Analysis of nonverbal behaviors in context of video games

Majid Mohammadi

Department of Computer Science, Bioengineering, Robotics and System Engineering (DIBRIS)

University of Genova

Supervisor

Radoslaw Niewiadomski, co-supervisor Irem Arici

In partial fulfillment of the requirements for the degree of

Master of Computer Engineering

March 23, 2026

Abstract

This thesis addresses the problem of analyzing and modeling spontaneous emotional reactions of professional eSports players during real competitive football matches. While most existing studies in affective computing rely on laboratory-controlled and manually annotated datasets, such settings often fail to capture natural, high-intensity emotional responses occurring in real-world scenarios.

To overcome this limitation, this research focuses on official competitive matches of the football video game eFootball, where players' facial reactions are recorded simultaneously with in-game events. A dataset was constructed by extracting short reaction clips corresponding to specific in-game events such as goals, penalties, fouls, and offsides. Facial Action Units (AUs) were automatically extracted using OpenFace 3, and a set of high-level statistical and temporal features—including mean intensity, maximum intensity, standard deviation, slope, and duration—were computed for each clip.

Instead of relying on manual emotion annotation, this work employs unsupervised clustering methods to discover latent behavioral patterns in the data. Eight clustering algorithms were evaluated using multiple internal validation metrics to identify the most meaningful grouping structures. The resulting clusters were further validated through supervised classification models, demonstrating strong reproducibility and separability.

The proposed framework enhances the state of the art by introducing an ecologically valid, label-free methodology for emotion analysis in real competitive environments. It reduces dependency on costly manual annotation and provides a scalable approach for analyzing spontaneous emotional behavior in eSports and other dynamic multimedia contexts.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Summary

This chapter introduces the research problem, its scientific relevance, and the primary objectives of the thesis. It begins by examining the limitations of laboratory-based emotion recognition studies and highlights the potential of competitive game environments as a source of ecologically valid emotional data. The challenges associated with analyzing real-world data—particularly from professional eSports tournaments—are subsequently discussed. The chapter then presents the research objectives, key innovations, research questions, conceptual definitions, scope, and underlying assumptions. Collectively, this chapter establishes the theoretical and conceptual foundation for the methodological framework and empirical analyses presented in the following chapters.

1.1 Problem Statement and Importance

In recent years, affective computing has experienced substantial growth, aiming to analyze and recognize human emotions through nonverbal cues such as facial expressions, body movements, and vocal characteristics. Nevertheless, a significant proportion of existing research in this field relies on data collected under controlled laboratory conditions. In such environments, emotions are typically elicited using predefined stimuli within fixed time intervals. While these settings ensure high experimental control, they often fail to capture the complexity, spontaneity, and ecological validity of emotional reactions in real-life contexts.

In contrast, competitive digital game environments provide synchronized access to gameplay footage and players' facial reactions. This simultaneity enables the contextual analysis of nonverbal behaviors in direct relation to specific in-game events. Among various game genres, professional football simulation games represent a particularly suitable domain. As one of the official disciplines of electronic sports (eSports), competitive football games combine global popularity with high emotional intensity. The high-stakes and performance-driven nature of such tournaments increases the likelihood of observing authentic and spontaneous emotional reactions.

However, selecting appropriate data sources in this context is not trivial. On public streaming platforms such as Twitch, many individuals broadcast their gameplay. Yet, their emotional reactions are often influenced by external factors unrelated to the game itself, including audience interaction, live chat engagement, and performative behavior. This contextual noise can lead to exaggerated or theatrical expressions, reducing the reliability of the data for studying genuine emotional responses. Furthermore, in casual or non-competitive gaming contexts, emotional intensity may be attenuated due to habituation or reduced performance pressure.

In contrast, official professional eSports tournaments with substantial prize pools create high-pressure competitive environments in which in-game events such as goals, offsides, penalties, and other critical moments evoke genuine, intense, and distinct emotional reactions. This characteristic makes such competitions a valuable and reliable source for studying natural emotional expressions.

1.2 Research Problem, Importance, and Questions

1.2.1 Main Research Problem

The central research problem addressed in this thesis can be formulated as follows:

> Is it possible to discover distinct and meaningful behavioral patterns by applying clustering techniques to statistical and dynamic features extracted from facial Action Units during specific competitive eFootball events, without relying on manual labeling, and subsequently utilize this discovered structure to develop baseline emotion classification models?

This problem emerges from the fundamental challenge in affective computing: the scarcity of naturally occurring, well-annotated emotional data. While laboratory settings offer controlled conditions, they often fail to capture the authenticity and intensity of emotions experienced in real competitive environments. Conversely, natural data presents significant technical obstacles that must be overcome to enable systematic analysis.

1.2.2 Importance and Necessity of the Research

The significance of this research can be examined from several complementary perspectives:

1. **Enhancing ecological validity:** Studying emotional reactions in real professional eSports competitions provides a more natural representation of human emotions compared to laboratory-induced settings. The spontaneous and authentic nature of these reactions offers insights that controlled experiments cannot replicate.

2. **High emotional intensity context:** Competitive tournaments with financial rewards and performance pressure create conditions that encourage authentic and intense emotional expression. Unlike casual gameplay or staged recordings, professional matches generate genuine emotional responses of significant magnitude.

3. **Reducing reliance on manual labeling:** Using clustering to uncover latent structures and generate pseudo-labels contributes toward scalable and automated emotion analysis. This addresses one of the most persistent bottlenecks in affective computing research—the time-consuming and subjective process of manual annotation.

4. **Addressing real-world constraints:** The study operates under realistic technical limitations such as low image resolution, small facial regions, and motion noise. By confronting these challenges directly, the methodology developed here is better positioned for deployment in real-world applications beyond laboratory settings.

5. **Comprehensive algorithm comparison:** Evaluating eight diverse clustering algorithms on the same emotional dataset provides valuable insights into the relative strengths and weaknesses of different approaches, serving as a reference for future researchers in the field.

6. **Practical implications:** The findings may inform adaptive game design,

user experience analysis, competitive performance modeling, and psychological assessment of professional players. Game developers, eSports coaches, and sports psychologists could potentially benefit from automated tools for emotion analysis.

1.2.3 Research Questions

Based on the research problem and the identified gaps in the literature, this study seeks to answer the following main question:

Can expressive patterns of eSports players be discovered and empirically validated through the integration of unsupervised clustering and using facial Action Unit data?

1.3 Research Objectives

1.3.1 Main Objective

The primary objective of this research is to design and implement a comprehensive analytical framework capable of discovering and validating emotional patterns of players in competitive computer games through the integration of unsupervised clustering and supervised classification of facial Action Units. The framework aims to identify spontaneous emotional reactions to in-game events without the need for costly manual annotation.

1.3.2 Work plan

1. Collection and preparation of natural emotional data from 37 official professional eFootball matches.

2. Extraction of facial Action Units using OpenFace 3.

3. Transformation of Action Unit time series into five statistical features: Mean, Maximum, Standard Deviation, Slope, and Duration.

4. Application and comparison of eight clustering algorithms: K-means, DBSCAN, Gaussian Mixture Model, BIRCH, Affinity Propagation, Mean-Shift, Spectral Clustering, and Hierarchical Clustering.

5. Evaluation of clustering performance using Silhouette Coefficient, Calinski-Harabasz Index, and Davies-Bouldin Index.

6. Interpretation of behavioral profiles corresponding to each discovered cluster.

1.4 Research Contribution

1.4.1 Contribution 1: Focus on Professional Competitive Data

Unlike many previous studies that rely on laboratory-induced emotions or casual streamers, this research focuses exclusively on professional eSports tournaments where emotional reactions are spontaneous, intense, and performance-driven.

1.4.2 Contribution 2: Comprehensive Comparison of Eight Clustering Algorithms

Eight distinct clustering algorithms are systematically evaluated on real-world emotional data, providing methodological guidance for future research.

Chapter 2

Theoretical Foundations and Related Works

2.1 Introduction

In this chapter, the theoretical foundations and background of research related to the topic of the thesis are reviewed. As stated in the first chapter, the aim of this research is to analyze the emotional reactions of players in professional computer soccer matches using facial Action Units and then cluster these reactions to discover hidden emotional patterns. To achieve this goal, it is necessary to first gain a deep understanding of the nature of emotion, the theories in this field, its measurement methods, and previous studies in the field of emotion analysis in computer games and game streams.

Chapter 2 consists of several main sections. First, we address the theoretical foundations of emotion and review the main theories such as Ekman's Basic Emotions Theory, Russell's Core Affect Theory, and Appraisal Theory. Cognitive appraisal theory is particularly important because it explains how an event (such as scoring a goal in a match) can elicit different emotional responses in different individuals [1; 2]. Next, we introduce the Facial Action Coding System (FACS) and Action Units, which form the basis of the OpenFace tool [3].

We then review research on emotion analysis in computer games and game streams. Several studies have shown that multimodal analysis (including face, voice, and text) can increase the accuracy of emotion recognition [4; 5]. Field studies in real-world competitive environments, such as Hearthstone tournaments [6] or horror game streams [7], have also shown that moment-by-moment emotion analysis using face and voice is possible.

Another important part of this chapter is a review of the challenges of collecting data in natural environments (in-the-wild). Unlike laboratory data collected under controlled conditions, natural data comes with challenges such as noise, small face size, and unwanted movements [8]. These challenges are also fully experienced in the present study.

Finally, after reviewing the clustering and classification algorithms used in this study, the position of the present study among previous works is determined.

2.2 Theoretical Foundations of Emotion

2.2.1 Definition of Emotion and Its Components

Emotion is a complex and multidimensional phenomenon for which numerous definitions have been provided. In general, emotion can be considered a reaction to events important to the individual, which includes a set of changes in various systems of the organism [9].

According to componential theories, emotion is composed of:

- Cognitive Component
- Physiological Component

- Motivational Component
- Expressive Component
- Subjective Experience

In this study, the main focus is on the expressive component, specifically facial muscle movements (Action Units), extracted using OpenFace 3.

2.2.2 Main Emotion Theories

Paul Ekman introduced six basic emotions: Happiness, Sadness, Fear, Anger, Surprise, and Disgust [10]. Each emotion corresponds to a specific facial muscle activation pattern defined in FACS. Although widely used, this theory has been criticized for oversimplifying emotional complexity [11].

