

# Chapter 1

## Introduction

### 1.1 Emotion

Emotions are complex, involving both psychological and physiological elements, and are deeply connected to instinctive responses triggered by stimuli such as social interactions, environmental cues, and internal cognitive processes (Liu et al. 2023). These emotional states significantly impact our well-being and decision-making. Emotions may influence many aspects of our lives, such as the ability to pay attention, memory, achieve goals, communicate effectively, learn, stay in good spirits, and be motivated (Gannouni et al. 2021).

In today's fast-paced society, individuals face increasing emotional pressures from various sources, including academic demands, professional responsibilities, and family obligations. These pressures lead to prolonged stress, which has significant mental health consequences (M. M. Hossain and Gupta 2023). Chronic emotional imbalance can result in mental health issues, affecting mood and personality, separate from physical illnesses (M. M. Hossain and Gupta 2023). Long-term emotional struggles, tension, negative feelings, and perceptions can potentially trigger mental disorders such as anxiety, depression, bipolar disorder, schizophrenia, substance use disorders, and neurodevelopmental disorders.

The consequences of mental health challenges are widespread throughout society. Individuals suffering from mental health issues often experience significant personal pain and distress, which can hinder their ability to lead fulfilling lives. They may struggle with relationships, miss educational or work opportunities, and withdraw from social activities (Meghrajani et al. 2023). The prevalence of mental health issues is rising annually, largely due to difficulties in recognizing and managing negative emotions effectively.

Statistics indicate that 792 million people worldwide suffer from mental health disorders. Depressive disorders, affecting approximately 300 million people globally, are the third leading cause of death. Additionally, the World Health Organization recently reported that over 284 million people suffer from anxiety disorders (Uban, Chulvi, and Rosso 2021), (Veeranki, Ganapathy, et al. 2024). Hence, early intervention can prevent significant mental health problems, underscoring the importance of recognizing and addressing emotions promptly for overall well-being and mental health (Liu et al. 2023).

## **1.2 Emotion Paradigms**

Human emotions are a complex multidimensional experience, and the assessment of emotions is a highly subjective task. Emotions are measured and quantified through various models in the literature, which can be classified into two main types: categorical and dimensional emotion models (Nandi et al. 2022) (Zhang et al. 2020). Categorical emotion models categorize and define emotions using specific keywords associated with each emotion. For instance, Ekman's six-basic-emotion model classifies emotions as happiness, sadness, anger, fear, surprise, and disgust, while other emotions are seen as combinations of these basic ones (Zhang et al. 2020). Similarly, Plutchik's wheel of emotions organizes eight primary emotions into opposing pairs: joy vs. sadness, anger vs. fear, disgust vs. trust, and surprise vs. anticipation. These primary emotions vary in intensity, and intermediate emotions arise from blends of these primary ones (Yilmaz et al. 2021).

In contrast, dimensional models analyze emotions across multiple dimensions, providing a comprehensive representation with continuous values that capture subtle emotional

