Daniela Soares de Pina Lopes

# Expressões faciais - Reconhecimento de emoções baseado em pontos de referência selecionados

# Facial expressions - Emotions recognition based on selected landmarks

Daniela Soares de Pina Lopes

# Expressões faciais - Reconhecimento de emoções baseado em pontos de referência selecionados

# Facial expressions - Emotions recognition based on selected landmarks

"*The greatest challenge to any thinker is stating the problem in a way that will allow a solution*"

— Bertrand Russell

Universidade de Aveiro
2022

**Daniela Soares de Pina Lopes**

# Expressões faciais - Reconhecimento de emoções baseado em pontos de referência selecionados

# Facial expressions - Emotions recognition based on selected landmarks

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica da Doutora Susana Manuela Martinho dos Santos Baía Brás, Investigadora no Instituto de Engenharia Eletrónica e Telemática de Aveiro, da Universidade de Aveiro, e do Doutor Ilídio Castro Oliveira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

Dedico este trabalho aos meus pais.

**o júri / the jury**

presidente / president

Professor Doutor Paulo Miguel de Jesus Dias
Professor Auxiliar, Universidade de Aveiro

vogais / examiners committee

Doutor Renato Eduardo Silva Panda
Investigador Auxiliar, Instituto Politécnico de Tomar - Centro de Investigação em Cidades Inteligentes

Doutora Susana Manuela Martinho dos Santos Baía Brás
Investigadora Doutorada (nível 1), Universidade de Aveiro

**agradecimentos / acknowledgements**

Esta dissertação assinala e conclusão de uma das etapas mais desafiantes na minha vida. Ao fim destes 6 anos, com altos e baixos, prevalece um sentimento de enorme gratidão perante todos os que me acompanharam.

Quero agradecer primeiramente aos meus pais que são as pessoas a quem tudo devo. Sempre me apoiaram incondicionalmente, sempre estiveram presentes nos bons e nos maus momentos, sempre me ampararam e depositaram em mim toda a sua confiança, sem nunca me deixarem desistir.

Um agradecimento à minha restante família, em especial aos meus avós, por todas as palavras de conforto e de força que me dirigiram, por toda a preocupação que sempre mostraram por mim e por todas a preces que depositaram em mim.

Agradeço também aos meus orientadores, Doutora Susana Brás e Doutor Ilídio Oliveira, pelo incrível trabalho de orientação, por toda a disponibilidade que sempre demonstraram, por todas as dicas e conselhos, pela forma atenciosa com que sempre me trataram e por toda a paciência que demonstraram.

Por fim, mas não menos importante, agradeço a todos os meus amigos com quem partilhei momentos memoráveis nesta jornada.

**Palavras Chave**

Computação afetiva, Reconhecimento de Expressões Faciais, Pontos de Referência na face, Processamento de imagem, Aprendizagem Automática.

**Resumo**

O reconhecimento automático de emoções baseado em expressões faciais é uma área de pesquisa ativa e desafiante, fazendo jus ao seu potencial académico e comercial. A maioria dos sistemas de reconhecimento de expressões faciais já existentes utilizam uma abordagem holística para retirar características emocionais da face, ou seja, adoptam métodos que fazem uso de todos os atributos faciais para reconhecer emoções.

Este projeto visa identificar as regiões do rosto que caracterizam cada emoção dentro de um conjunto de emoções básicas. Para tal, foi conduzida uma pesquisa para estudar os movimentos faciais que ocorrem desde a expressão neutra até ao ápice de uma determinada emoção, recorrendo à análise da evolução das distâncias euclidianas entre os pontos de referência na face ao longo dos frames de um vídeo. A partir desta análise, foram selecionados conjuntos de pontos de referência essenciais para identificar cada expressão facial. O modelo utilizado para realizar a classificação das expressões faciais foi o Support Vector Machine (SVM).

No decorrer deste trabalho foi possível verificar que o número de pontos descritores da face pode ser reduzido a um valor mínimo, possibilitando assim uma maior eficiência computacional, o uso de algoritmos mais simples e ainda, no caso de ser necessária guardar informação, preservar a privacidade das pessoas, reduzindo a probabilidade de re-identificar os intervenientes com base nas *features* selecionadas.

**Keywords**

Affective Computing, Facial Expressions Recognition, Facial Landmarks, Image Processing, Machine Learning.

**Abstract**

Automatic emotion recognition based on facial expressions is an active and challenging research field, owning to its academic and commercial potential. Most of the already existing facial expression recognition systems use a holistic approach to withdraw emotional characteristics from facial images, i.e., they adopt methods that make use of all facial attributes in order to recognize emotions.

This project aims to identify the regions of the face that characterize each emotion within a set of basic emotions. For this, research was conducted to study the facial movements that occur from the neutral expression to the apex of a given emotion, resorting to the analysis of the evolution of the euclidean distances between facial landmarks throughout the frames of a video. From this analysis, there were selected sets of facial landmarks essential to identify each facial expression. The model used to perform the facial expressions classification was the Support Vector Machine (SVM).

During this project, it was possible to verify that the number of facial landmarks can be reduced to a minimum value, promoting better computational efficiency, the usage of simpler algorithms, and also, in case it is necessary to store information, preserve people's privacy since it diminishes the probability of re-identifying the intervening subjects with only the selected features.

# Contents

# List of Figures

# List of Tables

# List of Acronyms

**ADFES-BIV** Amsterdam Dynamic Facial Expression Set - Bath Intensity Variations
**AUs** Action Units
**CLAHE** Contrast Limited Adaptive Histogram Equalization
**CNN** Convolutional Neural Network
**DCNN** Deep Convolutional Neural Network
**ECG** Electrocardiogram
**EEG** Electroencephalogram
**EMG** Electromyogram
**FACS** Facial Action Coding System
**FER** Facial Expressions Recognition
**HOG** Histogram of Oriented Gradient
**MLP** Multi-Layer Perceptron
**LBP** Local Binary Pattern
**LR** Logistic Regression
**SVC** Support Vector Classifier
**SVM** Support Vector Machine

CHAPTER 1

# Introduction

*"Computers are able to see, hear and learn. Welcome to the future."*

— *Dave Waters*

## 1.1 Context

Nowadays, technological systems are seen as an extension of human beings. The necessity of continuously building and improving human-computer interactions is undeniable. Hence, the interest in studying human behavior and responses to certain situations has been increasing in order to enhance affective systems.

Emotions are a crucial part of our everyday life. They influence decision-making and help us understand and be understood by others [1]. A significant segment in the perception of one's emotional state lies in assessing facial expressions [2] [3].

Facial expressions play a vital role in nonverbal communication. Humans show an enormous proficiency in reading the face and inferring emotional states, like joy, sadness, and anger [4].

Several facial expression recognition systems have been developed throughout the years to mimic the human ability to distinguish emotions, with several application areas, such as marketing, healthcare services, customer services, and education.

Most of the current methods make use of holistic representations of facial characteristics. Although this technique shows outstanding results, there is no full understanding of the important aspects of the face that allow a model to learn to distinguish human facial expressions. Some other approaches focus on a few facial structures considered to deliver the highest emotional content, namely the eyes, mouth, eyebrows, nose, and jawline. Still, in these cases, all these facial structures are used to train the model in any emotion.

The recognition of emotional states in video conference scenarios is also an active research area at the Institute of Electronics and Informatics Engineering of Aveiro (IEETA), from which the proposal for this dissertation emerged.

## 1.2 Objectives

Understanding the relevance of particular regions of the face during the expression of a certain emotion is important to understand the human mechanisms and ability to perceive emotions through the analysis of facial expressions. Therefore, this dissertation has three main goals:

1. Elicit the facial regions with the most significance during the expression of a specific emotion;
2. Evaluate which facial landmarks - that are used to map the facial structures - show a significant amount of displacement during the evolution of each emotion;
3. Specify the most significant region overall, i.e., the most expressive area of the face transversal to all emotions.

## 1.3 Outline

This document is divided into five chapters and is structured as follows:

- **Chapter 1 - Introduction:** presents an overview of the document, introducing the context to the problem and describing the main objectives;
- **Chapter 2 - Background concepts and related literature:** introduces important background concepts for this project, as well as existing approaches relative to machine learning in emotions recognition;
- **Chapter 3 - An exploratory study of facial landmarks:** describes the process of evaluating each facial region in each emotion and presents the selected landmarks to be used in the classification process;
- **Chapter 4 - Facial expressions classification:** includes a description of the features extracted from the facial landmarks and the model used for classification, as well as the results obtained;
- **Chapter 5 - Discussion and conclusions:** presents the final thoughts on the dissertation alongside possible future work proposals.

CHAPTER 2

# Background concepts and related literature

This chapter provides an overview of the background concepts and literature on the recognition of emotions.

## 2.1 Affective computing

Affective computing intends to train computers with human-like abilities. It is an emerging interdisciplinary research field spanning several domains of cognitive science, such as psychology, sociology, computer science, mathematics, physiology, and linguistics [5] [6].

The field of affective computing has multiple application areas. For example, several methods were developed in the education field to stimulate better learning techniques by monitoring the students' emotional states [7]. Also, in healthcare, diverse systems were designed to detect and monitor diseases, for instance, through wearable devices, like smartwatches, or video data and text mining approaches to detect conditions like depression or suicidal speech, among other research areas [8] [9] [10] [11]. Another area where affective computing is widely used is in marketing, where businesses monitor customers' reactions to new products, campaigns, and services to optimize their communication strategies [12].

The emotional state of a person can be detected from a variety of behavioral signals, such as facial expressions, voice, text, and body gestures, and physiological signals, for instance, the heart rate, skin temperature, Electrocardiogram (ECG), Electromyogram (EMG) and Electroencephalogram (EEG) [5]. This dissertation is focused on detecting and studying behavioral signals, more precisely, the recognition of emotions through facial expressions.

## 2.2 Emotion and its main properties

In [13], emotion is defined as "a complex and intense psycho-physiological experience of an individual's state of mind when reacting to biochemical (internal) and environmental

influences (external)". Therefore, emotions are a response to *stimuli*. They are instinctual and involuntary because they get triggered by specific events beyond human control, and how we respond also becomes reflexive [14].

Emotional experiences have three key elements: the subjective experience, the physiological response, and the behavioral response. The subjective experience also referred to as *stimulus*, is what triggers an emotion, like losing a loved one or getting a compliment. Then comes the physiological response, for example, an increased heart rate. Finally, the behavioral response happens, signaling to other people how we feel through facial expressions, voice, and hand or body gestures [14].

From the group of behavioral responses, verbal components convey one-third of human communication, and nonverbal components convey two-thirds [1]. Among several nonverbal components that carry emotional meaning, facial expressions are one of the main informative channels in interpersonal communication [2].

Faces are a ubiquitous part of everyday life for humans. The ability to perceive faces is one of the first capacities to emerge after birth. We know that a smiling face is usually interpreted as being happy, and a crying one may very well signal sadness [15] [14]. As an example, Table 2.1 summarizes the features we look for when assessing the facial expressions of the basic emotions for emotional cues.