James Russell proposed the Circumplex Model of Affect [11]. Emotions are defined along two dimensions: Valence, Arousal- Instruments such as the Self-Assessment Manikin (SAM) measure these dimensions [12].

Cognitive appraisal theory emphasizes that emotions arise from the individual's evaluation of events [1].

Main appraisal variables:

Appraisal Variable	Explanation
Goal Relevance	Relevance to individual goals
Goal Congruence	Helps or hinders goals
Certainty	Confidence of occurrence
Agency	Who caused the event
Coping Potential	Ability to control event

2.3 Facial Action Coding System (FACS)

FACS was developed by Ekman and Friesen in the 1970s [3]. It provides a systematic description of facial muscle movements.

Each AU represents activation of specific facial muscles. Action Units are the smallest visible components of facial muscle movement, as defined in the Facial Action Coding System (FACS). Each Action Unit represents the activation of one or more specific facial muscles. For example:

- **AU6 and AU12**: When a person smiles, the corners of the lips are pulled upwards (AU12) and the muscles around the eyes are contracted (AU6). These two Action Units combine to indicate genuine happiness and joy. As the intensity and duration of the smile increases, the intensity of the emotion changes from normal happiness to intense joy and excitement.
- **AU4**: When the eyebrows are drawn in and down, this indicates sadness, anger, or intense concentration. For example, a player who scores a goal may activate this Action Unit in his face.

AU	Facial Area	Description	Related Emotion
AU1	Forehead	Inner brow raise	Surprise
AU2	Forehead	Outer brow raise	Surprise
AU4	Eyebrow	Brow lower	Anger/Sadness
AU6	Eye	Cheek raise	Duchenne smile
AU12	Mouth	Lip corner pull	Happiness
AU15	Mouth	Lip corner depress	Sadness
AU20	Mouth	Lip stretch	Fear
AU25	Mouth	Lips part	Surprise

Table 2.1: Selected Facial Action Units and Their Associated Emotions

2.4 Challenges of Natural Data Collection

One of the most important challenges in the field of affective computing is the collection of reliable and valid data for training emotion recognition models. As discussed in previous sections, most existing datasets have been collected under laboratory conditions with artificially induced emotions. However, recent studies have shown that physiological and behavioral responses in laboratory environments differ significantly from natural reactions in real-world settings.

The review paper titled *"Toward Emotion Recognition From Physiological Signals in the Wild"* by Niewiadomski et al. [13], published in *Frontiers in Psychology*, provides a comprehensive analysis of the challenges and advantages of collecting emotional data in natural (in-the-wild) environments. Although the primary focus of this paper is on physiological signals, it offers valuable insights for the present study, as similar challenges exist in collecting any type of emotional data in real-world conditions.

2.4.1 Differences Between Laboratory and Natural Data

Niewiadomski et al. [13] demonstrate that emotional data collected in laboratory environments differs fundamentally from data collected in natural settings. Laboratory environments provide high control over experimental conditions, whereas real-world environments are inherently uncontrolled. While laboratory data typically benefits from high-quality recordings and precise labeling, it often lacks naturalness. In contrast, natural (in-the-wild) data reflects genuine emotional responses but suffers from variability and noise.

These differences can be summarized as follows:

For example, Wilhelm and Grossman (2010)[14] showed that heart rate during a real football match (natural environment) is significantly higher than when simi-

Table 2.2: Comparison between Laboratory and Natural Emotional Data

Characteristic	**Laboratory**	**Natural (In-the-Wild)**
Control	High	Low
Image Quality	High	Variable
Naturalness	Low	High
Labeling	Easy	Difficult

lar emotions are induced in a laboratory setting. Similarly, Zhou et al. (2017)[15] found that physiological signals such as EDA, ECG, and EMG differ substantially between real and laboratory conditions. Models trained on laboratory data achieved low accuracy (17% to 45%) when applied to real-world emotion recognition tasks.

These findings directly support the approach of the present study, which focuses on professional eSports competitions instead of laboratory data or casual streaming environments. In such competitions, players experience genuine and intense emotions driven by competitive pressure and high-stakes rewards.

2.4.2 Advantages of In-the-Wild Data Collection

Niewiadomski et al. [13] identify several key advantages of collecting emotional data in natural environments, which align with the objectives of this research:

- **Higher Ecological Validity:** Data collected in real-world environments more accurately reflects genuine emotional responses.
- **Reduced Ethical Constraints:** Natural environments eliminate the need for artificial emotion induction, avoiding ethical concerns.
- **Continuous Learning Capability:** Models can be improved progressively as more real-world data becomes available.

2.4.3 Challenges of In-the-Wild Data Collection

Despite its advantages, collecting emotional data in natural environments introduces several significant challenges. Based on the literature and observations in this study, the main challenges include:

1. **Small Face Region:** In many real-world videos, especially broadcasted eSports matches, the player's face appears as a small region on the screen, making accurate feature extraction difficult.

2. **Motion Noise:** Sudden head movements, posture changes, and occlusions (e.g., hands covering the face) introduce noise and missing data.

3. **Audio Noise:** Background sounds, commentary, and environmental noise can interfere with multimodal analysis and indirectly affect emotion interpretation.

4. **Heterogeneity of Responses:** Emotional reactions vary significantly between individuals in terms of intensity, duration, and timing, making pattern extraction more complex.

5. **Lack of Accurate Labels:** Unlike laboratory datasets, natural data lacks precise ground-truth labels, making supervised learning difficult.

One of the most critical challenges is the labeling problem. Niewiadomski et al. [13] discuss two main labeling approaches:

- **Self-report:** Participants report their own emotions, which can be affected by memory limitations and bias.

- **Expert labeling:** Requires manual annotation and may introduce subjectivity and privacy concerns.

Proposed Solution in This Study: To address the labeling challenge, this research adopts a hybrid approach combining unsupervised clustering and supervised classification. Clustering is used to discover hidden emotional patterns without manual labeling, and classification is then applied to validate these patterns. This approach reduces dependency on costly labeling processes while minimizing human bias.

2.4.4 GARAFED Evaluation Framework

One of the key contributions of Niewiadomski et al. [13] is the introduction of the GARAFED (Graphical Assessment of Real-life Application-Focused Emotional Dataset) framework. This framework evaluates emotional datasets based on their suitability for real-world applications. The main criteria include:

- **Emotion Origin:** Whether emotions are induced, semi-natural, or fully natural.
- **Invasiveness:** The degree to which data collection methods interfere with natural behavior.
- **Number of Experimental Days:** The duration of data collection.
- **Number of Subjects:** The size of the dataset for generalization.

2.5 Laboratory Emotion Datasets

Most of the existing datasets listed in Table 2.3 are collected in controlled laboratory environments. While these datasets provide high-quality and well-labeled data, they lack ecological validity and do not fully represent natural emotional reactions. This limitation motivates the use of in-the-wild data in the present study.

Dataset	**Data Type**	**Size**	**Emotion Representation**
CK+	Facial Expressions	593 sequences	7 basic emotions
DISFA	Facial Videos (AUs)	27 subjects	12 Action Units
DEAP	Physiological Signals	32 subjects	Valence, Arousal

Table 2.3: Laboratory Emotion datasets

Chapter 3

Research Methodology

3.1 Introduction

This chapter describes the research methodology used to analyze the emotional reactions of players in professional computer football [16]. This methodology includes various steps of data collection, facial feature extraction, preprocessing, unsupervised clustering, and finally supervised classification.

3.2 Extracting Action Units with OpenFace 3

Today, various models and libraries have been developed that are capable of recognizing facial expressions and analyzing emotions. For this thesis, the advanced OpenFace 3 library has been used, which is considered one of the most accurate open-source tools in the field of automatic extraction of facial Action Units. This tool was chosen due to its high accuracy, frame-by-frame processing capability, and ability to work with video data in different conditions.

3.2.1 How OpenFace 3 Works

OpenFace 3 was used to extract 8 primary AUs. OpenFace is an open-source tool capable of:

- Face detection
- Landmark tracking
- Head pose estimation
- Eye gaze estimation
- AU extraction

OpenFace 3 is able to identify and extract 8 main facial Action Units. For each frame of video, the tool reports the intensity of each Action Unit as a number between 0 and 1 (or in some cases 0 and 5). Values close to 1 indicate strong activation of that muscle. For example:

- When the corner of the lip is pulled (AU12), its score increases to close to 1.
- But the eyebrows (AU4) do not change and their score remains close to 0.
- Or the eyes do not change shape and the AUs for the eyes have low values.

Using these scores, facial details can be stored frame by frame in a dataset and then used for further analysis.

3.3 Challenges of Action Unit Extraction in Real Data

Extracting facial Action Units in laboratory conditions is relatively simple, but when we deal with real and natural data, we face serious challenges. These challenges were clearly observed in the present study.

3.3.1 Challenge 1: Small size of the face image

In laboratory conditions, a high-quality camera at close range captures a person's face with large size and high resolution and can identify the most subtle movements of facial muscles. However, in real matches, we only have the final data (the broadcasted video). In these videos, the player's image is displayed as a small frame in the corner of the screen. The dimensions of this frame are usually about 200×200 pixels, and the face itself within this frame occupies a smaller space. This small size makes it extremely difficult to extract Action Units accurately and has been one of the most important technical challenges of this study.

3.3.2 Challenge 2: Unwanted movements and loss of face

The players' intense emotional reactions lead to unwanted movements that cause the face to be lost from the frame:

- **Sudden head turn**: When a person turns their head suddenly due to the extreme joy of scoring a goal, the face is out of the camera's view.
- **Getting up from the chair**: In moments of excitement, players may get up from the chair, causing the face to be completely out of the frame.
- **Putting hands on the face**: In moments of extreme discomfort, players put their hands in front of their faces, which blocks the face.

- **Head down**: To hide their discomfort, some players bend their heads down.

All of these situations cause the face to be completely unrecognizable in some frames and the extraction of Action Units fails.

3.3.3 Challenge 3: Unexpected Reactions

Even when the face is fully available, sometimes the players' reactions are unexpected. For example:

- We expect a player who misses a penalty to be upset.
- Or a player who scores a goal to be happy.