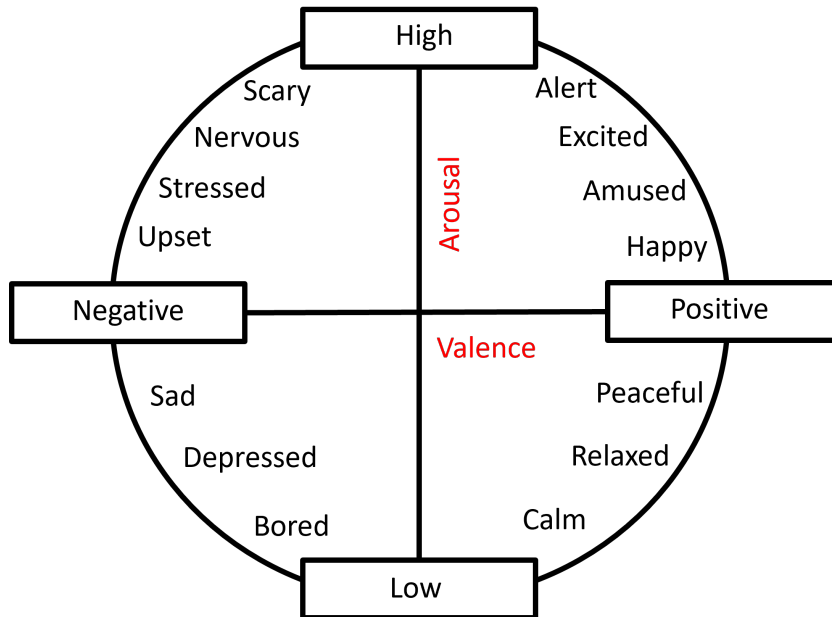


Figure 1.1: 2D Russel circumplex emotional model.

nuances. These models often depict human emotions in two-dimensional (2D) or three-dimensional (3D) structures as coordinated points within multi-dimensional space. Typically, emotions are portrayed through continuous or discrete levels, such as valence-arousal and valence-arousal-dominance. Valence signifies pleasantness, ranging from negative to positive, while arousal gauges excitement levels, from high to low (Ganapathy, Veeranki, H. Kumar, et al. 2021). Dominance evaluates an individual's control over their emotional states, indicating their capacity to manage emotional reactions. Given the challenge of quantifying dominance, the Valence-Arousal (VA) model, also known as Russell's circumplex model, emerges as a favored 2D emotion representation, focusing on valence and arousal as key aspects of emotional experience as shown in Figure 1.1 (Nandi et al. 2022).

### 1.3 Emotion Recognition

The methods for Emotion Recognition (ER) are primarily classified into trait-based and physiological state-based approaches. In the trait-based approach, researchers use human physical, behavioral, or non-physiological signals such as facial expressions, speech, gestures, posture, body movements, and self-reports (Ganapathy, Veeranki, and Swaminathan

2020). However, the reliability of these signals cannot be guaranteed, as individuals can easily control them to hide their true emotions, especially in social situations. The other category involves utilizing physiological signals such as Electroencephalogram (EEG), Electrodermal Activity (EDA), Electrocardiogram (ECG), Electromyogram (EMG), Skin Temperature (SKT), Blood Volume Pulse (BVP), and Respiration (RSP) (Ganapathy and Swaminathan 2019), (Sriram Kumar et al. 2023). The human nervous system encompasses the Central Nervous System (CNS) and the Peripheral Nervous System (PNS). The PNS comprises the Autonomic Nervous System (ANS) and the somatic nervous system. The ANS consists of sensory and motor neurons and regulates communication between the CNS and internal organs. Physiological signals such as EEG, EDA, ECG, EMG, SKT, BVP, and RSP respond to specific situations, reflecting CNS and ANS activity changes, consistent with Cannon's theory (Cannon 1927). Utilizing physiological signals offers an advantage due to their involuntary activation within the CNS and ANS, making them difficult to control consciously (Shu, Xie, et al. 2018).

Despite their utility, acquiring EEG and EMG signals can be complex, requiring specialized instrumentation and electrode systems. ECG signals are prone to interference from motion artifacts. At the same time, SKT measurements can be influenced by environmental factors such as ambient temperature, whereas BVP uses infrared light that requires ambient light cancellation to produce accurate measurements. Similarly, RSP measurements may be affected by factors like body position and breathing patterns, introducing variability into the results (Chunawale and Bedekar 2020; Sriram Kumar et al. 2023). In contrast, EDA stands out for its non-invasiveness, reliability, affordability, and non-intrusive nature, providing valuable insights for evaluating emotional states. EDA, measured by observing changes in skin conductivity, reflects the activity of eccrine sweat glands predominantly regulated by the sympathetic branch of the autonomic nervous system via sudomotor nerves. The activation of sudomotor nerves induces sweat production, altering the skin's electrical properties due to sweat secretion and variations in the ionic permeability of sweat gland membranes (Ganapathy and Swaminathan 2019; Greco, Valenza, Lanata, et al. 2015). These changes in skin conductivity, measured as EDA, serve as vital

markers for assessing emotional states. The simplicity of EDA and its ability to capture subtle emotional responses make it particularly attractive in emotion research, providing real-time data in naturalistic environments. Therefore, despite the challenges presented by other physiological signals, EDA remains a reliable and accessible method for studying human emotions(Gannouni et al. 2021).