Emotions are commonly divided into two groups: basic and complex emotions. Basic emotions tend to happen automatically; they are innate, universal, fast, and automatic and are usually associated with a certain facial expression. However, complex emotions are not so easily recognizable because they can differ from person to person, based, for example, on different cultures. Complex emotions are typically defined by combining two or more basic emotions [14].

Ekman and Friesen [16] proposed the Facial Action Coding System (FACS), which is a comprehensive system that breaks down facial expressions into individual components of muscle movement (Action Units (AUs)) in order to describe all visually discernible facial movements. The objective was to be able to measure facial behavior to distinguish all possible visible anatomically-based facial movements [17].

The facial AUs code the fundamental actions (46 AUs) of individual or groups of muscles typically seen when producing the facial expressions of a particular emotion. AUs are scored and analyzed as independent elements, but the underlying anatomy of many facial muscles constrains them so that they cannot move independently of one another, which generates dependencies between AUs. Therefore, in order to recognize facial emotions, an individual action unit is detected, and the system classifies facial categories according to the combination of AUs [1] [15].

Thus, it is possible to deduce that the FACS was created to provide a standard and systematic way to categorize the physical expression of emotions.

| Emotions | Characteristics |
|---|---|
| Anger | - Eyebrows pulled down and together<br>- Eyes opened wide<br>- Lips pressed tightly together |
| Contempt | - Tightened and raised lip corner on one side of the face |
| Disgust | - Lowered eyebrows<br>- Wrinkling on the side and bridge of the nose<br>- Upper lip raised<br>- Lower lip raised and slightly protruding |
| Fear | - Eyebrows raised and pulled together<br>- Raised upper eyelids<br>- Tensed lower eyelids<br>- Jaw dropped open, and lips stretched horizontally backwards |
| Enjoyment | - Eyes are narrowed and there is some wrinkling around the eyes<br>- Cheeks are raised<br>- Lips are pulled back and teeth are exposed in smile |
| Sadness | - Inner corners of the eyebrows pulled up and together<br>- Upper eyelids drooped and eye looking down<br>- Lip corners pulled downwards |
| Surprise | - Eyebrows raised but not drawn together<br>- Upper eyelids raised, lower eyelids neutral<br>- Jaw drooped down |

**Table 2.1:** Main facial characteristics of the basic emotions, adapted from [18].

## 2.3 Emotion models

Before exploring how emotions, or more precisely, facial expressions, are detected and recognized, it is essential to understand how emotions are described.

Usually, emotions are modeled in two ways: one way is to separate emotions into discrete categories, and the other is to use multiple dimensions to label emotions.

### 2.3.1 Discrete emotion model

In 1970, Paul Ekman [19] proposed the concept of universal emotions, which affirms that there are six basic emotions - anger, disgust, fear, joy, sadness, and surprise - that transcend language, region, culture, and ethnic differences. This method is known as the discrete model of emotions.

In the '90s, Ekman expanded the list of basic emotions, including a set of positive and negative emotions, which he considered a secondary class of emotions [20].

This model categorizes emotions using word descriptors instead of quantitative analysis.

### 2.3.2 Dimensional emotion model

The dimensional model introduces a quantitative analysis of emotions, providing ways to label a broader range of emotional states.

Russell [21] presented the circumplex model in 1980, in which an affective state is represented in a two-dimensional space, and the proposed dimensions are valence and arousal (Figure 2.1). Valence represents emotions' pleasure level, ranging from negative (unpleasant) to positive (pleasant), whereas arousal expresses how intense emotions are felt, varying from passive (low) to active (high).

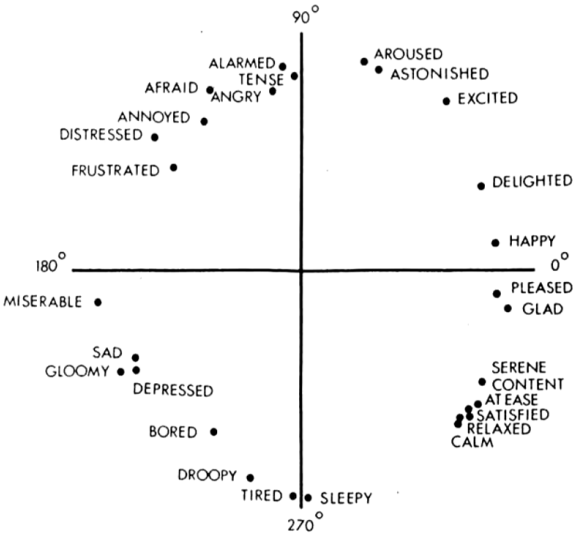



**Figure 2.1:** Russell's circumplex model [22]

Besides the bi-dimensional model, Mehrabian and Russell introduced a tri-dimensional model called the Valence-Arousal-Dominance (VAD) or Pleasure-Arousal-Dominance (PAD) [23] [24]. Figure 2.2 shows the distribution of the six basic emotions defined by Ekman within the tri-dimensional model.

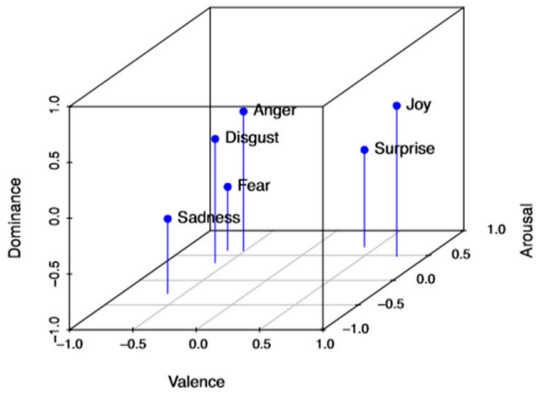



**Figure 2.2:** The VAD (Valence-Arousal-Dominance) model spanned across the six basic emotions [25]

## 2.4 Machine learning techniques for facial expressions recognition

The research interest in Facial Expressions Recognition (FER) has been fast growing. Researchers have been focusing on the recognition of the six basic emotional expressions, namely anger, disgust, fear, happiness, sadness, and surprise. Exploring the current work in

this field makes it possible to divide FER techniques into two categories: conventional and deep learning (DL)-based methods.

Conventional FER systems are usually deployed in three steps [1]:

1. **Face and/or facial components detection:** the faces are detected from an input image, and facial components (e.g., eyes, nose, and mouth) or landmarks are detected from the face region.
2. **Feature extraction:** spatial and temporal features are extracted from the facial components.
3. **Facial expressions classification:** the facial expressions are recognized using machine learning algorithms.

Qiu et al. [26] proposed a method to recognize emotions that relies solely on facial landmarks. In this work, 68 landmarks were detected to map the facial structures and compute spatial feature vectors with the distances between these landmarks. Then, a Multi-Layer Perceptron (MLP) classifier was used for expression recognition. Other similar works are presented in [27] and [28], where feature vectors were also computed from a group of 68 landmarks, and the facial expressions classification was done using Support Vector Machine (SVM) and Random Forest classifiers.

Ghimire and Lee [29] used a scheme of 52 landmarks to extract the angle and position of the face in image sequences. First, it was calculated the angle and the Euclidean distances between each pair of landmarks in a frame. Then the distances and angles of each frame were subtracted from the corresponding distances and angles in the first frame of the video sequence. For the classification, two methods were presented: a multi-class Adaboost and SVM with Boosted Features.

On a different note, Happy et al. [30] proposed a method that uses a Haar cascade classifier to detect the face and extracts features using a Local Binary Pattern (LBP) histogram of different block sizes of a face image as a feature vector.

While Happy et al. used a holistic representation of the face as a feature vector, Ghimire et al. [31] extracted region-specific features by dividing the entire face into domain-specific local regions. The relevant areas were defined using an incremental search approach, reducing feature dimension and improving recognition accuracy. This way, it is possible to affirm that different face regions have different levels of importance.

Deep learning algorithms have been gaining popularity in recent years and showing promising results in FER by blending end-to-end automatic feature extraction and classification into one step.

The most prominent algorithm in FER systems is the Convolutional Neural Network (CNN) since it specializes in processing data with a grid-like topology, which is the case of images. A CNN is defined by three types of heterogenous layers: convolution layer, max pooling layer, and fully connected layer [32].

Li et al. [33] proposed a method to recognize micro-expressions from videos using a 3D flow-based CNN, which extracts deeply learned features directly from images. Abdulsalam

et al. [3] proposed a Deep Convolutional Neural Network (DCNN) model to perform an end-to-end classification of ten emotional classes from videos.

Lopes et al. [34] proposed a solution that characterizes facial expressions using a combination of CNN and specific image pre-processing steps. Brewer and Kimmel [35] also used CNN visualization techniques to study and understand the relation between the feature maps these computational algorithms are using and the FACS method. In this work, Brewer and Kimmel reiterate the effectiveness of the method proposed by Ekman and demonstrate once again that specific regions on the face offer higher importance when perceiving emotional states.

CHAPTER 3

# An exploratory study of facial landmarks

## 3.1 Purpose of the study

Facial expressions convey lots of nonverbal information that is important to determine one's emotional state [13]. We tend to focus on different regions of the face while trying to perceive the emotion expressed by an individual.

Bearing this in mind, it was considered relevant to conduct an experimental study whose main goals are to determine which facial landmarks are the most instructive in each facial expression and understand which areas of the face better characterize a particular condition. This allows us to understand which key points and areas of the face should be analyzed to distinguish each facial expression and also determine which is the most expressive area of the face.

The data used in this investigation comes from the Amsterdam Dynamic Facial Expression Set - Bath Intensity Variations (ADFES-BIV) data set [36].

ADFES-BIV [36] gathers a set of video *stimuli* portraying three levels of emotional expressions, from low to high intensity. This data set consists of 10 facial expressions (anger, contempt, disgust, embarrassment, fear, happiness, neutral, pride, sadness, and surprise), and each emotion was expressed by 12 encoders: 7 males and 5 females.

Each video has 26 frames with a frame rate of 25/sec, resulting in videos of 1040 ms in length. Furthermore, all videos start with a neutral facial expression and continue until the end of a given expression, allowing to use the neutral expression as a point of reference.

Since we are dealing with videos, it is possible to predict that, as the transition from a neutral expression to any other expression occurs, there will also be significant changes in the behavior of specific facial landmarks.

By observing Figure 3.1, one can state that in the videos displaying low-intensity levels of an emotion, the subjects only show smooth expressions with little facial movement. On the other hand, the videos displaying high-intensity levels of an emotion show the apex of

each particular emotion, revealing a high amount of facial movement. In this case, we want to be able to detect the facial expression before its apex, but not at a moment where it is still difficult to discern the emotion expressed. For this reason, this study uses videos with intermediate intensity levels only.

**(a)** Neutral expression **(b)** Low intensity **(c)** Intermediate intensity **(d)** High intensity

**Figure 3.1:** The last frame of each one of the three videos of a female participant portraying the three different intensity levels of anger facial expression.