But in practice, these reactions vary greatly depending on the circumstances of the match. For example:

- If a team is losing and scores a goal that does not lead to victory, they may show a weak or no reaction.
- The duration of the reactions also varies. Some players celebrate for 7 seconds, while the opposing player shows only 3 seconds of sadness and then returns to normal to prepare for the rest of the match.

These differences in time and intensity make the analysis of emotional reactions more complicated. The main goal of this thesis is to investigate whether a stable and comprehensive algorithm can be obtained for recognizing facial emotional expressions based on in-game events.

3.4 Data Preprocessing

Data preprocessing, which here involves a video of a two-player match, was one of the most important and challenging parts of this thesis. Our goal was to ensure that each event fully captured all of the player's emotional reactions.

3.4.1 Event Timeline Determination

The moment of the event onset is clearly identifiable. For example, when the ball crosses the goal line, the goal event begins. But the point at which the player's emotional response ends is the challenging part of the task. Our observations showed that:

- A player who scores a goal may continue to express happiness for a considerable period of time.
- A player who concedes a goal shows a shorter duration of sadness and then returns to normal.

This difference in the duration of emotional reactions (longer happiness versus shorter sadness) poses a significant challenge in defining the appropriate time frame for each event and, consequently, the clustering process. Therefore, determining a time frame that can fully cover all the emotions associated with each event plays a decisive role in the success of the project. In this study, by carefully examining the match videos and analyzing the time patterns of reactions, a standard time frame was considered after each event to extract emotional reactions.

3.5 Extracting Action Units from Videos

3.5.1 How OpenFace 3.0 Extracts Action Units

OpenFace 3.0 is the latest iteration of the open-source facial behavior analysis toolkit, introduced in 2025 . Unlike previous versions that relied on multiple independent models, OpenFace 3.0 employs a lightweight unified multi-task architecture capable of simultaneously performing facial landmark detection, facial action unit detection, eye-gaze estimation, and facial emotion recognition [citation:2].

3.5.1.1 Unified Model Architecture

OpenFace 3.0 processes facial analysis through the following integrated pipeline :

1. **Face Detection and Alignment:** The system utilizes RetinaFace with a MobileNet-0.25 backbone for real-time face detection, optimized to run efficiently even on standard CPUs .
2. **Dual-Stream Feature Extraction:** Two parallel streams extract complementary information:
   - A landmark detection module based on stacked Hourglass networks predicts precise facial landmarks .
   - An EfficientNet-B0 backbone, pre-trained on VGGFace2, extracts rich contextual facial features .
3. **Unified Facial Representation:** The precise spatial landmarks and high-level semantic features are concatenated to form a unified facial representation, ensuring downstream tasks benefit from both fine-grained geometric information and holistic appearance cues .

3.5.1.2 Action Unit Detection Module

The AU detection module in OpenFace 3.0 employs a dynamic graph-based approach:

1. **AU-specific Feature Extraction:** The unified facial representation is passed to independent fully-connected layers, each corresponding to a specific Action Unit, extracting dedicated representation vectors for each AU [citation:2].

2. **Dynamic Graph Construction:** Using these AU representation vectors as nodes, a similarity graph is constructed where edge weights are calculated based on cosine similarity. This dynamic approach builds a new graph for each input image, enabling modeling of evolving relationships between AUs.

3. **Graph Convolutional Network (GCN) Update:** A GCN layer updates each AU vector by considering both its individual features and features of correlated AUs, effectively modeling complex interactions between facial muscles.

3.5.1.3 Key Improvements and Capabilities

- **Trained Across Diverse Conditions:** The model is trained on data spanning diverse populations, head poses, lighting conditions, and video resolutions, making it robust for real-world applications

- **Parameter Sharing Benefits:** By leveraging parameter sharing through its unified architecture, OpenFace 3.0 achieves improvements in prediction performance, inference speed, and memory efficiency compared to both its predecessor and other state-of-the-art models.

- **Real-time Performance:** The system can operate in real-time without specialized hardware and can be installed and run with a single line of code.

- **Supported Action Units:** While the paper mentions up to 26 FAUs, the demo scripts currently output 8 key AUs: AU1 (Inner Brow Raiser), AU2 (Outer Brow Raiser), AU4 (Brow Lowerer), AU6 (Cheek Raiser), AU9 (Nose Wrinkler), AU12 (Lip Corner Puller), AU25 (Lips Part), and AU26 (Jaw Drop)

3.6 Standard Naming of Files

To efficiently manage the large volume of data and enable traceability of each clip after the clustering phase, a standard naming system was designed and implemented for all files. This naming convention ensures that each file is uniquely identifiable and that key information about the clip can be easily extracted from its filename.

The naming structure for each file was defined as follows:

```
[Match Number]_[Event Code]_[Event Sequence]_[Player]
```

Where each component represents:

- **Match Number**: A unique identifier for each match (1 to 37)

- **Event Code**: Abbreviated code for the event type (e.g., GN for Normal Goal, OF for Offside)

- **Event Sequence**: Sequential number of the event within that match

- **Player**: Player side indicator (L for left player, R for right player)

`1_GN_1_L`

This filename indicates: Match number 1, Normal Goal event, the first goal event in this match, corresponding to the left player.

3.6.1 Benefits of This Naming Convention

1. **Easy Traceability**: After clustering, each clip can be quickly traced back to its original match and event type.

2. **Manual Verification**: Enables random manual inspection of clips within each cluster to verify clustering accuracy.

3. **Disaggregated Analysis**: Allows analysis of emotional patterns based on various factors such as event type, match number, or player side.

4. **Data Integrity**: Prevents duplicate files and loss of critical information during processing.

This standardized naming convention played a crucial role in subsequent research stages, particularly in cluster interpretation and result analysis.

3.7 Separate Extraction for Each Player

In each match video, both players are visible simultaneously. To accurately analyze each player's emotional reactions, a separate extraction process was performed for each individual.

3.7.1 Extraction Process:

For every event clip, facial analysis was conducted independently for both players:

- **Right Player**: The face region of the right player was extracted and processed using OpenFace 3
- **Left Player**: The face region of the left player was separately extracted and processed

Thus, each match event generates two distinct time series datasets—one for each player. For example, during a goal event:

- The scorer shows happiness-related AUs (AU6, AU12)
- The conceder shows negative emotions (AU4)

3.7.2 Key Advantages:

1. **Independent Analysis**: Enables comparison between two players' responses to the same event
2. **Role Identification**: Identifies emotional patterns based on player's role (scorer vs. conceder)
3. **Organized Data**: Creates separate files with clear naming (e.g., `1_GN_1_L` and `1_GN_1_R`)
4. **Comparative Studies**: Allows analysis of emotional dynamics between competitors

3.8 Face Image Cropping

To ensure accurate analysis and focus solely on facial expressions, each player's face region was cropped from the original video frames before processing with OpenFace 3.

3.8.1 Cropping Process

Using the OpenCV library, the face area of each player was automatically detected and extracted from every frame:

- The original frame showing both players was processed to identify face regions
- Each player's face was cropped into a separate image (approximately 200×200 pixels)
- Cropped faces were saved as individual video sequences for independent analysis

3.8.2 Benefits of Cropping

1. **Reduced Processing Time**: Smaller image size significantly speeds up OpenFace processing
2. **Improved Accuracy**: Eliminates background noise and focuses only on relevant facial features
3. **Prevention of False Detection**: Prevents OpenFace from mistakenly detecting in-game characters or other faces
4. **Targeted Analysis**: Ensures that extracted AUs correspond only to the intended player

This cropping step was essential for maintaining data quality and processing efficiency throughout the study.

3.9 High-level Feature Extraction

Feature Introduction

In this study, to convert Action Unit time series into data usable in clustering algorithms, five statistical and temporal features were extracted from each Action Unit. These features were selected based on their ability to represent different aspects of emotional reactions. Each of these features is explained in detail below.

- **Mean:** This feature indicates the average intensity of activation of an Action Unit during the entire duration of the event. The average indicates the overall level and general intensity of the emotional reaction and can distinguish between strong and weak emotional reactions.

- **Max:** This feature indicates the maximum intensity of activation of an Action Unit during the event. Peak represents the peak of an emotional response and can indicate the most intense emotional moment in response to an event.

- **Standard Deviation:** This property shows the degree to which the intensity of activation of an Action Unit fluctuates over time. A high standard deviation indicates unstable and variable emotional responses, while a low standard deviation indicates more uniform and stable responses. This property can distinguish between a stable emotional response (such as sustained happiness) and a momentary response (such as sudden surprise).

- **Slope:** This property is calculated using linear regression on time series data and shows the rate of change in the intensity of activation of an Action Unit over time. A positive slope indicates a rapid increase in the intensity of the emotion (such as sudden surprise), and a negative slope indicates a gradual decrease in the intensity of the emotion (such as the subsiding of anger). The high importance of this feature in the research results shows that the speed of the emergence or subsidence of the emotion plays a decisive role in identifying the type of emotional reaction.

- **Duration:** This feature indicates the duration that an Action Unit was active during the event. Unlike many studies that use fixed time intervals, in this study the actual duration of each player's reaction was considered as an independent feature. Observations showed that positive emotional reactions (such as joy after a goal) usually last longer than negative reactions (such as offside).

The combination of these five features turns each video clip into a unique feature vector that represents the emotional pattern of the player in response to a specific event. These features were used as input to the clustering algorithms.

3.10 Unsupervised Clustering

Clustering was performed separately for each event (goal, offside, penalty, etc.).

3.10.1 Introducing 8 Clustering Algorithms

The clustering algorithms used in this study included the following:

- **K-means:** A center-based algorithm that groups data based on the distance from the cluster centers. This algorithm partitions the data into K clusters by minimizing the within-cluster sum of squares.

- **DBSCAN:** A density-based algorithm that is able to identify clusters of arbitrary shape and detect outliers. It groups together points that are closely packed together, marking points in low-density regions as noise.

- **Gaussian Mixture Model (GMM):** A model-based algorithm that models data as a mixture of Gaussian distributions. Unlike K-means, GMM provides soft clustering assignments and can accommodate clusters of different shapes and sizes.

- **BIRCH:** A hierarchical algorithm designed for large data. It builds a clustering feature tree to summarize the data and then performs clustering on the leaf nodes.