## **1.4 Electrodermal activity**

### **1.4.1 Background**

The study of psychological effects on electrical changes in human skin began over 100 years ago. Féré discovered that passing a small electrical current across two electrodes on the skin could measure momentary decreases in skin resistance in response to various stimuli (Fere 1888). Tarchanoff reported changes in electrical potential between two electrodes on the skin could be measured without applying external current (Tarchanoff 1890). Féré and Tarchanoff are credited with discovering the two basic methods of recording EDA: the exosomatic and endosomatic methods.

### **1.4.2 Skin Anatomy**

The skin acts as a dynamic barrier, preventing foreign substances from entering the body while facilitating the exchange of materials, maintaining water balance, and regulating body temperature through processes like vasoconstriction/dilation and sweat production. Figure 1.2 shows the anatomy of human skin, and it contains two types of sweat glands: apocrine and eccrine, with eccrine glands particularly interesting to psychophysiology researchers. Eccrine glands on the palms and soles respond more to psychological stimuli than changes in temperature due to their higher density and specific location. The EDA measurement focuses on psychologically induced sweat gland activity, involving peripheral mechanisms within the skin layers, especially the eccrine sweat gland. Sweat ducts serve as variable resistors, with sweat rising in varying amounts and numbers of sweat glands, depending on sympathetic nervous system activation. A more conductive path

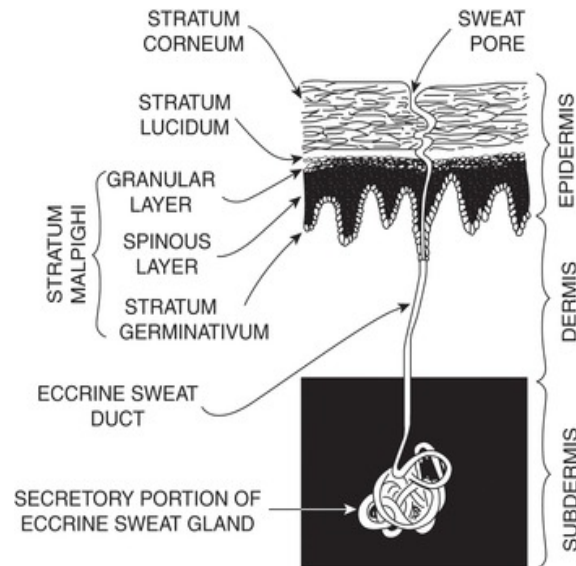


Figure 1.2: Anatomy of the eccrine sweat gland in various layers of skin.

forms through the resistant corneum as sweat fills the ducts. Higher sweat levels result in lower resistance in the variable resistor. Changes in sweat level alter resistor values, causing observable EDA changes (Boucsein and Boucsein 2012). This relationship enables observation of changes corresponding to sweat level variations, offering insights into physiological responses to psychological stimuli.

### 1.4.3 Electrodermal activity

EDA measures changes in skin conductance, reflecting responses to brain signals and offering insights into physiological reactions to various stimuli, especially those with psychological significance. The regulation of eccrine sweat glands, essential for EDA measurement, involves the autonomic nervous system's sympathetic and parasympathetic divisions. While acetylcholine, typically linked to parasympathetic functions, mediates eccrine gland activity, norepinephrine, associated with sympathetic activation, has a lesser direct impact. Human sweat glands are primarily innervated by sympathetic cholinergic fibers, though adrenergic fibers are also present nearby. Boucsein 1999 proposed pathways for skin conductance responses (SCRs) highlight contralateral influences from the cortex and basal ganglia, ipsilateral influences from the hypothalamus and limbic system, and mechanisms within the brainstem's reticular formation. Neuroimaging studies re-

veal consistent brain activation patterns associated with EDA, involving regions like the ventromedial prefrontal cortex, amygdala, and orbitofrontal cortex. The hypothalamus, crucial for regulating thermoregulatory sweating, integrates sympathetic activity patterns with emotional responses and limbic structures (Boucsein 1999).