## 3.2 Procedure

As already described in the previous chapter, the FACS is the most popular method to describe facial expressions, using combinations of AUs.

Taking this into account, it was used the *Dlib* library [1] to map the key points of the face and replicate the effects of Ekman's AUs.

Each video went through a set of stages, from the video splitting into static frames up to the extraction of the desired metrics. This process is illustrated in Figure 3.2.

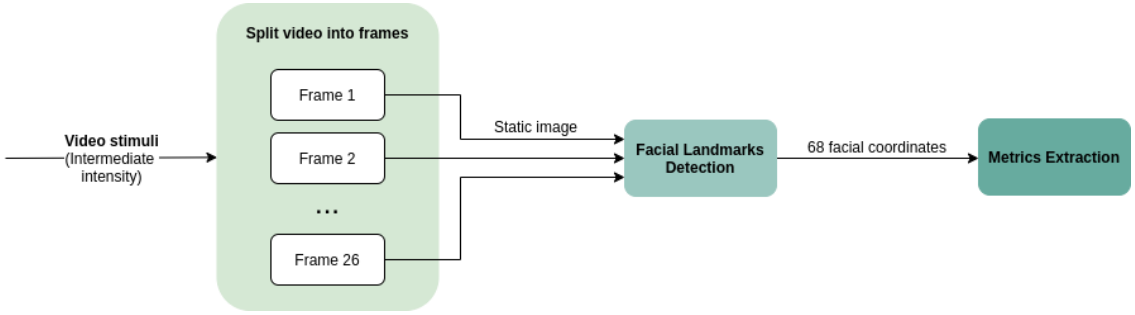



**Figure 3.2:** Diagram illustrating the methodology used to extract the necessary metrics.

### 3.2.1 Splitting the video into static frames

Splitting each video into static frames was the first stage of this research. This was easily done using the *OpenCV* library that already provides methods to perform this task, by creating *VideoCapture* objects for each video.

Each video resulted in 26 frames. All the static images were stored in categorized folders and then used for the facial landmarks detection stage.

[1] http://dlib.net/

### 3.2.2 Facial landmarks detection

After having all the static frames, the next stage was to detect the facial landmarks.

Facial landmarking is a computer vision task used to track the salient regions of the human face. This technique can have various applications, such as human head pose estimation or face replacement. In this case, this method was used to help measure the amount of movement that occurred on the face while a person was expressing an emotion.

At this stage, the human face and facial landmarks detection were done using the Python wrapper for *Dlib* [37], which is an open source library written in *C++*. *Dlib* provides numerous machine learning algorithms, including classification, regression, clustering, data transformation, structured prediction, and other functionalities like image processing, numerical algorithms, etc.

In this case, *Dlib* is used to estimate the location of 68 key points that map the facial structures. A visualization of the 68 facial landmarks detected with *Dlib* library is presented in Figure 3.3.

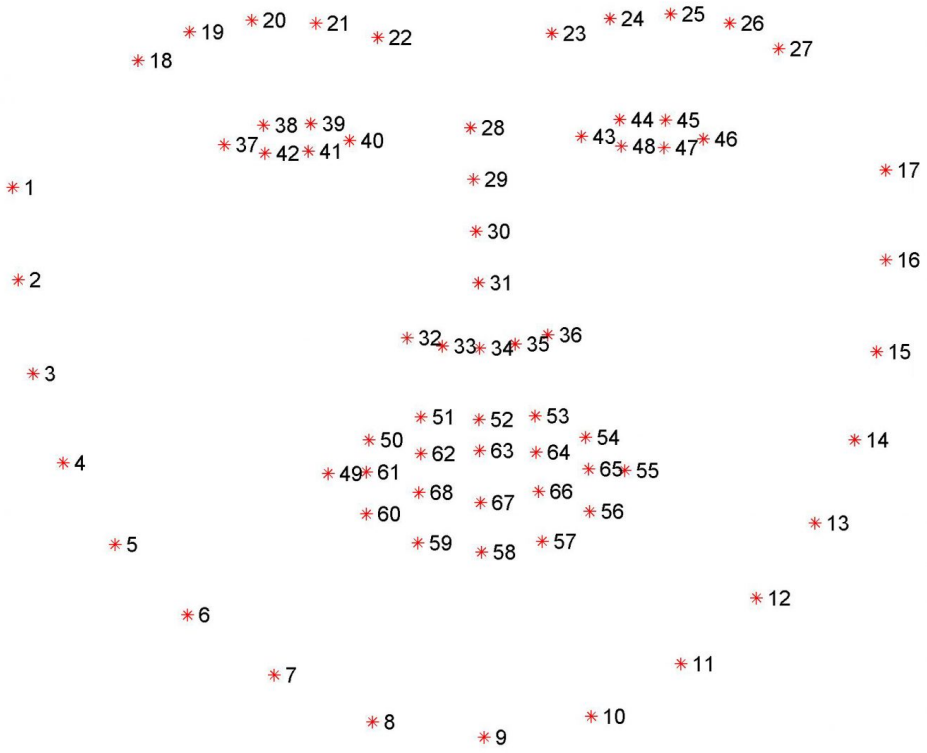



**Figure 3.3:** The 68 facial landmarks detected with *Dlib* library.

In some cases, the landmarks detection might occur under complex lighting conditions. To assure the lighting equalization of the images, each static frame was converted to grayscale (Figure 3.4b). Then, the Contrast Limited Adaptive Histogram Equalization (CLAHE) filter was applied. In CLAHE, the image is divided into small blocks, called "tiles", and later it is applied normal histogram equalization in each one of these blocks. In Figure 3.4c it can be seen the application of CLAHE filter over the grayscale image.

The last step involves detecting the human face and the facial landmarks.

The identification of human faces on the images was achieved using the built-in *Dlib* function *get_frontal_face_detector*, which returns an object *detector* that can be used to identify faces in an image. This detector was developed using the Histogram of Oriented Gradient (HOG) feature, to extract important features from the image, combined with a linear SVM classifier, to determine if there is a face in that image or not, based on the information provided by the previously extracted features.

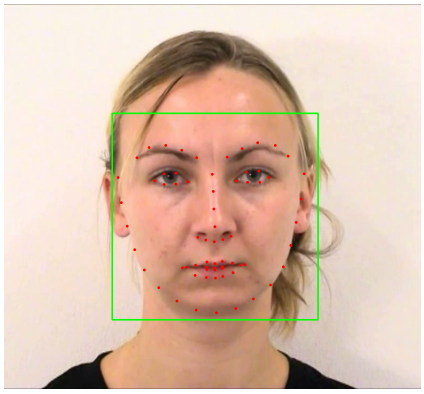
After detecting the face, the next measure was to detect the facial landmarks. For that, it was also used a *Dlib* built-in function (*shape_preditor*). The location of the 68 (x,y)-coordinates that map to facial structures on the face is achieved by using the popular pre-trained model *shape_predictor_68_facial_landmarks*. The final result of this process can be observed in Figure 3.4d.

The facial landmarks coordinates were stored in separate files to be used in the metrics extraction stage.

**(a)** Original image

**(b)** Grayscale image

**(c)** CLAHE filter applied over grayscale image



**(d)** Detection of the human face and facial landmarks

**Figure 3.4:** Different stages of facial landmarks detection process.

### 3.2.3 Metrics extraction

Having all the facial landmarks coordinates, it is now viable to extract metrics that allow analyzing and evaluating the evolution of each emotion, starting from a neutral expression.

When considering which metrics to extract, the entropy and the Euclidean distance were the first to be selected. By definition, the entropy measures the degree of disorder in a system,

which we assumed is suitable for the problem at hand. However, this metric conduced to an evaluation that does not accomplish the study proposals, and was not suitable for frames description. Thus, it was decided to proceed only with the computation of the Euclidean distances.

First of all, it is essential to explain how the data were organized in order to obtain the desired outcome. To better understand the process that the data went through, Figure 3.5 shows a diagram that illustrates the steps taken, from reading the facial landmarks coordinates of each frame of a given video up to the extraction of the desired metrics.

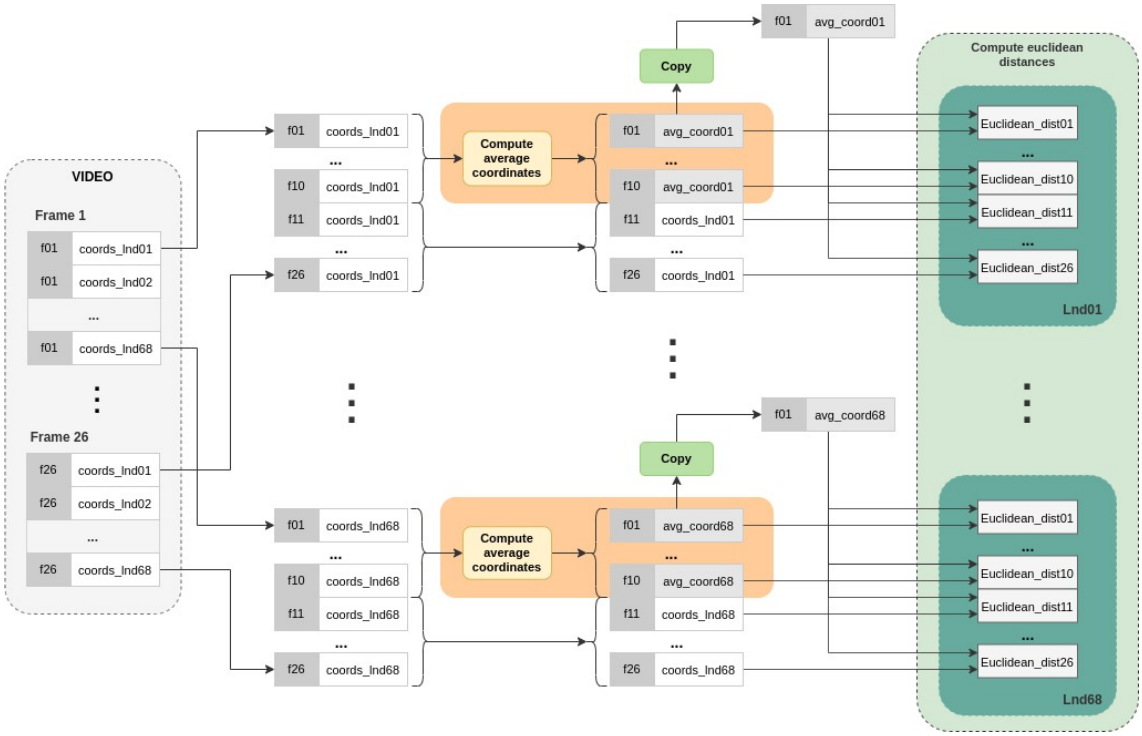



**Figure 3.5:** Metrics extraction steps.

The coordinates belonging to each frame were stored in order in separate files and categorized folders. Since each video has 26 frames, there are 26 files with 68 (x, y)-coordinates for each participant and emotion.