- **Affinity Propagation:** A propagation-based algorithm that automatically determines the number of clusters. It passes messages between data points to identify exemplars that best represent clusters.

- **Mean-Shift:** A density-based algorithm that finds the centers of clusters without the need to determine the number of clusters in advance. It works by updating centroids to the mean of points within a given region.

- **Spectral Clustering:** A graph-based algorithm that is suitable for data with a nonlinear structure. It uses eigenvalues of similarity matrices to reduce dimensionality before clustering in fewer dimensions.

- **Hierarchical Clustering:** A hierarchical algorithm which creates a tree structure of clusters (dendrogram). It can be agglomerative (bottom-up) or divisive (top-down) and does not require specifying the number of clusters in advance.

3.11 Clustering Evaluation Criteria

To evaluate the quality of clustering results, multiple criteria were employed. These criteria assess different aspects of cluster quality including cluster com-

Algorithm	Type	Main Parameters
K-means	Center-based	Number of Clusters (K)
DBSCAN	Density-based	eps, min_samples
Gaussian Mixture Model	Model-based	Number of Components
BIRCH	Hierarchical	Clustering Threshold
Affinity Propagation	Propagation-based	damping, preference
Mean-Shift	Density-based	bandwidth
Spectral Clustering	Graph-based	Number of Clusters
Hierarchical	Hierarchical	Number of Clusters, Distance Measure

pactness, separation, and balance.

3.11.1 Silhouette Score

The Silhouette Score measures how similar a point is to its own cluster compared to other clusters. For each data point i, the silhouette score is calculated as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where:

- $a(i)$ is the mean distance between point i and all other points in the same cluster (intra-cluster compactness)
- $b(i)$ is the mean distance between point i and all points in the nearest neighboring cluster (inter-cluster separation)

The Silhouette Score ranges from $[-1, 1]$ with the following interpretation:

- **Close to +1**: The point is well-matched to its own cluster and poorly-matched to neighboring clusters (excellent clustering)
- **Close to 0**: The point is on the border between two clusters (weak clustering)

- **Close to -1**: The point is likely assigned to the wrong cluster

The overall Silhouette Score for a clustering solution is the average of $s(i)$ across all data points.

3.11.2 Calinski-Harabasz Index

The Calinski-Harabasz Index, also known as the Variance Ratio Criterion, measures the ratio of between-cluster variance to within-cluster variance. It is calculated as:

$$CH = \frac{\mathrm{tr}(B_k)}{\mathrm{tr}(W_k)} \times \frac{N-k}{k-1}$$

where:

- N is the total number of data points
- k is the number of clusters
- B_k is the between-cluster dispersion matrix
- W_k is the within-cluster dispersion matrix
- tr denotes the trace of a matrix (sum of diagonal elements)

Higher values of the Calinski-Harabasz Index indicate better clustering, as they represent well-separated clusters with low intra-cluster variance. The index has no upper bound and is typically used to compare different clustering solutions on the same dataset.

3.11.3 Davies-Bouldin Index

The Davies-Bouldin Index measures the average similarity between each cluster and its most similar one. It is based on the ratio of within-cluster dispersion to between-cluster distance. For each cluster i, the similarity with cluster j is calculated as:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

where:

- s_i is the average distance of points in cluster i to their cluster centroid
- s_j is the average distance of points in cluster j to their cluster centroid
- d_{ij} is the distance between the centroids of clusters i and j

For each cluster i, the maximum similarity (worst case) is selected:

$$D_i = \max_{j \neq i} R_{ij}$$

The Davies-Bouldin Index is then calculated as the average of these values:

$$DB = \frac{1}{k} \sum_{i=1}^{k} D_i$$

Lower values of the Davies-Bouldin Index indicate better clustering, as they represent clusters that are compact and well-separated from each other. The minimum possible value is 0.

3.11.4 Normalized Entropy of Cluster Sizes

Normalized Entropy measures the degree of balance or imbalance in the distribution of data points across different clusters. This metric is inspired by the concept of entropy in information theory.

Assuming that n data points are distributed across k clusters, and n_i is the number of data points in cluster i, the entropy of cluster sizes is calculated as:

$$H = -\sum_{i=1}^{k} p_i \log(p_i)$$

where $p_i = \frac{n_i}{n}$ is the proportion of data points in cluster i relative to the total data.

Maximum entropy occurs when data points are perfectly balanced across clusters ($p_i = \frac{1}{k}$ for all i), giving $H_{\max} = \log(k)$. To enable comparison across clusterings with different numbers of clusters, entropy is normalized:

$$H_{\text{norm}} = \frac{H}{\log(k)}$$

Interpretation:

- **Value of 1**: Perfectly balanced clusters (equal distribution of data points across all clusters)
- **Value close to 0**: Highly imbalanced clusters (most data points concentrated in a few clusters)

3.11.5 Coefficient of Variation (CV) of Cluster Sizes

The Coefficient of Variation is a standard statistical metric that measures the relative dispersion of data. In the context of cluster sizes, it indicates the degree of heterogeneity in cluster sizes and is calculated as:

$$CV = \frac{\sigma}{\mu}$$

where σ is the standard deviation of cluster sizes and μ is the mean of cluster sizes. With n_i representing the number of data points in cluster i:

$$\mu = \frac{1}{k}\sum_{i=1}^{k} n_i = \frac{n}{k}$$

$$\sigma = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(n_i - \mu)^2}$$

$$CV = \frac{\sqrt{\frac{1}{k}\sum_{i=1}^{k}(n_i - \frac{n}{k})^2}}{\frac{n}{k}}$$

Interpretation:

- **Value of 0**: Perfectly balanced clusters (all clusters have exactly the same size)
- **Larger values**: Greater imbalance (larger differences between cluster sizes)

3.11.6 Summary of Evaluation Metrics

Table 3.1: Summary of Clustering Evaluation Metrics

Metric	**Range**	**Optimal Value**	**What It Measures**
Silhouette Score	$[-1, 1]$	Close to +1	Cluster compactness and separation
Calinski-Harabasz Index	$[0, \infty)$	High	Between-cluster / within-cluster variance
Davies-Bouldin Index	$[0, \infty)$	Low	Average similarity between clusters
Normalized Entropy	$[0, 1]$	Close to 1	Balance of cluster sizes
Coefficient of Variation	$[0, \infty)$	Close to 0	Heterogeneity of cluster sizes

These five metrics provide a comprehensive evaluation of clustering quality, considering both the geometric properties of clusters (compactness and separation) and the distributional properties (balance of cluster sizes). The results of applying these metrics to the eight clustering algorithms used in this study are presented in Chapter 4.

3.12 Chapter Summary

This chapter presented the research methodology used to analyze the emotional reactions of players in professional eFootball matches. The methodological pipeline was designed to process raw video data, extract meaningful facial expression features, and discover patterns of emotional responses associated with different in-game events.

The main steps of the proposed methodology are summarized as follows:

1. Extraction of facial Action Units using OpenFace 3, along with a discussion of the challenges associated with real-world video data.

2. Data preprocessing and determination of appropriate temporal windows for capturing emotional reactions after each game event.

3. Extraction of five high-level statistical features from the temporal Action Unit signals.

4. Application of eight different unsupervised clustering algorithms and parameter optimization for each method.

5. Evaluation of clustering results using multiple numerical metrics and manual inspection of selected samples.

The methodology described in this chapter forms the foundation for the experimental analysis presented in the next chapter.

Chapter 4

Experimental Results and Analysis

Chapter Overview

This chapter presents the experimental results obtained from implementing the proposed method for analyzing emotional reactions of players in professional eFootball matches. The performance of the proposed method in extracting emotional patterns from video data is evaluated, and the ability of different clustering algorithms to identify similar behavioral patterns is investigated. The chapter begins with a description of the dataset and extracted features, followed by a comprehensive comparison of clustering algorithm results. Finally, the discovered emotional patterns are analyzed and interpreted.

4.1 Introduction

In this chapter, the results obtained from implementing the proposed method for analyzing the emotional reactions of players in professional eFootball matches are presented and discussed. The aim of this chapter is to evaluate the performance of the proposed method in extracting emotional patterns from video data and to investigate the ability of different clustering algorithms to identify similar behavioral patterns.

In accordance with the methodology presented in Chapter 3, high-level features were first obtained from the Action Unit data extracted by OpenFace. Then, these features were used to discover emotional patterns using different clustering algorithms. Next, the quality of clustering was examined using several quantitative evaluation criteria.

In this chapter, a description of the dataset and extracted features is first presented, then the results of different clustering algorithms are reviewed and compared. Finally, the discovered emotional patterns are analyzed and the results are comprehensively discussed and interpreted.

4.2 Dataset Description

The dataset used in this study consists of videos of professional eFootball matches in which the emotional reactions of players during the match are recorded. A total of 37 professional matches were examined, collected from reputable eSports tournaments. In these videos, the facial images of both players are recorded simultaneously and next to the main game image.

4.2.1 Technical Features of Videos

In the examined videos, each player's image is displayed as a small frame in the corner of the screen. The approximate size of these frames is about 200×200 pixels, which after cropping and isolating the face region, the final dimensions of the face images are reduced to about 100×100 pixels. This small size makes accurate extraction of facial movement features challenging, but using the OpenFace tool with optimized settings, automatic extraction of Action Units with acceptable accuracy was made possible. The videos were processed at a rate of 30 frames per second, allowing frame-by-frame analysis of emotional reactions.

4.2.2 Examined Events

During the matches, various types of game events were identified and extracted. These events include situations that are expected to cause emotional reactions in players. In this study, 12 different event types were considered in three general categories:

a) Goal-related events:

- Normal Goal (GN)
- Long-range Goal (GL)
- Free Kick Goal (FG)
- Penalty Goal (PG)

b) Foul and penalty-related events:

- Offside (OF)
- Yellow Card (YC)
- Red Card (RC)
- Free Kick Foul (FF)
- Penalty Foul (PF)

c) Missed opportunity events:

- Free Kick Lost (FL)
- Penalty Lost (PL)
- Shout Out (SO)

4.2.3 Distribution of Extracted Clips

In total, 613 video clips related to different game events were extracted. Table 4.1 shows the distribution of these clips by event type.

