based emotion recognition from multichannel eeg signals using artificial neural network', *Computational intelligence and neuroscience* **2022**(1), 6000989.

Ayata, D., Yaslan, Y. & Kamasak, M. E. (2020), 'Emotion recognition from multimodal physiological signals for emotion aware healthcare systems', *Journal of Medical and Biological Engineering* **40**, 149–157.

Cai, Y., Li, X. & Li, J. (2023), 'Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review', *Sensors* **23**(5), 2455.

Cannon, W. B. (1927), 'The james-lange theory of emotions: A critical examination and an alternative theory', *The American journal of psychology* **39**(1/4), 106–124.

Dobrokhvalov, M. O. & Filatov, A. Y. (2024), Two-branch convolutional networks for stress detection from biomedical data, *in* '2024 XXVII International Conference on Soft Computing and Measurements (SCM)', pp. 449–452.

Du, Y., Crespo, R. G. & Martínez, O. S. (2023), 'Human emotion recognition for enhanced performance evaluation in e-learning', *Progress in Artificial Intelligence* **12**(2), 199–211.

Dzedzickis, A., Kaklauskas, A. & Bucinskas, V. (2020), 'Human emotion recognition: Review of sensors and methods', *Sensors* **20**(3), 592.

Fang, A., Zhong, P., Pan, F., Li, Y. & He, P. (2024), 'Impact of emotional states on tinnitus sound therapy efficacy based on ecg signals and emotion recognition model', *Journal of Neuroscience Methods* **409**, 110213.

Guo, R., Guo, H., Wang, L., Chen, M., Yang, D. & Li, B. (2024), 'Development and application of emotion recognition technology - a systematic literature review', *BMC Psychology* **12**(1), 95.
URL: *https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-024-01581-4*

Gupta, S., Kumar, P. & Tekchandani, R. K. (2023), 'Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models', *Multimedia Tools and Applications* **82**(8), 11365–11394.

Hofmann, J., Troiano, E., Sassenberg, K. & Klinger, R. (2020), 'Appraisal theories for emotion classification in text', *arXiv preprint arXiv:2003.14155* .

Htun, H. H., Biehl, M. & Petkov, N. (2023), 'Survey of feature selection and extraction techniques for stock market prediction', *Financial Innovation* **9**(1), 26.

Khan, M. U., Aziz, S., Hirachan, N., Joseph, C., Li, J. & Fernandez-Rojas, R. (2023), 'Experimental exploration of multilevel human pain assessment using blood volume pulse (bvp) signals', *Sensors* **23**(8).
URL: *https://www.mdpi.com/1424-8220/23/8/3980*

Khare, S. K., Blanes-Vidal, V., Nadimi, E. S. & Acharya, U. R. (2024), 'Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations', *Information Fusion* **102**, 102019.
URL: *https://www.sciencedirect.com/science/article/pii/S1566253523003354*

Kostev, R. S. (2023), Challenges and problems of the moscow method application in erp system implementation, *in* '2023 International Scientific Conference on Computer Science (COMSCI)', pp. 1–4.

Le, H.-D., Lee, G.-S., Kim, S.-H., Kim, S. & Yang, H.-J. (2023), 'Multi-label multimodal emotion recognition with transformer-based fusion and emotion-level representation learning', *IEEE Access* **11**, 14742–14751.

Lebaka, L. N., Govarthan, P. K., Rani, P., Ganapathy, N., Ronickom, A., Fredo, J. et al. (2023), Automated emotion recognition system using blood volume pulse and xgboost learning, *in* 'Healthcare transformation with informatics and artificial intelligence', IOS Press, pp. 52–55.

Liu, X., Shi, T., Zhou, G., Liu, M., Yin, Z., Yin, L. & Zheng, W. (2023), 'Emotion classification for short texts: an improved multi-label method', *Humanities and Social Sciences Communications* **10**(1), 1–9.

Lutin, E., Hashimoto, R., De Raedt, W. & Van Hoof, C. (2021), Feature extraction for stress detection in electrodermal activity., *in* 'BIOSIGNALS', Vienna, Austria, pp. 177–185.

Park, C. Y., Cha, N., Kang, S., Kim, A., Khandoker, A. H., Hadjileontiadis, L., Oh, A., Jeong, Y. & Lee, U. (2020), 'K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations', *Scientific Data* **7**(1), 293.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X. & Yang, X. (2018), 'A review of emotion recognition using physiological signals', *Sensors (Basel)* **18**(7), 2074.
URL: *https://www.mdpi.com/1424-8220/18/7/2074*

Singh, J., Saheer, L. B. & Faust, O. (2023), 'Speech emotion recognition using attention model', *International Journal of Environmental Research and Public Health* **20**(6), 5140.

Stark, L. & Hoey, J. (2021), The ethics of emotion in artificial intelligence systems, *in* 'Proceedings of the 2021 ACM conference on fairness, accountability, and transparency', pp. 782–793.

Veeranki, Y. R., Diaz, L. R. M., Swaminathan, R. & Posada-Quintero, H. F. (2024), 'Nonlinear signal processing methods for automatic emotion recognition using electrodermal activity', *IEEE Sensors Journal* **24**(6), 8079–8093.

Wang, Z., Ho, S.-B. & Cambria, E. (2020), 'A review of emotion sensing: categorization models and algorithms', *Multimedia Tools and Applications* **79**, 35553–35582.