#### **1.4.4 EDA Recording**

EDA is assessed using Ohm's law principles, passing a small electric current through skin surface electrodes. Ohm's law states that skin resistance  $R$  equals the voltage  $V$  divided by the current  $I$ , expressed as  $R = V/I$ . By maintaining a constant current, variations in voltage between the electrodes are measured, directly reflecting skin resistance. Conversely, if voltage remains constant, changes in current flow correspond to the reciprocal of skin resistance, termed skin conductance. Skin conductance is quantified in Siemens, with EDA readings commonly reported in microSiemens ( $\mu S$ ). Skin conductance measurement involves placing two electrodes on active skin sites for bipolar recording. Typically, EDA signals are recorded from areas like the palms of the hands, with common positions including the thenar eminences or the volar surface of the fingers as shown in Figure 1.3. The distal phalange site is preferred unless specific constraints exist, such as small fingertips in children or issues like cuts or heavy calluses (Freedman et al. 1994). Recording EDA on the non-dominant hand is preferable due to reduced artifacts and greater stability compared to the dominant hand. With less movement variability and fewer precise movements, non-dominant hand recordings yield cleaner data and minimize the risk of distortions. The ipsilateral limbic system primarily controls EDA during emotional tasks, while during non-emotional tasks, the contralateral system may have a greater influence (Boucsein and Boucsein 2012). In Figure 1.3, placement 1 involves volar surfaces on medial phalanges, placement 2 involves volar surfaces of distal phalanges, and Placement 3 involves the thenar and hypothenar eminences of palms.

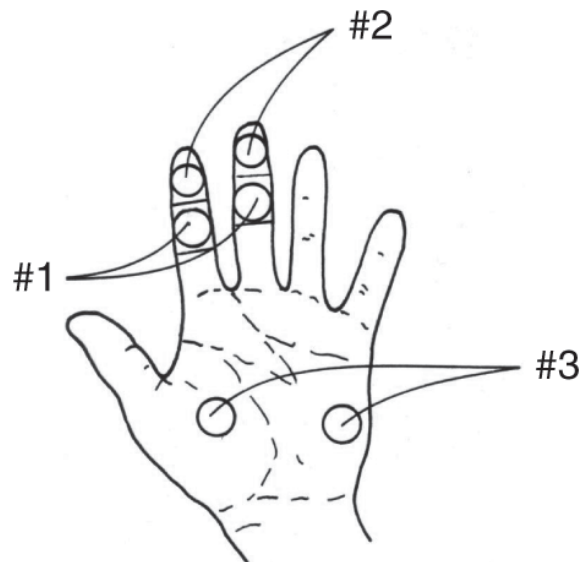


Figure 1.3: Electrode placements for recording EDA.

### 1.4.5 EDA Components

The raw EDA comprises two fundamental components: tonic and phasic, as shown in Figure 1.4 (Alexander et al. 2005). The tonic signal, known as skin conductance level, reflects slowly varying activity, forming a stable baseline unique to each individual. This baseline varies significantly among individuals and is influenced by psychological states, physical conditions, and autonomic regulation. In contrast, the phasic signal, or Skin Conductance Response (SCR), refers to rapid changes in the EDA signal, representing sudden increases in skin conductance triggered by environmental stimuli. These increases typically occur within 1 to 5 seconds following a stimulus presentation, such as a video, image, or audio. The phasic component is closely linked to cognitive processes like decision-making and reasoning, typically occurring in response to short-term events. These rapid changes reflect the brain's engagement in these cognitive activities, indicating heightened attention and processing related to the immediate environment. SCRs may be called peaks or spikes if they exceed a specific threshold. These peaks can be event-related, attributed to a specific eliciting stimulus or task, or non-specific, occurring without an identifiable stimulus (Boucsein and Boucsein 2012; Horvers et al. 2021).