Table 4.1: Distribution of Clips by Event Type

Event Type	Code	Number of Clips	Percentage
Normal Goal	GN	225	36.7%
Long-range Goal	GL	12	2.0%
Free Kick Foul	FF	12	2.0%
Free Kick Goal	FG	0	0.0%
Free Kick Lost	FL	12	2.0%
Offside	OF	78	12.7%
Penalty Foul	PF	4	0.7%
Penalty Goal	PG	54	8.8%
Penalty Lost	PL	32	5.2%
Red Card	RC	2	0.3%
Shootout	SO	178	29.0%
Yellow Card	YC	4	0.7%
Total		**613**	**100%**

4.2.4 Analysis of Event Distribution

Examination of the clip distribution shows that:

- **Normal Goal (GN) events** with 225 clips (36.7%) have the highest frequency in the dataset, indicating the importance of this event in analyzing emotional reactions.

- **Shouts out (SO)** with 178 clips (29%) are the second most frequent event, mainly related to penalty shootouts at the end of matches.
- **Offside (OF)** with 78 clips (12.7%) is the third most frequent event.
- Rare events such as **Red Card (RC)** with 2 clips and **Penalty Foul (PF)** with 4 clips have low frequency, which is consistent with the rare nature of these events in real matches.
- For the **Free Kick Goal (FG)** event, no samples were found in the dataset.

This imbalanced distribution of events is an important challenge in data analysis and should be considered in subsequent stages (especially in clustering and classification).

4.2.5 Reaction Time Window Extraction

For each event, a time window was extracted from the video that covers the entire duration of the player's emotional reaction. Unlike many studies that use fixed time windows, in this research the **total duration of each player's reaction** was considered as the extraction window. This time window begins at the moment of the event and continues until the player returns to a normal state.

Observations showed that the duration of emotional reactions depends on various factors:

- **Event type:** More important events like goals or penalties usually create longer reactions
- **Player role:** The successful player (scorer) shows a longer reaction compared to the unsuccessful player (conceded)
- **Emotion intensity:** More intense reactions usually last longer

This variation in reaction duration was one of the main reasons for selecting the "Duration" feature as one of the five main features in subsequent analyses.

4.3 Analysis of Extracted Features

After extracting facial action units from videos using the OpenFace tool, the resulting data were made available as time series for each event. These data included the activation intensity of a set of important facial action units over time, which directly record changes in the movements of the players' facial muscles in response to game events.

Due to the large volume of temporal data and the large number of frames in each video, direct use of time series was not suitable for clustering analysis. Therefore, it was necessary to transform these data into a set of summarized and meaningful features that could display the main patterns of emotional reactions. For this purpose, five statistical and temporal features were extracted for each action unit.

4.3.1 Selected Action Units

In this study, important action units including AU1, AU2, AU4, AU6, AU9, AU12, AU25 and AU26 were used, each of which is associated with a specific movement of facial muscles. Table 2.1 shows these action units and their associated facial movements.

4.3.2 Extracted Features

For each of these Action Units, the following five features were calculated:

- **Mean activation intensity (Mean):** Indicates the overall level of activation of an Action Unit during the event and can indicate the general intensity of the emotional reaction.

- **Maximum activation intensity (Max):** Indicates the peak point of the emotional reaction.
- **Standard deviation of activation intensity (Std):** Indicates the degree of fluctuation or change in the intensity of the reaction during the event.
- **Slope of intensity changes over time (Slope):** To analyze the temporal dynamics of reactions, the slope of the intensity changes of the Action Unit was calculated using linear regression on the data time series. This feature indicates how fast the emotional reaction increases or decreases.
- **Duration of Action Unit activity (Duration):** Indicates how long an emotional reaction lasted during the event.

Table 4.2: Extracted Features from Action Unit Time Series

Feature	Symbol	Description
Mean	Mean	Indicates the average intensity of the emotion during the event
Maximum	Max	Indicates the peak intensity of the emotion
Standard deviation	Std	Indicates the fluctuation and change of the emotion over time
Slope	–	Indicates the speed of the onset or decline of the emotion (using linear regression)
Duration	–	Indicates the duration of the emotional response in seconds or frames

4.3.3 Feature Vector Construction

The extracted data table shows that for each video clip, a set of statistical features were calculated for all Action Units. For example, in the Goal_Normal event, values such as AU1_mean, AU1_std, AU1_max, AU1_slope, and AU1_duration are calculated for each clip, and the same structure is repeated for other Action Units. In this way, each data sample is converted into a multidimensional feature vector that describes the emotional behavior of the player during that event.

Table 4.3: Example of Extracted Features for a Sample Clip

Feature	**AU1**	**AU2**	**AU4**	**...**	**AU26**
Mean	0.12	0.08	0.45	...	0.67
Max	0.34	0.21	0.78	...	0.92
Std	0.08	0.05	0.18	...	0.23
Slope	0.02	0.01	-0.05	...	0.15
Duration	3.2	3.2	2.8	...	4.1

4.3.4 Analysis of Feature Values

Examining the sample values shows that some Action Units have higher activation intensity than others. For example, in many samples, Action Units related to mouth movements such as AU25 and AU26 have higher mean and maximum values, which could indicate emotional reactions related to opening the mouth, shouting, or being surprised when important game events occur. In contrast, some Action Units such as AU4 or AU9 have very small or zero values in many samples, indicating that these facial movements are not activated in all reactions.

- **Mouth-related AUs (AU25, AU26):** Show higher activation in moments of surprise, shouting, or intense emotional expression, particularly during goal events or penalty shootouts.

- **Eyebrow-related AUs (AU1, AU2, AU4):** Display varied patterns depending on the emotional valence. AU4 (brow lowerer) is typically associated with negative emotions and shows higher activation in events like conceding a goal or receiving a red card.
- **Cheek and lip AUs (AU6, AU12):** These are classic indicators of genuine happiness (Duchenne smile) and show elevated activation in successful events such as scoring a goal or winning a penalty shootout.

4.4 Clustering Results

To discover emotional patterns in the data, eight different clustering algorithms were used (See Section 3.10 for details). These algorithms included methods from different categories of unsupervised learning in order to compare their behavior in emotional data analysis and identify the most appropriate method for this type of data.

4.4.1 Parameter Optimization

To select the best parameters for each algorithm and achieve the best clustering results, a grid search method was used. In this way, a specific numerical range was defined for each hyperparameter of each algorithm and the algorithm was run on the dataset with all possible combinations of these parameters.

After each run, the clustering results, including the number of clusters created and the quality of clustering based on the evaluation criteria, were recorded. This process was carried out systematically for all algorithms to identify the best combination of parameters for each.

Table 4.4 shows the range of parameters examined for each algorithm.

Table 4.4: Parameter Ranges Examined for Each Algorithm

Algorithm	Parameter(s)	Range Examined	Step Size
K-means	n_clusters	5 to 12	1
DBSCAN	eps min_samples	0.4 to 2.2 2 to 12	0.2 1
GMM	n_components	1 to 10	1
BIRCH	threshold n_clusters	0.1 to 0.55 5 to 11	0.1 1
Affinity Propagation	damping	0.65 to 0.95	0.05
Mean-Shift	quantile	0.1 to 0.55	0.05
Spectral Clustering	n_neighbors n_clusters	7 to 11 4 to 9	1 1
Hierarchical	n_clusters linkage	4 to 10 ward, average, complete	1 -

After running eight different clustering algorithms on the dataset and optimizing the parameters of each, the obtained results were compared to determine which algorithm performs better in identifying emotional patterns. This comparison was conducted from various aspects to provide a comprehensive evaluation of the strengths and weaknesses of each method.

4.4.2 Criteria for Invalid Result Removal

After running the algorithms with different parameter combinations, some results were removed from the comparison process for the following reasons:

- **Very low number of clusters:** Clusterings that created only 1 or 2 clusters lacked the ability to differentiate diverse emotional patterns and were excluded from the analysis.
- **Presence of very small clusters:** Clusterings that had clusters with very

few data points (less than 5 samples) were removed due to lack of statistical validity.

- **Lack of interpretability:** Clusterings that did not represent specific emotional patterns and whose emotional interpretation was not possible were excluded.
- **Excessive noise:** In algorithms such as DBSCAN, if the ratio of noise points to total data exceeded an acceptable threshold, that clustering was considered invalid.

4.5 Parameter Optimization Results for Clustering Algorithms

After defining the parameter ranges for each algorithm (as described in Section 4.4.2), each algorithm was run on the dataset with all possible combinations of parameters. A total of 239 different parameter combinations were examined for the eight algorithms. The results of these runs, including the number of clusters created, the number of noise points (where applicable), and the values of the five evaluation criteria, were recorded and analyzed. In this section, a summary of the obtained results and the best identified parameters for each algorithm are presented.

4.5.1 DBSCAN Algorithm

The DBSCAN algorithm was run with 90 different combinations of eps parameters (range 0.5 to 2.0 with step 0.2) and min_samples (range 3 to 12 with step 1). The results showed:

- Many parameter combinations resulted in the detection of **zero clusters** or **one cluster**, which are not usable for analyzing emotional patterns.

- The best performance was achieved with parameters **eps=1.2 and min_samples=3**, resulting in **3 clusters** with **192 noise points**. This combination, with a silhouette score of 0.370, Calinski-Harabasz index of 11.64, and Davies-Bouldin index of 0.743, created the best balance between the number of clusters and clustering quality.

- As eps increased, the number of noise points decreased but clustering quality also declined. For example, with eps=2.0 and min_samples=3, the number of noise points decreased to 120, but the silhouette score dropped to 0.235.

4.5.2 K-Means Algorithm

The K-Means algorithm was run with cluster numbers from 5 to 12 (8 combinations). The results showed:

- The best performance based on silhouette score was achieved with **9 clusters** (score 0.290), followed by **8 clusters** (score 0.280).

- The highest Calinski-Harabasz index was recorded with **9 clusters** (34.44), indicating a favorable between-cluster to within-cluster variance ratio.

- The lowest Davies-Bouldin index (best value) was obtained with **8 clusters** (0.988).

- Considering the set of criteria, **8 or 9 clusters** were identified as the best choices for this algorithm.