The goal is to determine the Euclidean distance between coordinates of the same landmark throughout the several frames, always using the coordinate of the first frame as a reference point. For example, the Euclidean distances of landmark 1 of a given video are always calculated between the coordinate of landmark 1 of the first frame (Frame 1) and the coordinates of landmark 1 of each frame, including the initial frame as well. For this to be achievable, the data had to be re-arranged, as it is shown in the first interaction of the diagram depicted in Figure 3.5.

This way, it is possible to obtain a stochastic process of the Euclidean distances between the first and the last frame of each video and examine the evolution of each landmark.

### 3.3 Evaluation of the most informative landmarks in each emotion

Determining the most informative facial landmarks results not only from the graphical analysis of the values of Euclidean distances obtained in the metrics extraction phase but also from an exhaustive analysis of all the videos.

As already mentioned, all videos start with a neutral facial expression that then evolves into an expression of anger, contempt, disgust, embarrassment, fear, joy, pride, sadness, or surprise.

During the assessment of each video, it was possible to point out the frame where the transition from the neutral expression to one of the other expressions mentioned begins. We call it the activation frame. It is after this activation frame that the most significant facial movements occur. The activation frame of each participant in each emotion is shown in Table 3.1.

| **Subjects** | **Classes** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Anger | Contempt | Disgust | Embarrass | Fear | Joy | Pride | Sadness | Surprise |
| F01 | 22 | 21 | 20 | 21 | **8** | 21 | 20 | 18 | 19 |
| F02 | 20 | 19 | 20 | 20 | 19 | 20 | 11 | 15 | 21 |
| F03 | 12 | 15 | 20 | 15 | 19 | 21 | 11 | 14 | 19 |
| F04 | 21 | 19 | 19 | 19 | 19 | 16 | 14 | 14 | 19 |
| F05 | 11 | 15 | 18 | 14 | 21 | 23 | **6** | 15 | 14 |
| M02 | 16 | 21 | 20 | 12 | 23 | 19 | **2** | 14 | 22 |
| M03 | 19 | 21 | 20 | 18 | 21 | 19 | 12 | 11 | 20 |
| M04 | 12 | 20 | 21 | 17 | 19 | 20 | **8** | 15 | 21 |
| M06 | 14 | 17 | 20 | 21 | 20 | 19 | 10 | 16 | 17 |
| M08 | 20 | 18 | 19 | **1** | 18 | 12 | **6** | 11 | 21 |
| M11 | 14 | 19 | 19 | **4** | 19 | 18 | 10 | 18 | 22 |
| M12 | 14 | 19 | 19 | 17 | 19 | 18 | **8** | 13 | 22 |

**Table 3.1:** Activation frame of each subject in each emotion.

The facial regions that showed alterations through the emotional manifestation were also registered during this process, allowing us to get an overview of the most meaningful areas of the face in each condition. By looking at the facial landmarks depicted in Figure 3.3, it is possible to identify seven facial areas/regions:

- Jawline (landmarks 1 to 17);
- Left eyebrow (landmarks 18 to 22);
- Right eyebrow (landmarks 23 to 27);
- Nose (landmarks 28 to 36);
- Left eye (landmarks 37 to 42);
- Right eye (landmarks 43 to 48);
- Mouth (landmarks 49 to 68);

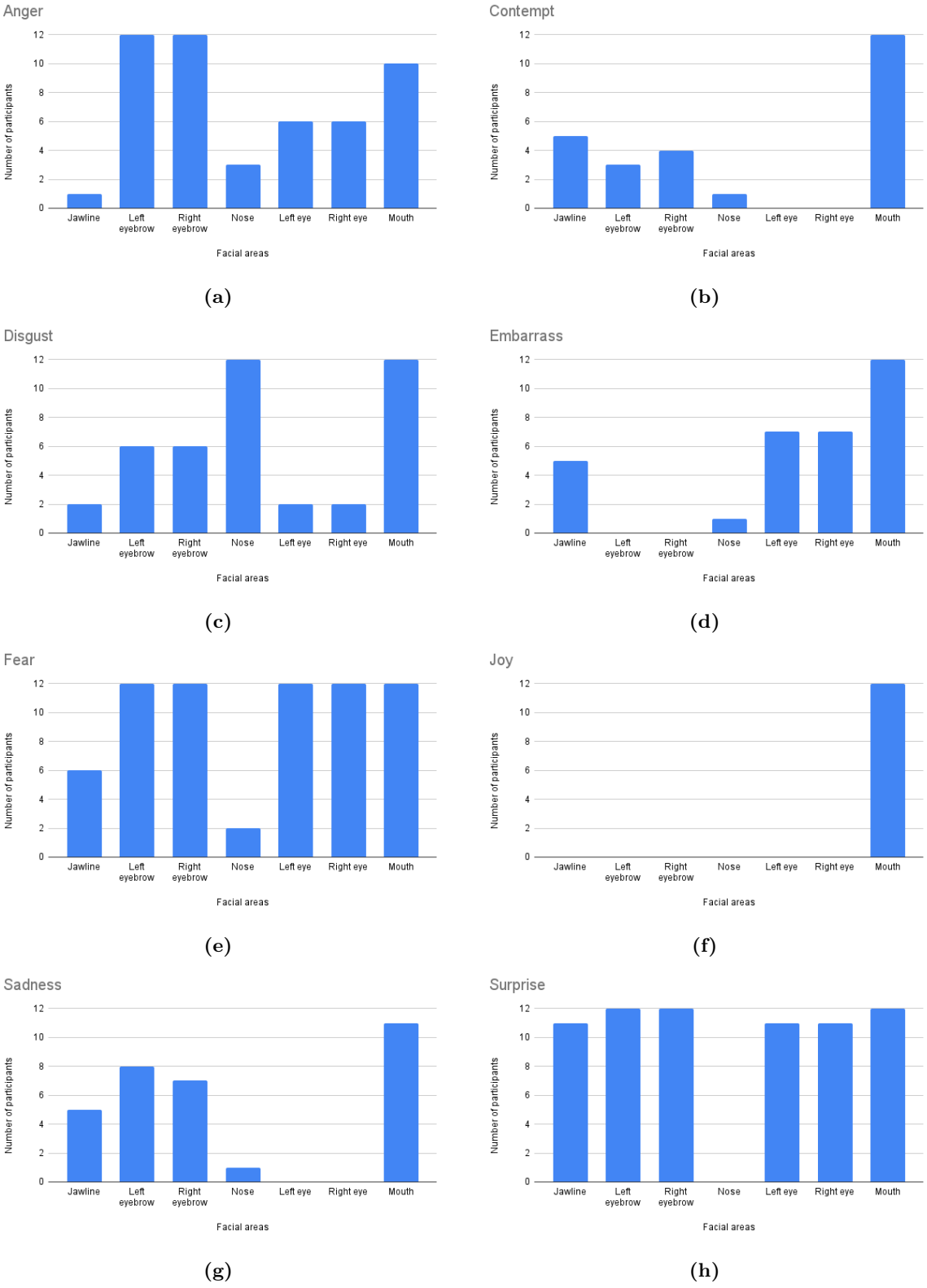



**Figure 3.6:** Facial regions that experience more alterations on each emotion after an assessment of all videos.

Figure 3.6 shows which areas experienced the evidenced changes in each emotion, summarizing the data for all participants. This graphical evaluation does not include pride's

facial expression because, in this case, all facial areas experience changes. This condition is characterized by the upward movement of the head, not specific regions of the face. For this reason, it was decided to drop out of this facial expression analysis in particular.

Now using the extracted Euclidean distance values, it is possible to make a similar analysis. Once again, the goal is to analyze which facial areas have experienced changes in each condition and verify if these results resemble those obtained by the assessment of the videos.

Since there are 68 landmarks to analyze in each static frame and the process of selecting the most informative landmarks in each condition is the same in all cases, only the analysis of one representative case will be discussed.

Taking the anger condition as example.

Figure 3.7 shows charts describing the information of each facial region, considering the facial expression of anger. Each chart corresponds to the average of the Euclidean distances in each frame of the landmarks of the analyzed region and aggregates the information of all participants. The blue shaded band represents the respective standard deviation.

Looking at Table 3.1, one can notice that in all the cases, except for some particular situations, the participants started the transition from the neutral expression to any other condition after frame 10. Thus, to emphasize the increase of the Euclidean distances of specific landmarks and perhaps reduce some of the noise caused by the subject's natural movements during the neutral expression (like breathing or some small involuntary movements of the head), it was decided to calculate the average value of the landmarks coordinates of the first ten frames. This explains the straight horizontal line in the first 10 frames in the following plots describing the evolution of the Euclidean distances.

A threshold was computed to establish the point at which the change from the neutral expression to one of the other facial expressions starts. This threshold was defined based on the emotion activation frame identified by visual evaluation of the emotional videos. For each emotion, the landmarks belonging to the regions that experience changes were selected, and then only the frames from the activation frame onward of each of the subjects were selected. From this set of data, four descriptive statistical values were computed: mean, median, standard deviation, and interquartile range of the Euclidean distances.

In most cases the mean and median values are very similar, which indicates that the Euclidean distances present a symmetrical distribution of values. However, there are 2 cases (disgust and embarrassment) with a high standard deviation due to outliers that cause the dispersion of the data.

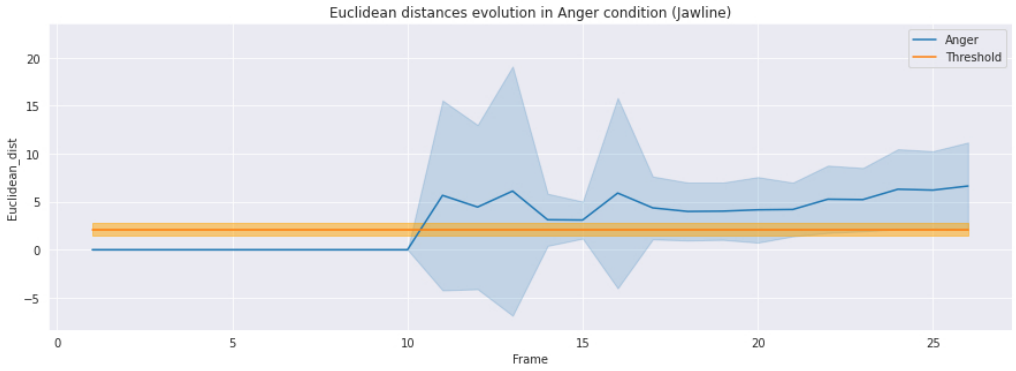
Since the mean and standard deviation have already been used to describe the behavior of Euclidean distances between facial landmarks along with the frames, it was decided to use the same approach for the threshold. Therefore, in the plots of the next figures, the threshold is represented by the orange line and the same colored band is the associated standard deviation.