## 1.5 Decomposition of EDA: Tonic and Phasic

Decomposition, or deconvolution, of EDA involves breaking down the EDA signal into its components, allowing for a more thorough and detailed analysis of the physiological response. Researchers have developed various decomposition methods, often in the form of mathematical models or software packages, to separate tonic and phasic signals. These methods include Continuous Decomposition Analysis (CDA), Discrete Decomposition Analysis (DDA), Dynamic Causal Modeling (DCM), Convex Optimization based on EDA (cvxEDA), Compressed Sensing, Nonnegative Sparse Deconvolution (SparsEDA), and the Poral Valve Model (BayesianEDA).

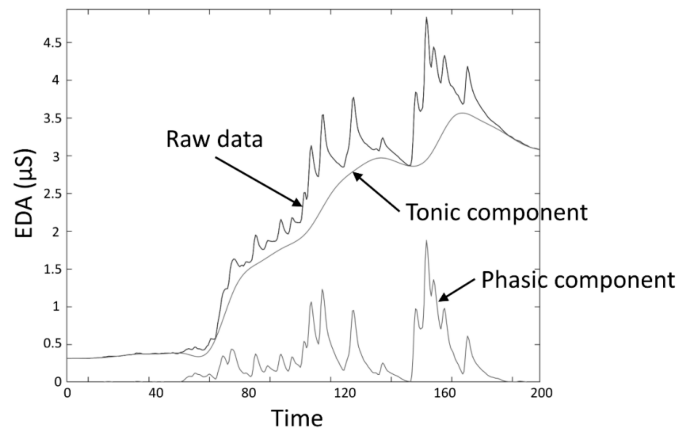


Figure 1.4: Raw EDA and tonic and phasic components of EDA signal.

Benedek et al. (2010) proposed two methods: CDA using a deconvolution approach and DDA using a nonnegative deconvolution approach (Benedek and Kaernbach 2010a; Benedek and Kaernbach 2010b). These methods are based on the Edelberg model (Edelberg, 1993) and provide a rational explanation for estimating the shape of the phasic response. Later, Bach et al. 2011 developed DCM to characterize the EDA with underlying sudomotor nerve activity using a linear time-invariant model. In another study, Alberto Greco proposed the cvxEDA method, which estimates the phasic and tonic signals using the Bateman function and cubic B-spline basis function, respectively. This method found that the features extracted from the estimated phasic signal are good predictors for ER (Greco, Valenza, Lanata, et al. 2015). Hernando developed a fully automated nonnegative sparse deconvolution method (Hernando-Gallego, Luengo, and Artés-Rodríguez 2017).

Jain et al. 2016 proposed a compressed sensing-based decomposition approach to extract SCR events from the EDA signal. In a recent study, Amin et al. (2022) introduced the BayesianEDA method, which relies on the functioning of the autonomic system's activation to induce changes in skin sweating (R. Amin and Faghieh 2022).

Analyzing entire EDA signals can be computationally expensive, often requiring significant computational resources. Identifying optimal segments is critical for optimizing resource usage and enhancing the efficiency of emotion detection models in real-time applications. Selecting an optimal segment allows for the effective capture of temporal dynamics associated with physiological changes across various emotional states, ensuring the accuracy and reliability of emotion detection systems. A key factor in the success of an emotion-based approach is the careful selection of an appropriate segment or window size. If a window is too short, we might miss out on important subtle changes in physiological responses that indicate different emotions. On the other hand, if the window is too long, we might end up averaging out or smoothing over dynamic fluctuations, losing the finer details of emotional shifts (W. Yang et al. 2018). Therefore, meticulous calibration of segment size is imperative for the robustness of emotion detection algorithms.