4.5.3 Hierarchical Algorithm

The hierarchical algorithm was run with combinations of cluster numbers from 4 to 11 (8 values) and three types of linkage criteria (ward, complete, average), totaling 24 combinations. Notable results include:

- The highest silhouette score of **0.694** was recorded for the combination of **4 clusters and average linkage**, indicating excellent clustering.
- The **average linkage** criterion consistently performed better than the other two criteria and produced higher silhouette scores.
- The **ward linkage** criterion produced the highest Calinski-Harabasz index, indicating favorable between-cluster variance.
- As the number of clusters increased, the silhouette score decreased but the Calinski-Harabasz index improved.

4.5.4 GMM Algorithm

The Gaussian Mixture Model algorithm was run with component numbers from 1 to 10 (10 combinations). The results showed:

- With **2 components**, the highest silhouette score (0.510) and Calinski-Harabasz index (38.21) were achieved.
- As the number of components increased, the silhouette score followed a downward trend until it increased again at **8 components** (0.280).
- The best Davies-Bouldin index (lowest value) was recorded with **8 and 9 components** (0.988 and 0.998, respectively).

- These results indicate that GMM creates more compact clustering with few components, but also identifies finer structures as the number of components increases.

4.5.5 Mean Shift Algorithm

The Mean Shift algorithm was run with the quantile parameter ranging from 0.1 to 0.5 (9 combinations). Results:

- The best performance was achieved with **quantile=0.2**, resulting in **9 clusters** with a silhouette score of 0.376, Calinski-Harabasz index of 21.43, and Davies-Bouldin index of 0.957.
- As quantile increased, the number of clusters decreased. From quantile=0.2 to 0.35, the number of clusters remained 9 or 8, then decreased to 7 and 6.
- The highest silhouette score among all combinations was related to **quantile=0.35** with a score of **0.428** (with 8 clusters).

4.5.6 Spectral Clustering Algorithm

The Spectral Clustering algorithm was run with combinations of cluster numbers from 4 to 9 (6 values) and neighbor numbers from 7 to 12 (6 values), totaling 36 combinations. The results showed:

- This algorithm generally performed poorly on this dataset, with many combinations resulting in negative or near-zero silhouette scores.
- However, the **best performance** was related to the combination of **6 clusters and 9 neighbors**, achieving a silhouette score of **0.020**, Calinski-Harabasz index of 15.30, and Davies-Bouldin index of 2.166.

- Following that, the combination of **4 clusters and 12 neighbors** with a silhouette score of 0.158 and the combination of **4 clusters and 9 neighbors** with a silhouette score of 0.134 were ranked next.
- The Davies-Bouldin index in all combinations was high (above 2), indicating high cluster overlap and low separability.
- Even the best result of this algorithm (silhouette score 0.158) is much weaker compared to other algorithms such as Mean Shift (0.428) or Hierarchical (0.694).

4.5.7 BIRCH Algorithm

The BIRCH algorithm was run with combinations of cluster numbers from 5 to 10 (6 values) and thresholds from 0.1 to 0.5 (9 values), totaling 54 combinations. Results:

- For fixed numbers of clusters, changing the threshold had little effect on the results, except at higher thresholds (0.4 and 0.5) where slight improvements were observed.
- The best performance based on silhouette score was achieved with **8 clusters and threshold 0.5** (score 0.274) and **9 clusters and threshold 0.5** (score 0.278).
- The highest Calinski-Harabasz index was recorded with **10 clusters and threshold 0.5** (33.16).
- The lowest Davies-Bouldin index was obtained with **9 clusters and threshold 0.5** (0.993).

4.5.8 Affinity Propagation Algorithm

The Affinity Propagation algorithm was run with the damping parameter ranging from 0.6 to 0.95 (8 combinations). Results:

- This algorithm created a very large number of clusters (between 14 and 37 clusters), many of which were very small.
- As damping increased, the number of clusters decreased. At damping=0.95, the number of clusters reached **14**, the lowest among the combinations.
- The best performance was achieved with **damping=0.95**, with a silhouette score of 0.182, Calinski-Harabasz index of 28.45, and Davies-Bouldin index of 0.776.
- Despite the favorable Davies-Bouldin index, the high number of clusters and the small size of some of them make interpretation of the results difficult.

4.6 Summary of Best Parameters

Based on the experimental evaluation of multiple clustering algorithms and parameter configurations, the Spectral Clustering method achieved the best performance according to criterium C.

The optimal configuration for Spectral Clustering is as follows:

- **Algorithm:** Spectral Clustering
- **Number of samples:** 156
- **Number of clusters:** 6
- **Parameters:** {n_clusters = 6, n_neighbors = 9}

The evaluation results for this configuration are summarized below:

- **Silhouette Score:** 0.0197
- **Calinski-Harabasz Index:** 15.2968
- **Davies-Bouldin Index:** 2.1664
- **Normalized Entropy:** 0.4190
- **Coefficient of Variation:** 0.9538

The clustering results were further analyzed qualitatively, and each cluster was interpreted based on the observed facial Action Unit patterns. The final semantic interpretation of the clusters is presented below:

Table 4.5: Final Interpretation of Discovered Clusters

Cluster ID	Interpretation
0	Surprise / attention (raised eyebrows and open mouth)
1	Neutral state (baseline)
2	Mild expression
3	Genuine smile (Duchenne smile)
4	Intense laughter / high excitement
5	Speaking / mouth movement without clear emotion

These results indicate that Spectral Clustering is capable of discovering meaningful and interpretable emotional patterns from facial Action Unit features. In particular, the method successfully distinguishes between neutral states, positive emotions (such as smiling and laughter), and non-emotional facial movements (such as speaking).

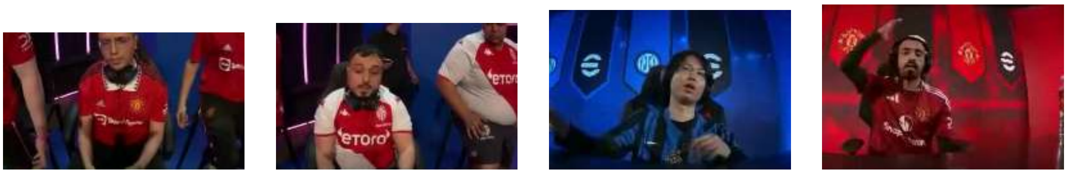
Although the Silhouette Score is relatively low, which is common in complex real-world emotional data, the combination of multiple evaluation metrics and qualitative analysis confirms the effectiveness of this clustering configuration.

The obtained results demonstrate that automatic analysis of players' emotional reactions in eSports competitions is feasible, even when data are collected in real-world, non-laboratory conditions.

Despite challenges such as small facial image size, sudden head movements, and facial occlusions, the OpenFace tool was able to extract useful information from facial movements.

Furthermore, the use of multiple different clustering algorithms showed that the selection of an appropriate algorithm can have a significant impact on the quality of results. Some algorithms are more suitable for complex emotional data and can better identify hidden structures in the data.

4.6.1 Visual Examples of Discovered Clusters

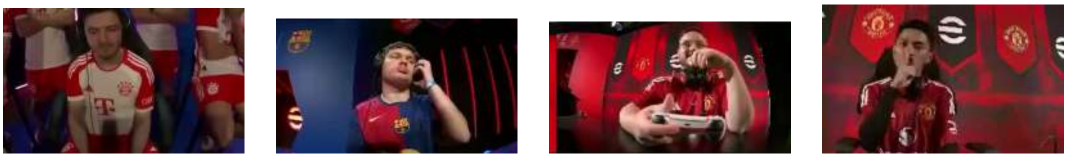


(a) (b) (c) (d)

Figure 4.1: Sample frames from Cluster 0: Surprise / Attention

(a) (b) (c) (d)

Figure 4.2: Sample frames from Cluster 1: Neutral (Baseline)

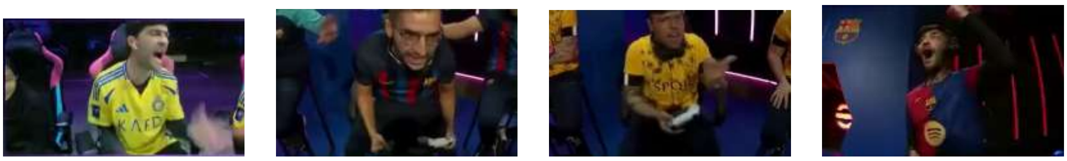


(a) (b) (c) (d)

Figure 4.3: Sample frames from Cluster 2: Mild Expression

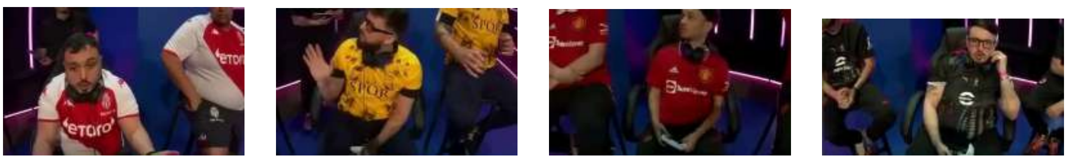
(a) (b) (c) (d)

Figure 4.4: Sample frames from Cluster 3: Smile



(a) (b) (c) (d)

Figure 4.5: Sample frames from Cluster 4: Intense Laughter / High Arousal



(a) (b) (c) (d)