Looking at Figure 3.6a, it is possible to observe that the changes occur in the eyebrows, mouth, and eyes regions. Comparing this with the results of Figure 3.7, we can conclude that those regions indeed experienced changes. Furthermore, it is noticeable that the eyebrows area experienced more emphasized variations than the mouth and eyes area since it presents

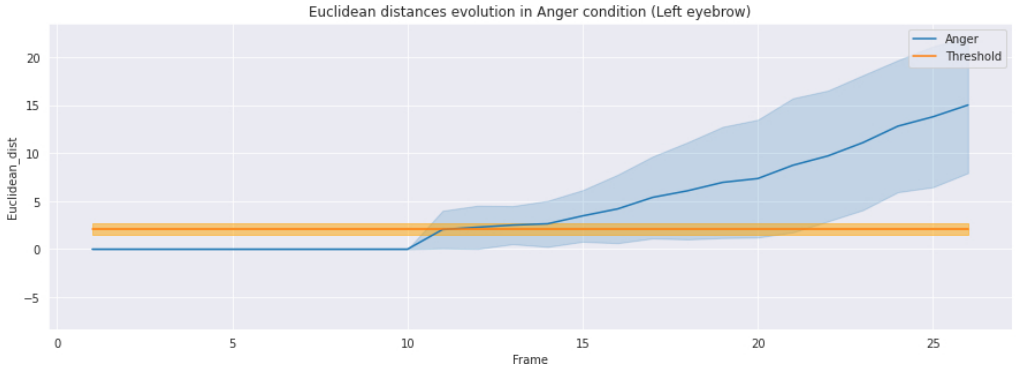
a higher evolution of Euclidean distances. This is also due to the fact that it is the most used area by the participants when they express anger.

Although the mouth region is also one of the most used to express anger, according to the video's assessment process, the evolution of Euclidean distances in this area is not so evident because the mouth movements are more delicate. The same goes for the eye regions.

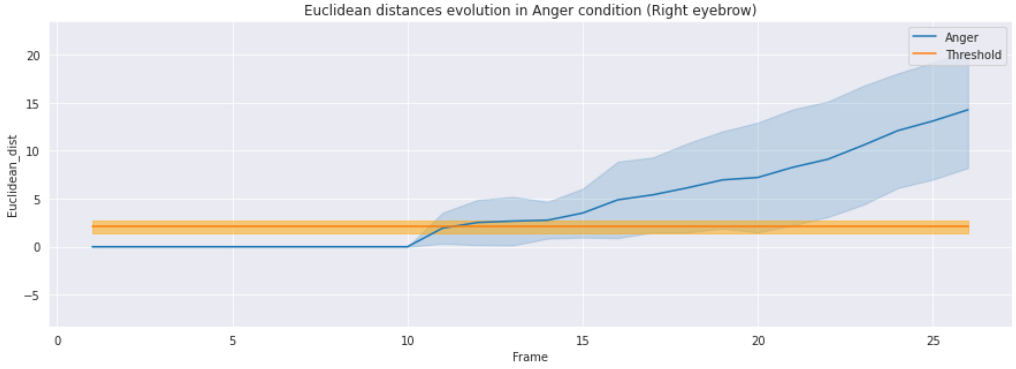
Just by analyzing the plots in Figure 3.7, the areas of the jawline and nose also seem to experience slight changes, but Figure 3.6a shows that these regions were less used by the subjects while expressing anger. Thus, as less than half of the participants used the jawline and nose areas to express anger, it was decided that these regions are not essential to characterize this condition.



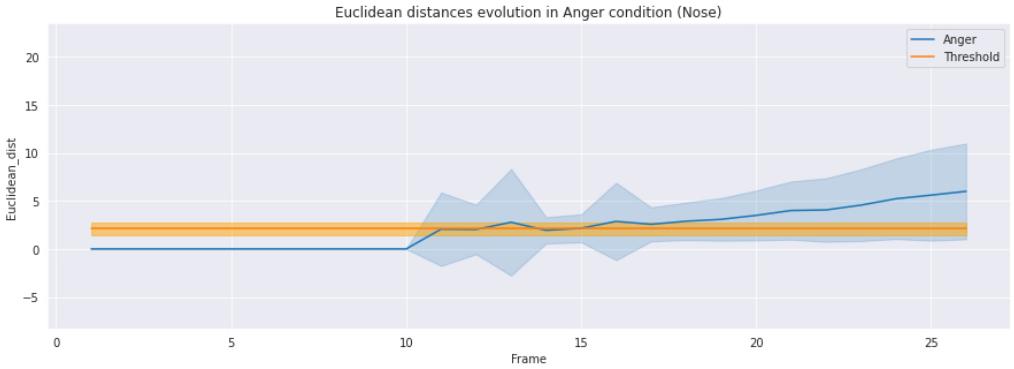

**(a)** Euclidean distances evolution of the jawline region in the facial expression of anger.



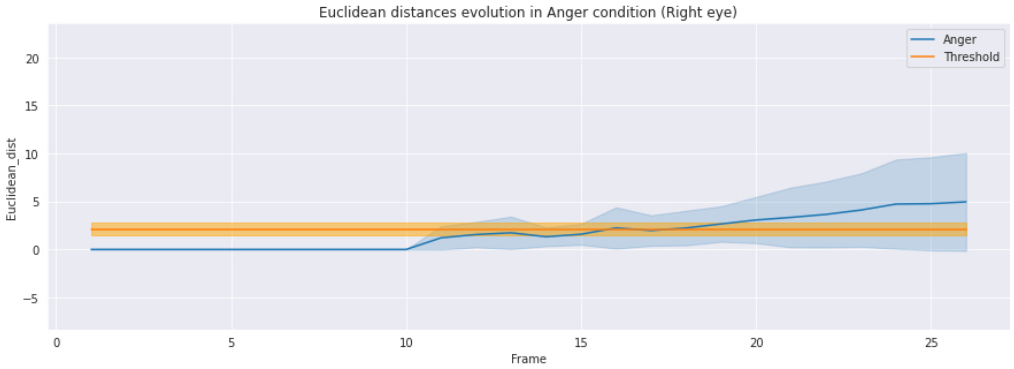

**(b)** Euclidean distances evolution of the left eyebrow region in the facial expression of anger.



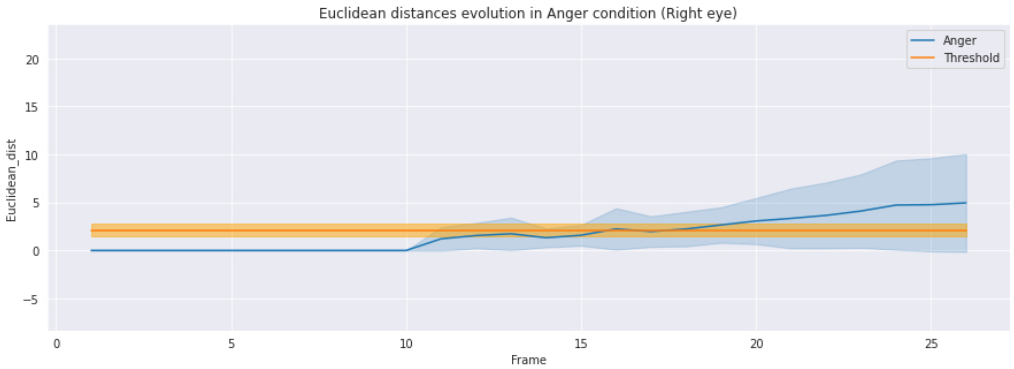

**(c)** Euclidean distances evolution of the right eyebrow region in the facial expression of anger.



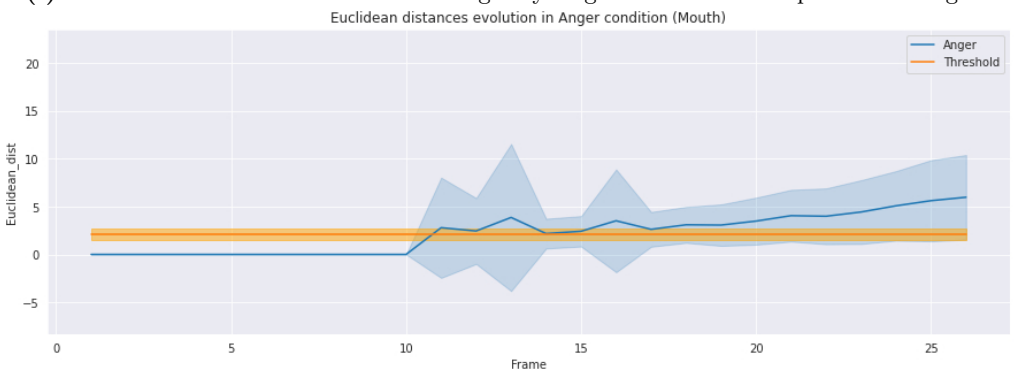

**(d)** Euclidean distances evolution of the nose region in the facial expression of anger.




**(e)** Euclidean distances evolution of the left eye region in the facial expression of anger.




**(f)** Euclidean distances evolution of the right eye region in the facial expression of anger.



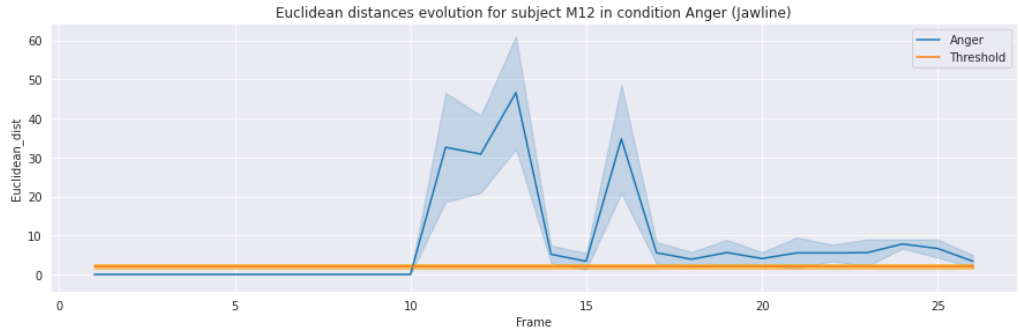

**(g)** Euclidean distances evolution of the mouth region in the facial expression of anger

**Figure 3.7:** Plots representing the evolution of the Euclidean distances by facial region in the facial expression of anger.

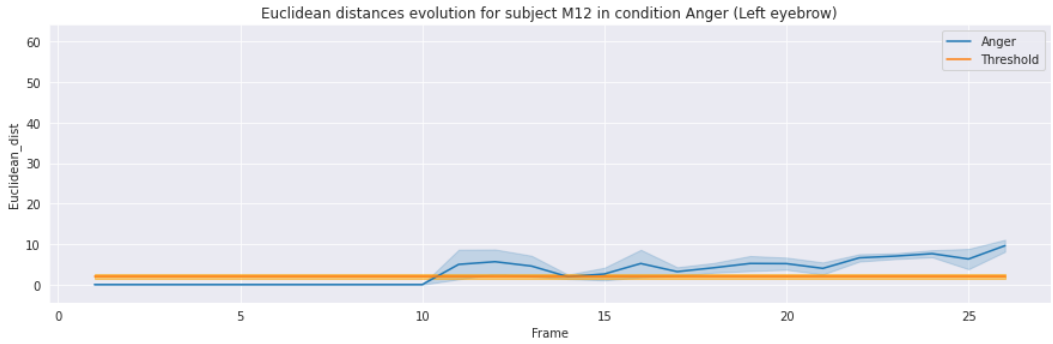
There are some cases where subjects present Euclidean distances values beyond normal, causing irregularities in the data that later become evident in the graphical results. An example of one of these cases is shown in Figure 3.8 and happens with subject M12 in the anger condition. These events may occur due to errors in the detection of landmarks and are treated as outliers.