## **1.6 Importance of Feature Extraction in Emotion Recognition**

Assessing EDA signals for ER in three major domains, namely time, frequency, and time-frequency domains, can provide greater insight into the underlying physiological process. Time domain features, encompassing amplitude variations at different instances, statistical metrics, event-related attributes, higher-order crossings, and Hjorth features, have shown effectiveness in EDA emotion recognition (Shukla et al. 2019). However, these features primarily capture changes within a signal over time. In contrast, frequency domain features, such as sum of spectral power, mean of spectral components, median of spectral components, frequency non-specificity of SCR, and fast Fourier transform for

bandwidths F1 (0.1, 0.2), F2 (0.2, 0.3), and F3 (0.3, 0.4), along with frequency bands within ranges [0.02–0.25 Hz], [0.25–0.4 Hz], and [0.4–1 Hz], are utilized to reveal information about the presence and magnitude of various frequencies within a signal (Bianco and Napoletano 2019), (Sánchez-Reolid, Martínez-Rodrigo, et al. 2020). Nonetheless, they may not adequately capture rapid changes or temporal aspects of events.

Time-frequency domain features are employed to address these limitations of the above two representations by capturing rapid changes over time within signals that may not be fully captured or illustrated solely in either the time or frequency domains (Rao Veeranki, Ganapathy, and Swaminathan 2021). Various time-frequency spectrogram techniques, including the short-time Fourier transform (STFT), continuous wavelet transform (CWT), Wigner-Ville distribution (WVD), Choi-Williams distribution (CWD), and Mel-frequency cepstrum (MFC) have been proposed in literature and utilized to gain a comprehensive understanding of the time-frequency characteristics exhibited by signals. The CWT faces challenges in choosing the right mother wavelet, computational demands, resolution adaptability, and interpreting complex wavelet coefficients, while the WVD and CWD contend with issues like cross-term interference, sensitivity to noise, computational complexity, and susceptibility to edge effects (Ingle, Rakshe, Ronickom, et al. 2023). In contrast, STFT helps in analyzing the frequency content over short time intervals, allowing us to capture temporal variations in the EDA signals (Ganapathy and Swaminathan 2019; Elbir et al. 2018). Further, the MFC helps in analyzing EDA signals by capturing frequency information over time using a Mel-scale. This enables a concise representation, facilitating enhanced analysis and pattern recognition in EDA data (Benito-Gorron et al. 2019). In this thesis, STFT and MFC are used to obtain spectrograms of phasic EDA signals.

In addition to time-frequency methods, diverse signal-to-image conversion techniques are employed. The signal-to-image module, Brain2Images, utilizes a convolutional neural network for conversion (Kavasidis et al. 2017). Signal-to-image mapping involves forming 2D images by computing correlations (W. Li et al. 2022), and graph networks utilize

matrices such as adjacency and Laplacian (Xue, Adriaenssens, and Mao 2023). However, some of these methods may lack preservation of signal space characteristics or temporal information, necessitating the integration of novel techniques (Patel and Yildirim 2023). Advanced algorithms such as the Gramian Angular Field (GAF), Markov Transition Field (MTF), and Recurrence Plot (RP) address limitations in existing methods. GAF preserves temporal dynamics, MTF encodes intricate signal details through transition probabilities, and RP visualizes recurrence patterns. The hand-crafted features computed from the time-encoded images can be crucial for distinguishing between different classes or emotional states in a signal. Notably, the Gray-Level Co-occurrence Matrix (GLCM) and Gray-Level Run-Length Matrix (GLRLM) have emerged as widely used techniques for quantifying image texture information. GLCM focuses on capturing spatial relationships among pixels sharing the same gray level. At the same time, GLRLM describes the lengths and orientations of pixel runs with identical gray levels (Öztürk and Akdemir 2018; Haralick, Shanmugam, and Dinstein 1973; Shabu and Jayakumar 2020; Ronicko et al. 2020). Fractal Dimension Texture Analysis (FDTA) provides another avenue for evaluating the complexity and irregularity of patterns within an image (Tsiaparas et al. 2010). Other techniques exploring 2D representation-based feature extraction have gained prominence alongside these methods. Zernike’s Moments (ZM) and Hu’s Moments (HM) are two widely adopted techniques for image analysis, capturing global and local shape information, respectively (Shiyi Chen et al. 2023; Anandhalli, Tanuja, and Baligar 2022). Additionally, First Order Statistics (FOS), which encompass mean, variance, skewness, and kurtosis, have found application as features in recognizing emotions from physiological signals (Ahmed, Bari, and Gavrilova 2019).