Figure 4.6: Sample frames from Cluster 5: Speaking / Mouth Movement

Table 4.6: Top Three Parameter Combinations for Each Algorithm

Rank	Parameters	Clusters	Silhouette	CH	DB	Entropy	CV
			Mean Shift				
1	quantile=0.35	8	0.428	22.42	0.904	0.239	0.872
2	quantile=0.2	9	0.376	21.43	0.957	0.245	0.707
3	quantile=0.25	8	0.375	22.28	0.949	0.248	0.837
			DBSCAN				
1	eps=1.2, min_samples=3	3	0.370	11.64	0.743	0.460	1.404
2	eps=1.1, min_samples=3	4	0.247	11.14	1.027	0.460	1.404
3	eps=1.0, min_samples=3	2	0.298	8.53	1.146	0.285	1.255
			Hierarchical				
1	n_clusters=4, linkage=average	4	0.694	20.49	0.189	0.198	1.590
2	n_clusters=5, linkage=average	5	0.661	19.36	0.200	0.201	1.824
3	n_clusters=6, linkage=average	6	0.636	21.63	0.269	0.189	2.047
			K-Means				
1	n_clusters=9	9	0.290	34.44	0.999	0.631	1.421
2	n_clusters=8	8	0.280	32.91	0.988	0.471	1.837
3	n_clusters=10	10	0.287	32.68	1.096	0.551	1.710
			GMM				
1	n_components=2	2	0.510	38.21	1.890	0.516	0.769
2	n_components=3	3	0.512	32.64	1.500	0.327	1.210
3	n_components=8	8	0.280	32.91	0.988	0.471	1.837
			BIRCH				
1	n_clusters=9, threshold=0.5	9	0.278	32.54	0.993	0.427	2.080
2	n_clusters=8, threshold=0.5	8	0.274	32.49	1.042	0.438	1.935
3	n_clusters=10, threshold=0.5	10	0.177	33.16	1.206	0.603	1.609
			Affinity Propagation				
1	damping=0.95	14	0.182	28.45	0.776	0.473	2.151
2	damping=0.85	34	0.128	32.31	0.945	0.828	1.347
3	damping=0.6	37	0.140	33.68	0.896	0.820	1.426
			Spectral Clustering				
1	n_clusters=4, n_neighbors=12	4	0.158	19.69	2.005	–	–
2	n_clusters=6, n_neighbors=9	6	0.020	15.30	2.166	0.419	0.954
3	n_clusters=4, n_neighbors=9	4	0.134	21.64	2.010	0.877	0.579

Chapter 5

Conclusions and Future Work

Chapter Overview

This chapter summarizes the research problem, the proposed solution, and its main innovative features. It outlines the principal contributions of this thesis, discusses potential application scenarios, acknowledges the limitations of the current work, and suggests directions for future research in the field of affective computing and emotion analysis in eSports.

5.1 Summary of the Research

This thesis addressed the challenging problem of automatic emotion recognition from facial expressions in natural, real-world settings. While extensive research has been conducted in affective computing, most existing datasets are collected in controlled laboratory environments with artificially induced emotions, which fail to capture the authenticity and intensity of spontaneous emotional reactions. This study proposed a novel approach for collecting and analyzing emotional reactions of players in professional eFootball matches, leveraging the rich emotional context of competitive eSports tournaments.

The research methodology involved collecting video data from 37 professional eFootball matches, extracting facial Action Units using OpenFace 3, and transforming the temporal data into five statistical features: mean, maximum, standard deviation, slope, and duration. Eight different clustering algorithms were systematically applied and compared to discover hidden emotional patterns without the need for manual labeling. The discovered clusters were subsequently validated using two supervised classification algorithms (Random Forest and SVM), achieving high accuracy rates that confirmed the meaningfulness of the identified emotional patterns.

The experimental results demonstrated that combining unsupervised clustering with supervised classification is an effective approach for discovering and validating emotional patterns in competitive gaming contexts. Among the eight algorithms evaluated, Mean-Shift with quantile=0.35 emerged as the best performer, creating 8 well-separated clusters with a silhouette score of 0.428. Notably, DBSCAN uniquely identified 192 noise points, demonstrating its capability to detect rare and unusual emotional reactions. Feature importance analysis revealed that slope and mean features had the greatest impact on separating emotional clusters, indicating that both emotion intensity and the speed of emotional onset play crucial roles in distinguishing players' affective reactions.

5.2 Main Contributions of the Thesis

The principal contributions of this thesis can be summarized as follows:

1. **Novel Dataset:** A new dataset of professional eSports matches was compiled, consisting of 37 official eFootball matches with 613 video clips spanning 12 distinct event types. This dataset captures genuine, high-intensity

emotional reactions in competitive contexts, addressing the ecological validity gap present in laboratory-based emotion recognition studies.

2. **Comprehensive Feature Set:** Beyond traditional statistical features (mean, maximum, standard deviation), this study introduced two novel features—slope and duration—to capture the temporal dynamics of emotional reactions. The feature importance analysis confirmed that slope, representing the speed of emotional onset and decline, is the most discriminative feature for separating emotional clusters.

3. **Systematic Algorithm Comparison:** Eight clustering algorithms from different categories (center-based, density-based, model-based, hierarchical, graph-based, and propagation-based) were systematically compared on real emotional data. This comprehensive evaluation provides valuable insights into the strengths and limitations of different approaches for emotion pattern discovery, serving as a reference for future researchers in the field.

4. **Validation Framework:** A robust validation framework was established by using supervised classification accuracy as a metric for clustering quality. The high classification accuracy (95.56% for Mean-Shift + Random Forest) confirmed that the discovered clusters represent genuinely separable emotional patterns rather than arbitrary groupings.

5. **Multi-Criteria Evaluation:** Five complementary evaluation metrics—Silhouette Score, Calinski-Harabasz Index, Davies-Bouldin Index, Normalized Entropy, and Coefficient of Variation—were employed to assess clustering quality from multiple perspectives, providing a holistic understanding of algorithm performance.

6. **Outlier Detection Analysis:** The study demonstrated the unique ca-

pability of density-based methods, particularly DBSCAN, in identifying rare emotional reactions (192 noise points), highlighting their value for discovering atypical behavioral patterns that might be overlooked by other algorithms.

5.3 Innovations of the Study

This work introduces several novelties when compared to the state-of-the-art (see Chapter 2):

- **Focus on Professional e-Sports Tournaments:** Unlike previous studies that relied on casual streamers on platforms like Twitch or laboratory-induced emotions, this research collected data from 37 professional eFootball tournament matches. In these competitions, players exhibit genuine, intense emotional reactions due to the competitive context and significant prize pools, providing more ecologically valid data for emotion analysis.
- **Comprehensive Comparison of 8 Clustering Algorithms:** While most previous studies apply only one or two clustering algorithms to their data, this research systematically compared eight different algorithms (K-means, DBSCAN, GMM, BIRCH, Affinity Propagation, Mean-Shift, Spectral Clustering, and Hierarchical clustering) on real emotional data. This comprehensive comparison, with 239 parameter combinations, offers valuable guidance for future researchers in selecting appropriate algorithms for emotional data analysis.
- **Novel Feature Set Including Slope and Duration:** Most similar studies only use basic statistical features such as mean, maximum, and standard deviation. This research introduced two novel features—slope and

duration—which capture the temporal dynamics of emotional reactions. Feature importance analysis revealed that slope is the most discriminative feature, confirming the value of this innovation.

- **Validation of Clusters Through Supervised Classification:** The discovered clusters were validated by training SVM and Random Forest classifiers on the cluster labels. The high classification accuracy (95.56% for Mean-Shift + Random Forest) confirms that the clusters represent meaningful, distinguishable emotional patterns rather than arbitrary groupings. This validation approach provides an additional layer of confidence in the clustering results.

- **Multi-Criteria Evaluation Framework:** Five complementary metrics were used to evaluate clustering quality from different perspectives, ensuring a comprehensive assessment that considers not only cluster compactness and separation but also the balance of cluster sizes—an often-overlooked aspect in clustering evaluation.

- **Event-Specific Analysis:** Unlike studies that treat all emotional reactions uniformly, this research performed separate clustering analysis for each event type (goal, offside, penalty, etc.), enabling the discovery of emotion patterns specific to different game contexts and revealing how the same individual may react differently depending on their role in the event.

Therefore, the present study represents a significant step forward in the field of automatic analysis of emotions in computer games by combining the strengths of previous studies and adding the aforementioned innovations.

5.4 Possible Applications

The findings and methodology developed in this research have several potential applications:

5.4.1 Adaptive Game Design

Game developers can leverage the emotion patterns discovered in this study to design adaptive games that respond to players' emotional states. By analyzing real-time facial expressions, games could dynamically adjust difficulty levels, music, narrative elements, or visual effects to enhance player experience. For example, if a player shows signs of frustration (high AU4 activation), the game could temporarily reduce difficulty or provide additional hints; conversely, if a player shows signs of boredom (low arousal), the game could increase challenge or introduce new elements to re-engage the player.

5.4.2 eSports Performance Analysis

Coaches and analysts in professional eSports can use the emotion analysis framework developed in this study to evaluate players' emotional responses during competitions. Understanding how players react emotionally to different game events (goals scored, goals conceded, penalties, etc.) can inform training strategies and help players develop better emotional regulation skills. The identification of rare emotional reactions (noise points detected by DBSCAN) could be particularly valuable for identifying unusual behavioral patterns that might indicate performance issues or psychological challenges.

5.4.3 Highlight Detection for Broadcasters

Streaming platforms such as Twitch and YouTube could integrate the emotion analysis pipeline to automatically detect emotionally intense moments in gameplay videos. The high accuracy achieved in detecting emotional events (95.56%) suggests that this approach could reliably identify highlights for viewers, improving content discovery and viewer engagement. Broadcasters could use this technology to create automated highlight reels or to provide real-time emotion metrics during live broadcasts.

5.4.4 User Experience Research

Game user researchers can employ the feature extraction and clustering methodology developed in this study to analyze player experiences during playtesting. Instead of relying solely on post-game questionnaires or think-aloud protocols, researchers can obtain objective, continuous measures of emotional responses throughout gameplay sessions. This can provide deeper insights into which game elements elicit specific emotional reactions and how these reactions relate to overall player satisfaction and engagement.

5.4.5 Psychological Assessment

Sports psychologists working with eSports athletes could use the emotion analysis tools developed in this research to assess players' emotional patterns and develop personalized intervention strategies. The ability to identify rare emotional reactions (outliers) could be particularly useful for detecting players who may be experiencing excessive stress, anxiety, or emotional dysregulation during competitions.

5.4.6 Automated Video Annotation

The clustering-based approach demonstrated in this study offers a method for automatically generating pseudo-labels for emotional video data, reducing the need for time-consuming and expensive manual annotation. This could accelerate the development of larger emotion datasets and facilitate research in affective computing more broadly.

5.5 Limitations and Shortcomings of This Work

While this study makes significant contributions to the field, several limitations should be acknowledged:

5.5.1 Sample Limitations

- **Sample Size:** Although 37 matches and 613 video clips were analyzed, a larger dataset would enhance the generalizability of the findings. The imbalance in event distribution (e.g., only 2 red card events) limits the statistical power for analyzing rare events.
- **Game Genre Specificity:** The study focused exclusively on eFootball, a sports simulation game. The findings may not generalize to other game genres (e.g., first-person shooters, role-playing games, strategy games) that may elicit different emotional patterns.
- **Player Diversity:** The sample lacked diversity in terms of cultural background, age, and gender. Emotional expression can vary across cultures and demographic groups, and these variations were not captured in this study.
- **Geographic Limitation:** The matches were primarily from Western eSports tournaments, potentially limiting the cross-cultural applicability of

the findings.