This subject's behavior explains the oscillations witnessed in Figure 3.7, in the jawline, nose, and the mouth areas, since the peaks of Euclidean distances observed in specific frames are relatively high. However, the impact of this subject's data does not invalidate the results obtained or the conclusions retrieved. As evidence, Figure 3.9 shows the same plots presented in Figure 3.7, but without the data of subject M12. Comparing the two sets of plots, it is possible to state that the results do not experience noteworthy changes.

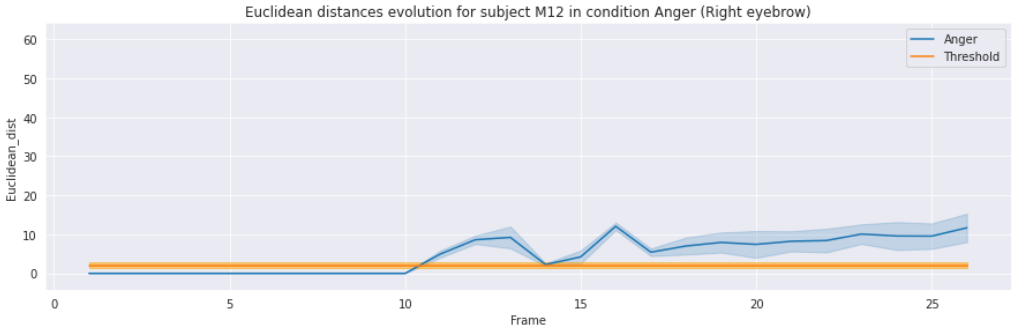
On the other hand, there was in fact anomalous behavior of subjects while expressing other emotions that influenced the results in a negative way. The way to deal with these events was precisely to identify such behavior and remove the data related to these subjects from the analysis since the number of participants who present behavior within the expected is always higher than the ones that present unwanted behavior.



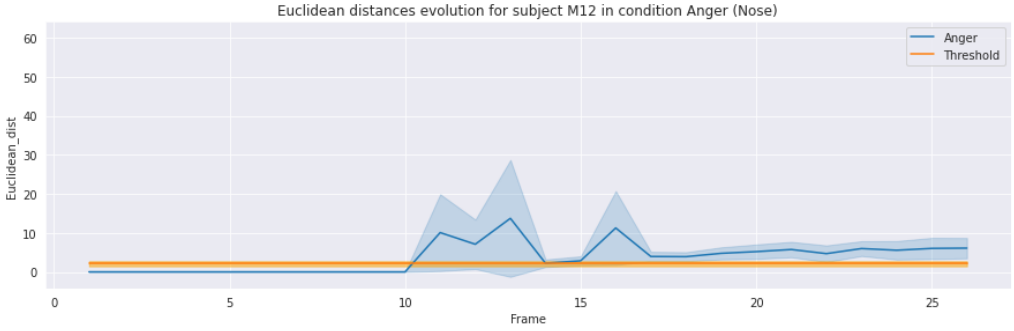

**(a)** Euclidean distances evolution on the jawline region in the anger condition



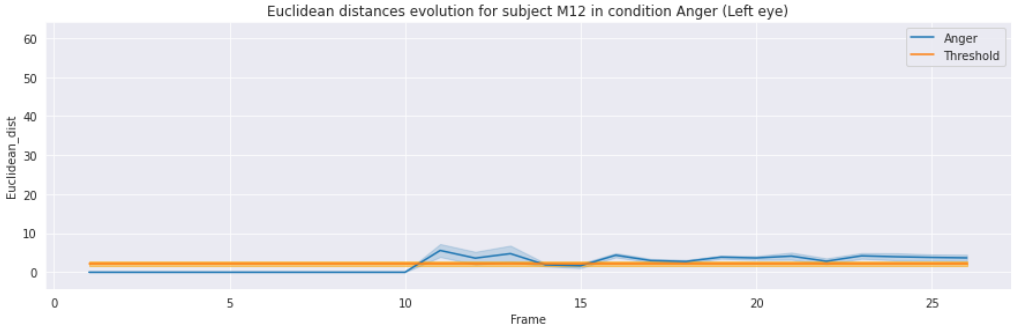

**(b)** Euclidean distances evolution on the left eyebrow region in the anger condition



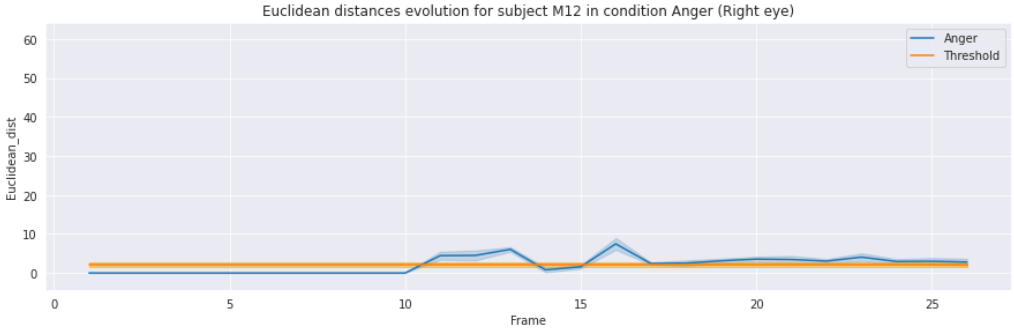

**(c)** Euclidean distances evolution on the right eyebrow region in the anger condition




**(d)** Euclidean distances evolution on the nose region in the anger condition



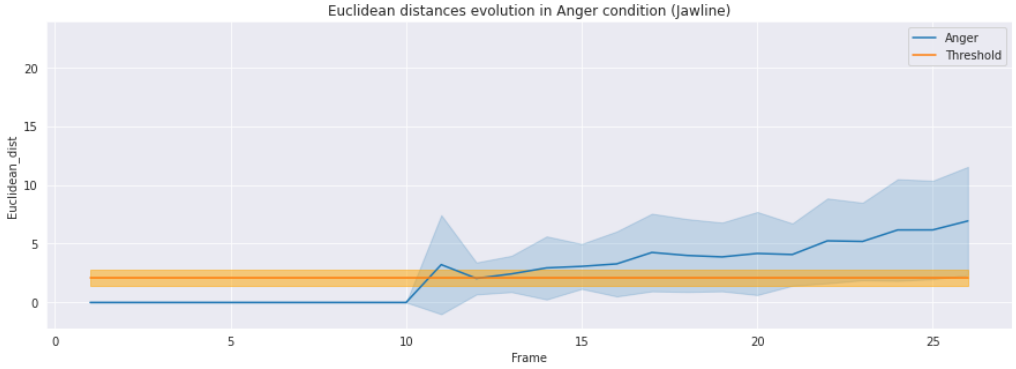

**(e)** Euclidean distances evolution on the left eye region in the anger condition



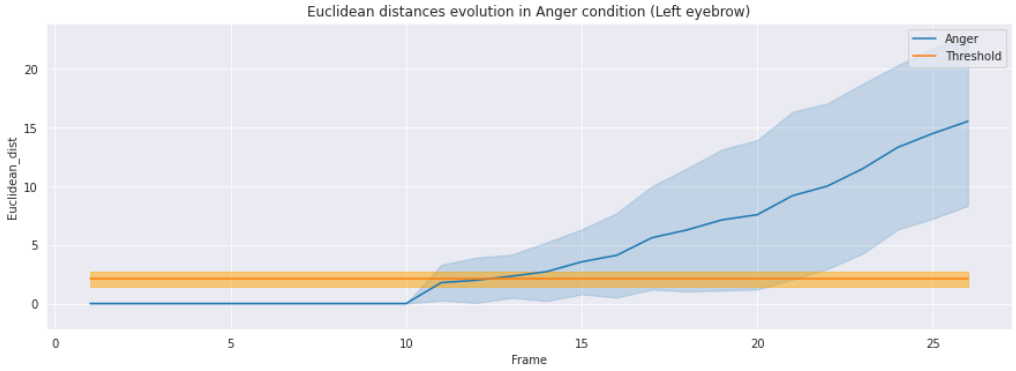

**(f)** Euclidean distances evolution on the right eye region in the anger condition


**(g)** Euclidean distances evolution on the mouth region in the anger condition

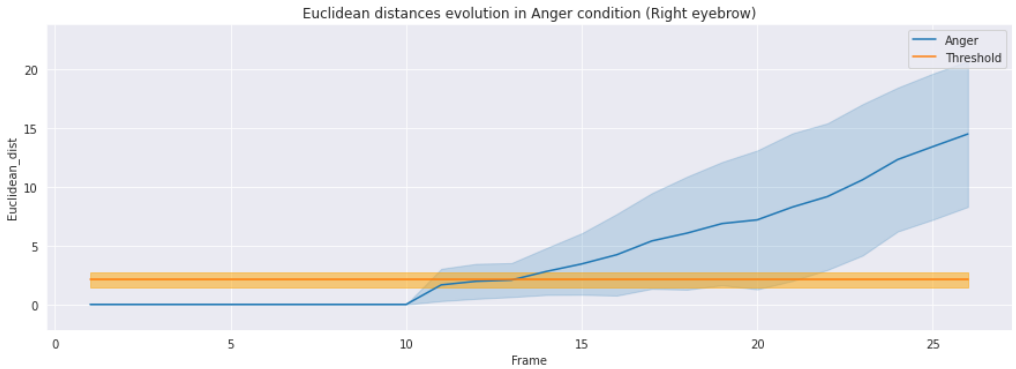
**Figure 3.8:** Plots representing the Euclidean distances evolution by facial region for a single participant (M12) in the anger condition.