## **1.7 Machine Learning in Emotion Recognition**

Numerous studies have explored the use of Machine Learning (ML) algorithms in the realm of ER using EDA. These studies have utilized various classifiers, including linear, nonlinear, ensemble, and deep learning models, to distinguish and categorize different emotional states (Tronstad et al. 2022; Sánchez-Reolid, López de la Rosa, et al. 2022).

Linear classifiers, such as Naive Bayes, linear discriminant analysis, linear support vector classifiers, and Logistic Regression (LR), have been utilized in these investigations. Similarly, nonlinear classifiers, including decision trees, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), have been explored to capture the nuanced patterns within EDA data. Additionally, ensemble classification algorithms, such as AdaBoost, Extreme Gradient Boosting (XGB), and Random Forest (RF), have demonstrated efficacy in this context (Rao Veeranki, Ganapathy, and Swaminathan 2021; Burns, Adeli, and Buford 2020; Ellis et al. 2017; Sriram Kumar et al. 2023). Deep learning models, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have also made notable contributions to emotion detection using EDA signals (Ganapathy and Swaminathan 2019; Hassanpour et al. 2019; Moradi et al. 2020). However, acknowledging the drawbacks of deep learning models, challenges encompass their sensitivity to high-resolution images, the requirement for substantial datasets, and the involvement of numerous parameters necessitating meticulous adjustment during training. A study highlighted the significance of providing texture patterns extracted from spectrograms of signals to ML algorithms instead of directly supplying spectrograms to CNN for optimal performance (Costa, Oliveira, and Silla Jr 2017). In this thesis, we used non-linear and two ensemble classifiers, including boosting and bagging algorithms such as SVM, RF, and XGB, to build an emotion recognition system. SVM utilizes a hyperplane to effectively separate data into distinct classes by maximizing margins in multidimensional spaces. RF employs an ensemble of decision trees to enhance performance and prevent overfitting, while XGB, a well-distributed gradient-boosting library, offers high efficiency and flexibility in classification tasks. This collective use of diverse algorithms allows for a thorough exploration of model capabilities, ensuring a balanced assessment and potentially improving the accuracy of emotion recognition based on phasic EDA features.

**Chapter 2** presents the literature review on emotion assessment studies, the significance of EDA for emotion assessment, time-frequency representation, and ML techniques. Based on this, the motivation for the thesis is derived.

**Chapter 3** outlines the dataset and methodology utilized in this study, along with the approach employed for feature extraction and ML models for the classification of emotional states.

**Chapter 4** elaborates on the outcomes of successive stages in-depth as a selection of optimal decomposition methods for four emotional states: amusing, boring, relaxing, and scary. Additionally, it explores the optimal selection of EDA segment signals for emotional classification and the selection of the windowing approach for the latter part of the EDA signal. The chapter also presents the results of the strategies employed for efficient classification.

**Chapter 5** discusses each stage analysis based on the results obtained in Chapter 4.

**Chapter 6** explores limitations and potential avenues for future research.

**Chapter 7** outlines the conclusion of this study. The references section contains citations to relevant research publications.