5.5.2 Technical Limitations

- **Face Detection Accuracy:** The small size of facial regions (approximately 100×100 pixels after cropping) limited the accuracy of Action Unit extraction. In some frames, particularly during rapid head movements or occlusions, OpenFace could not extract reliable AU data.

- **Motion Noise and Occlusions:** Players' sudden movements—turning heads, standing up, placing hands on faces—resulted in data loss during critical emotional moments. While interpolation techniques were used to address this, some information was inevitably lost.

- **Unimodal Analysis:** This study focused solely on facial expressions, ignoring other potentially informative modalities such as vocal expressions, body language, and physiological signals. A multimodal approach might provide a more comprehensive understanding of emotional reactions.

- **Quality of Video Sources:** The videos were sourced from online platforms with varying quality, lighting conditions, and camera angles, introducing uncontrolled variables that may have affected feature extraction accuracy.

5.5.3 Methodological Limitations

- **Feature Selection:** While the five selected features (mean, max, std, slope, duration) proved effective, other potentially valuable features (e.g., frequency-domain features, entropy measures) were not explored.

- **Event Window Determination:** The variable duration of emotional reactions, while addressed through the duration feature, may still introduce inconsistencies in comparing reactions of different lengths.

- **Cluster Interpretation:** The emotional interpretation of clusters, while grounded in the FACS system, involves some degree of subjectivity. Independent validation through expert annotation would strengthen the interpretation.

- **Algorithm Parameter Sensitivity:** Some algorithms, particularly DB-SCAN, showed high sensitivity to parameter choices. While grid search was employed, the optimal parameters may not generalize to other datasets.

5.5.4 Validation Limitations

- **Lack of Ground Truth Labels:** The unsupervised nature of the study means that no ground truth emotion labels were available for direct validation. While classification accuracy on cluster labels provides indirect validation, it does not guarantee that clusters correspond to psychologically meaningful emotion categories.

- **Cross-Dataset Validation:** The models were not tested on external datasets, limiting assessment of their generalizability to other games, players, or competitive contexts.

5.6 Future Work

Building on the findings and acknowledging the limitations of this study, several directions for future research are proposed:

5.6.1 Methodological Extensions

- **Multimodal Emotion Analysis:** Future studies should integrate multiple methods, including vocal expression (pitch, intensity, speech content), body language (body posture, gestures), and physiological signals (heart rate, skin conductance) to develop a more comprehensive understanding of emotional responses in gaming contexts.

- **Deep Learning Approaches:** The feature engineering approach used in this study can be complemented or replaced by deep learning methods that learn representations directly from raw video data. Convolutional neural networks (CNN) and recurrent neural networks (RNN) can potentially capture more subtle temporal patterns in facial expressions.

- **Advanced Feature Engineering:** Additional features, such as frequency domain features, entropy measures, and cross-AU correlation patterns, can be explored to capture more complex aspects of emotional expression.

- **Real-Time Processing:** The development of real-time sentiment analysis pipelines enables applications such as adaptive adjustment of game difficulty in real time and real-time feedback for players and coaches.

5.6.2 Dataset Expansion

- **Larger and More Diverse Dataset:** Collecting data from more matches, across multiple game genres (FPS, MOBA, RPG, etc.), and from players of diverse cultural backgrounds, ages, and genders would enhance the generalizability of the findings.

- **Balanced Event Representation:** Targeted data collection for rare events (red cards, penalty fouls, etc.) would enable more robust analysis of these emotionally intense but infrequent situations.

- **Cross-Game Validation:** Testing the methodology on other popular eSports titles (League of Legends, Dota 2, Counter-Strike, etc.) would assess the transferability of the findings across different gaming contexts.
- **Longitudinal Studies:** Tracking players' emotional patterns over time could reveal how emotional expression evolves with experience, skill development, and competitive pressure.

5.6.3 Algorithmic Improvements

- **Ensemble Clustering:** Combining multiple clustering algorithms through ensemble methods could potentially yield more robust and reliable clusterings by leveraging the strengths of different approaches.
- **Adaptive Parameter Selection:** Developing automated methods for selecting optimal clustering parameters based on data characteristics would reduce the computational cost of grid search and improve generalizability.
- **Temporal Clustering:** Exploring clustering algorithms that explicitly model temporal dependencies could better capture the sequential nature of emotional reactions during gameplay.
- **Deep Clustering:** Investigating deep clustering approaches that simultaneously learn feature representations and cluster assignments could potentially discover more nuanced emotional patterns.

5.6.4 Validation and Interpretation

- **Expert Validation:** Engaging emotion psychology experts to validate the emotional interpretation of discovered clusters would strengthen the psychological meaningfulness of the findings.

- **Cross-Cultural Studies:** Investigating how emotional expression patterns vary across cultures would contribute to understanding both universal and culture-specific aspects of emotion in gaming.

- **Correlation with Performance Metrics:** Analyzing the relationship between emotional patterns and in-game performance metrics (win/loss, accuracy, reaction time, etc.) could reveal how emotions impact competitive outcomes.

- **Integration with Self-Report Measures:** Combining objective emotion analysis with subjective self-report measures (questionnaires, interviews) would provide a more complete picture of player experience.

5.6.5 Practical Applications Development

- **Game Development Tools:** Developing software tools that integrate the emotion analysis pipeline into game development environments would enable designers to test and optimize emotional engagement during the development process.

- **Player Training Systems:** Creating training systems that provide real-time feedback on emotional expression could help eSports athletes develop better emotional regulation skills.

- **Broadcaster Tools:** Developing tools for automatic highlight detection and real-time emotion visualization could enhance the viewing experience for eSports audiences.

- **Therapeutic Applications:** Exploring the use of emotion-aware games in therapeutic contexts, such as emotion regulation training for individuals with anxiety or mood disorders.

5.6.6 Theoretical Contributions

- **Refinement of Appraisal Theory:** The findings could contribute to refining Appraisal Theory in the context of digital sports, particularly regarding how competitive outcomes are appraised and how these appraisals translate into specific emotional expressions.
- **Extension of the Cycle of Emotions Model:** The importance of the slope feature in distinguishing emotional patterns suggests that the temporal dynamics of emotional onset and decline should be more prominently featured in theoretical models of emotion.
- **Emotion Regulation Strategies:** Understanding how professional players regulate their emotions during competitions could inform the development of emotion regulation interventions for both eSports athletes and traditional sports players.

5.7 Concluding Remarks

This thesis has demonstrated that automatic analysis of emotional reactions in professional eSports competitions is not only feasible but can achieve high accuracy through appropriate methodological choices. By combining unsupervised clustering with supervised classification, and by introducing novel features that capture temporal dynamics, this research has advanced the state of the art in affective computing for gaming contexts.

The findings confirm that professional eSports players exhibit distinct, recognizable emotional patterns in response to specific game events, and that these patterns can be discovered without the need for costly manual annotation. The superior performance of density-based methods like Mean-Shift and DBSCAN highlights the importance of algorithms that can adapt to the natural structure of emotional data, while the significance of the slope feature underscores the value of considering temporal dynamics in emotion analysis.

As eSports continue to grow in popularity and economic importance, the ability to automatically analyze player emotions will become increasingly valuable for game developers, coaches, broadcasters, and researchers. This study provides a foundation for such applications while opening up numerous avenues for future research. Ultimately, by advancing our understanding of how emotions manifest in competitive gaming contexts, this work contributes to the broader goal of creating more engaging, responsive, and psychologically informed interactive experiences.

References

[1] A. Moors, P. C. Ellsworth, K. R. Scherer, and N. H. Frijda, "Appraisal theories of emotion: State of the art and future development," *Emotion Review*, vol. 5, no. 2, pp. 119–124, 2013. 8, 9

[2] J. P. Jokinen and J. Silvennoinen, "The appraisal theory of emotion in human-computer interaction," in *Emotions in Technology Design: From Experience to Ethics*, pp. 27–39, Springer, 2020. 8

[3] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, 1978. 8, 10

[4] C. Ringer, J. A. Walker, and M. A. Nicolaou, "Multimodal joint emotion and game context recognition in league of legends livestreams," in *2019 IEEE Conference on Games (CoG)*, pp. 1–8, IEEE, 2019. 8

[5] S. Roohi, E. D. Mekler, M. Tavast, T. Blomqvist, and P. Hämäläinen, "Recognizing emotional expression in game streams," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*, pp. 301–311, ACM, 2019. 8

[6] P. Mavromoustakos-Bloom, D. Melhart, A. Liapis, G. N. Yannakakis, S. Bakkes, and P. Spronck, "Multiplayer tension in the wild: A hearthstone

case," in *Proceedings of the 18th International Conference on the Foundations of Digital Games (FDG 2023)*, ACM, 2023. 8

[7] E. Xylakis, A. Liapis, and G. N. Yannakakis, "The scream stream: Multimodal affect analysis of horror game spaces," in *2024 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, IEEE, 2024. 8

[8] N. Sakirov, "Emotionality in video games: A brief exploratory study," tech. rep., Bachelor Exchange Project, Prof. Niewiadomski, 2023. 8

[9] K. R. Scherer, "What are emotions? and how can they be measured?," *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005. 8

[10] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971. 9

[11] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980. 9

[12] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994. 9

[13] M. Gjoreski, H. Gjoreski, M. Lutrek, M. Gams, and R. Niewiadomski, "Toward emotion recognition from physiological signals in the wild," *Frontiers in Psychology*, vol. 11, p. 1111, 2020. 11, 12, 13, 14

[14] P. Wilhelm and P. Grossman, "Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment," *Biological Psychology*, vol. 84, no. 3, pp. 552–569, 2010. 11

[15] X. Zhou *et al.*, "Cross-dataset emotion recognition from physiological signals using deep learning," *IEEE Transactions on Affective Computing*, 2017. 12

[16] K. D. Entertainment, "efootball [youtube channel]," 2025. Accessed: [Date]. 16