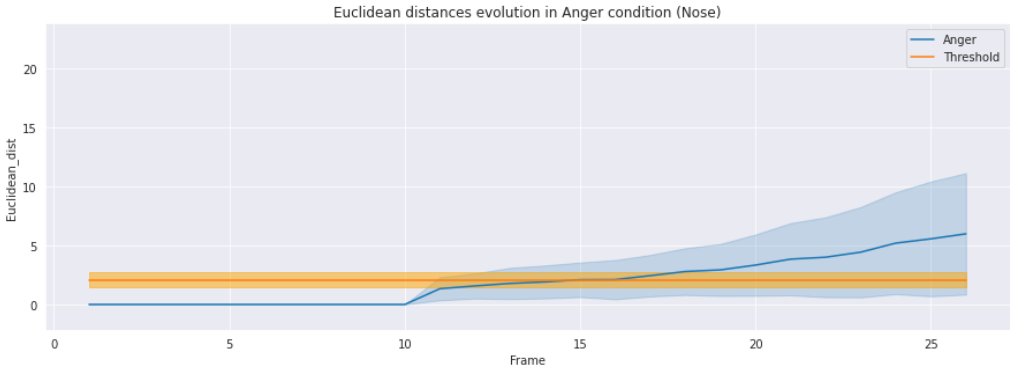

**(a)** Euclidean distances evolution on the jawline region in the anger condition



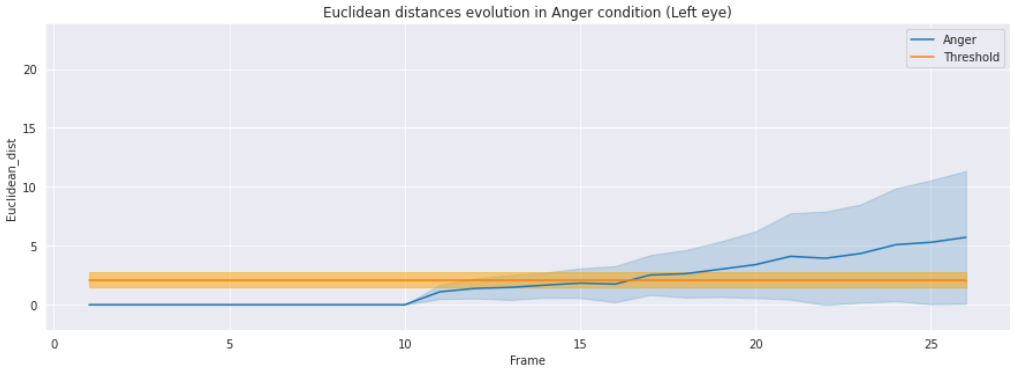

**(b)** Euclidean distances evolution on the left eyebrow region in the anger condition



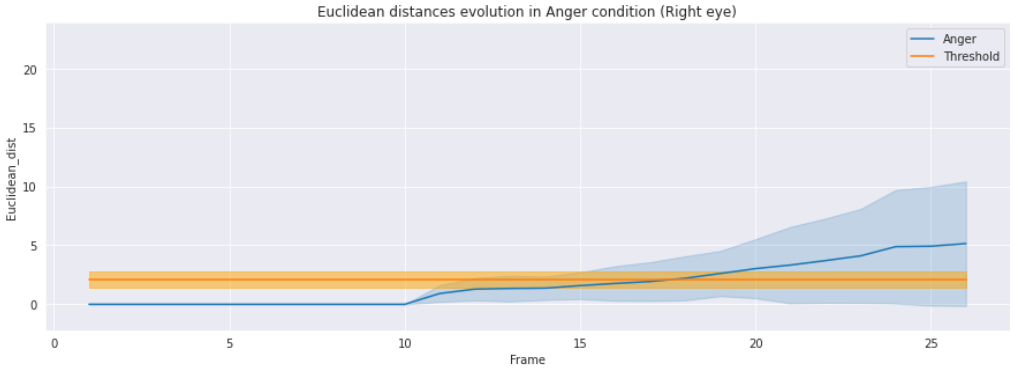

**(c)** Euclidean distances evolution on the right eyebrow region in the anger condition



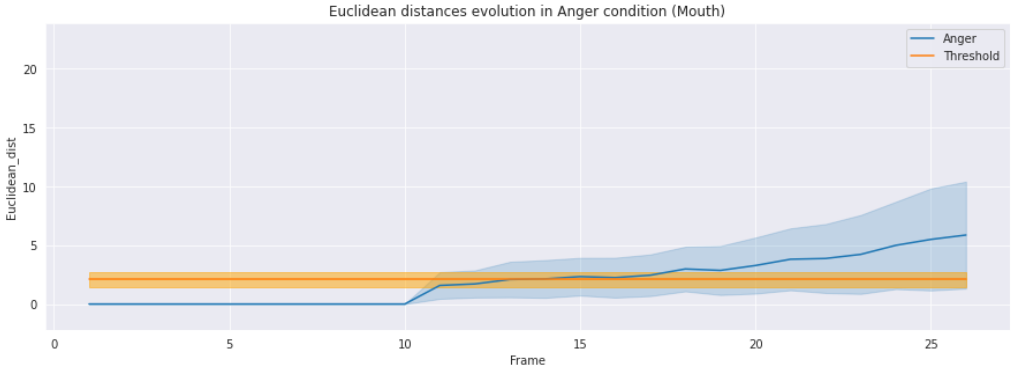

**(d)** Euclidean distances evolution on the nose region in the anger condition




**(e)** Euclidean distances evolution on the left eye region in the anger condition




**(f)** Euclidean distances evolution on the right eye region in the anger condition



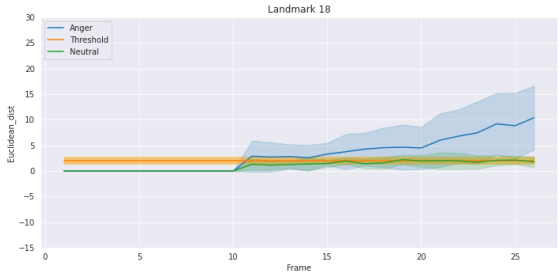

**(g)** Euclidean distances evolution on the mouth region in the anger condition

**Figure 3.9:** Plots representing the evolution of the Euclidean distances by facial region in the anger condition., without the data belonging to participant M12.

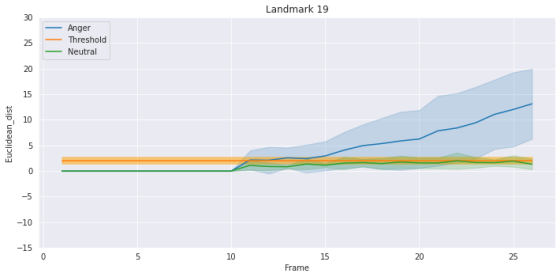
Having a global vision of the areas that better characterize the anger facial expression, the next step was to analyze the evolution of each landmark of the regions that were pointed out as relevant in this emotion.

Figure 3.10 presents the plots with the evolution of the Euclidean distances of the landmarks belonging to the eyebrows region. Similar to the previous figures, each chart represents the average value of the Euclidean distances in each frame for each landmark belonging to the eyebrows area, and the blue shaded band is the associated standard deviation. Furthermore, it is also represented the evolution of the Euclidean distances for each landmark in the neutral expression as a baseline to simplify the comparison between the neutral and, in this case, the anger expression and emphasize the displacement of these landmarks.

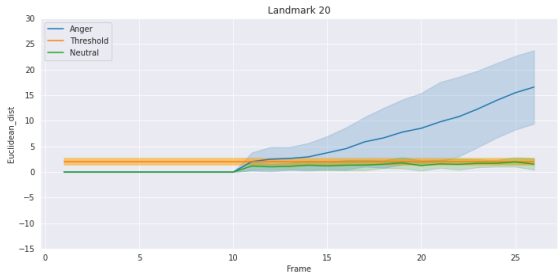
Observing these charts makes it possible to affirm that all the landmarks in the eyebrows area suffer substantial changes. Thus, all of them are essential to characterize the facial expression of anger.



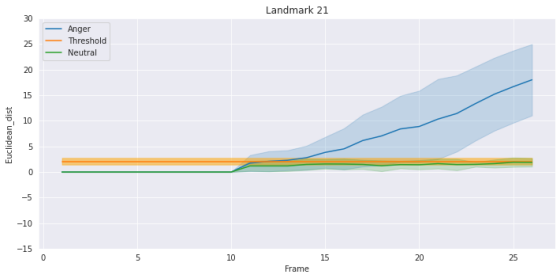

**(a)** Euclidean distances evolution of landmark 18



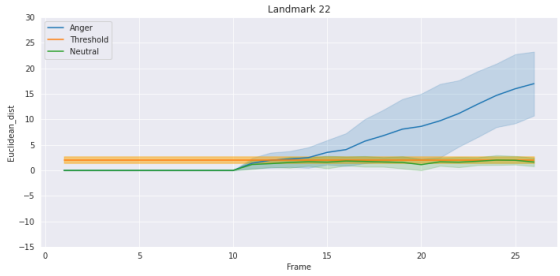

**(b)** Euclidean distances evolution of landmark 19



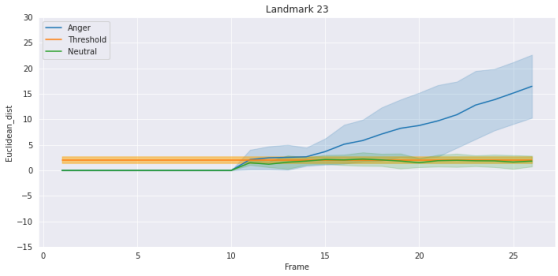

**(c)** Euclidean distances evolution of landmark 20



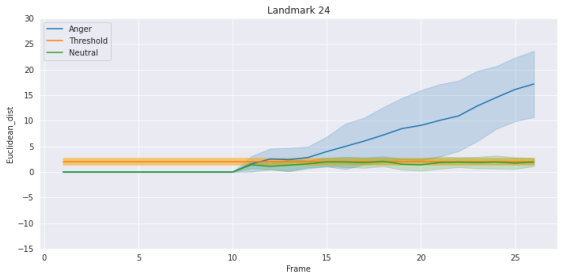

**(d)** Euclidean distances evolution of landmark 21



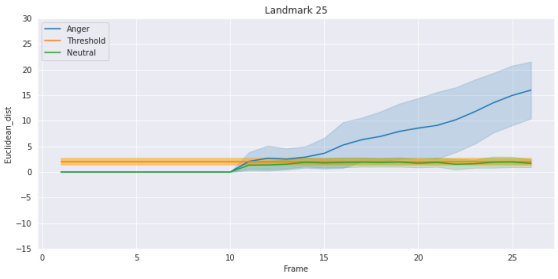

**(e)** Euclidean distances evolution of landmark 22



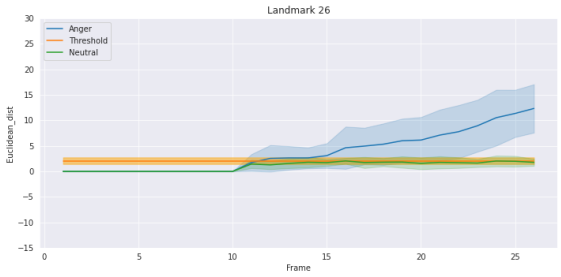

**(f)** Euclidean distances evolution of landmark 23



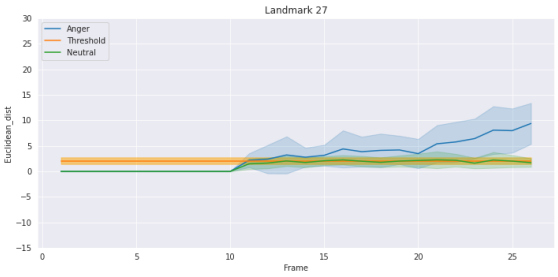

**(g)** Euclidean distances evolution of landmark 24




**(h)** Euclidean distances evolution of landmark 25




**(i)** Euclidean distances evolution of landmark 26



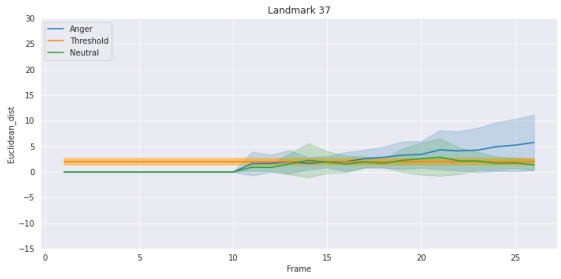

**(j)** Euclidean distances evolution of landmark 27

**Figure 3.10:** Plots representing the evolution of the Euclidean distances of the landmarks of the eyebrows region in anger condition.

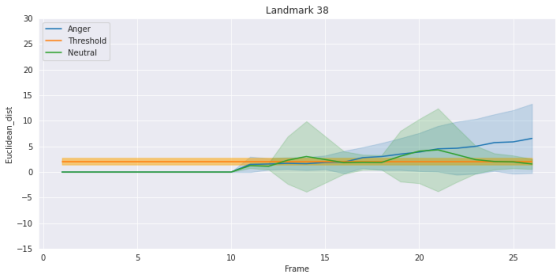
Figure 3.11 shows the evolution of the Euclidean distances for each landmark of the eye regions. It is possible to notice that the displacement in the eye regions is much smaller than in the eyebrows regions because the subjects displayed softer movements with the eyes which consequently generates smaller values of Euclidean distances. Also, all the landmarks present a similar evolution of the Euclidean distances, making it hard to select specific landmarks. However, combined with the previous assessment of the videos, it was possible to identify that landmarks 38, 39, 42, and 43, which map the upper eyelids, suffer small changes during the progression of the emotion and show a slightly higher evolution of the Euclidean distances in the charts.

Also, while examining the charts of landmarks 38, 39, 42, and 43, it was possible to notice that the green line and, more specifically, the green band, which represent the average value of the Euclidean distances in the neutral expression and the associated standard deviation, respectively, show some inconsistencies, revealing a decrease followed by an increase on the standard deviation values. This behavior happens due to some subjects blinking their eyes during the video, provoking some spikes in the data.

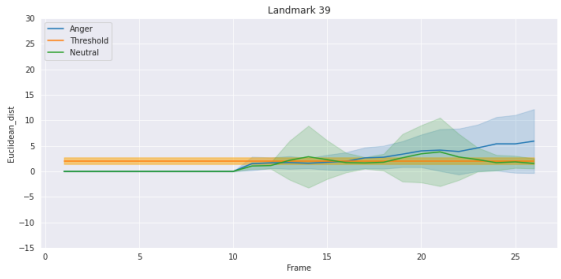
Figure 3.12 shows an example of a participant that blinked their eyes between frames 18 and 24. As it can be seen, there are steep alterations in the landmarks that map the upper eyelids, and, when compared with the behavior of the subject during the visualization of the video, it was possible to determine that the cause of these oscillations was the eye blinking. Although there were other situations in which the subjects blinked their eyes, these fluctuations caused by the eye blinking do not challenge the correct examination of the figures and do not invalidate the conclusions retrieved.



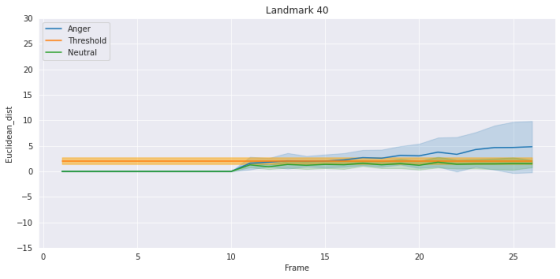

**(a)** Euclidean distances evolution of landmark 37



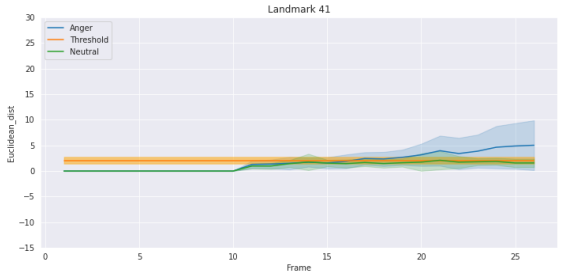

**(b)** Euclidean distances evolution of landmark 38



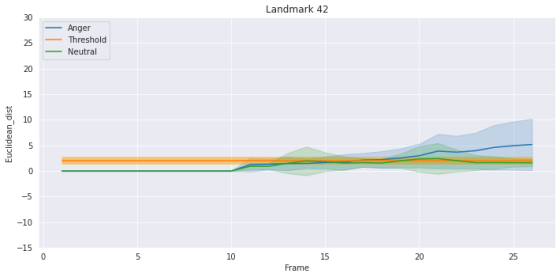

**(c)** Euclidean distances evolution of landmark 39



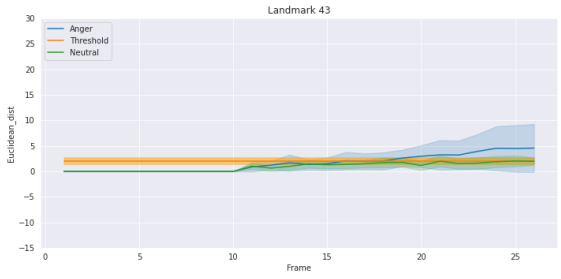

**(d)** Euclidean distances evolution of landmark 40



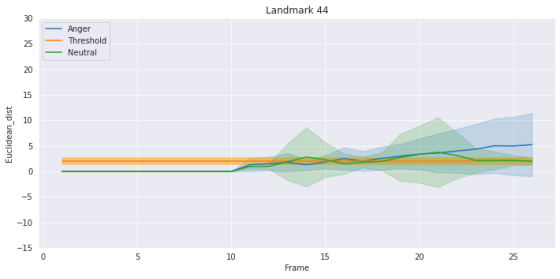

**(e)** Euclidean distances evolution of landmark 41



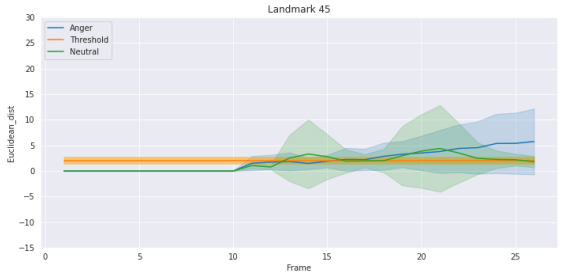

**(f)** Euclidean distances evolution of landmark 42



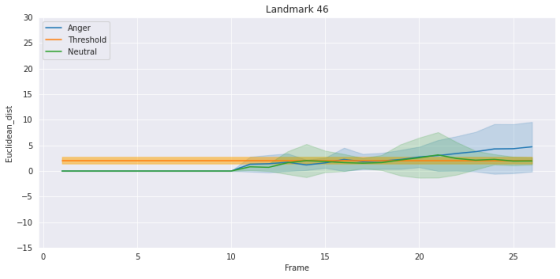

**(g)** Euclidean distances evolution of landmark 43



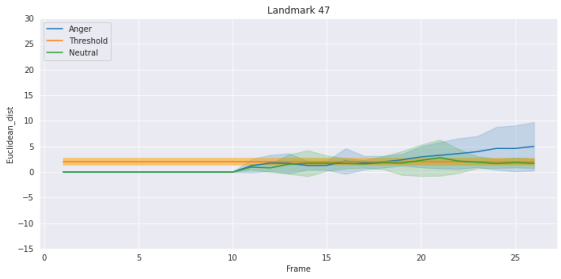

**(h)** Euclidean distances evolution of landmark 44



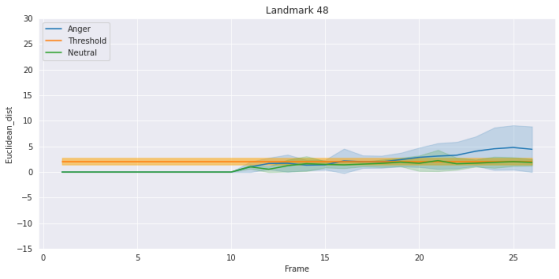

**(i)** Euclidean distances evolution of landmark 45




**(j)** Euclidean distances evolution of landmark 46

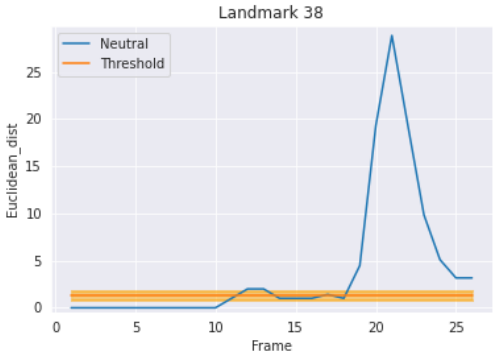



**(k)** Euclidean distances evolution of landmark 47

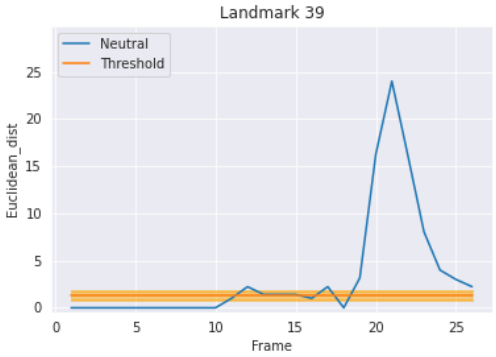



**(l)** Euclidean distances evolution of landmark 48

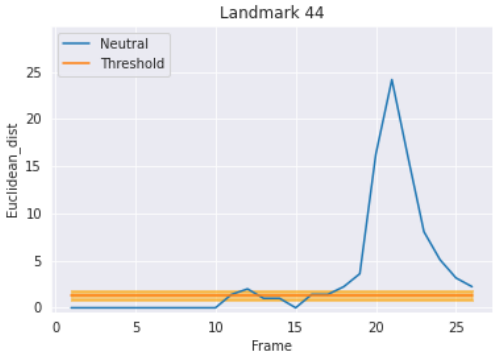
**Figure 3.11:** Plots representing the evolution of the Euclidean distances of the landmarks of the eyes region in anger condition.



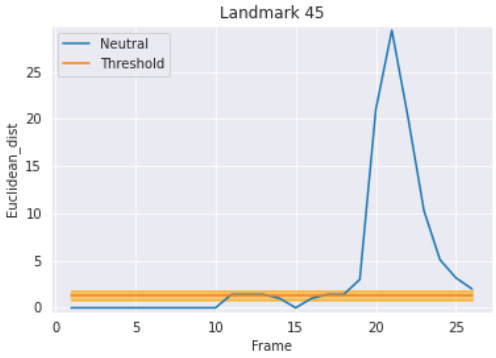

**(a)** Euclidean distances evolution of landmark 38




**(b)** Euclidean distances evolution of landmark 39



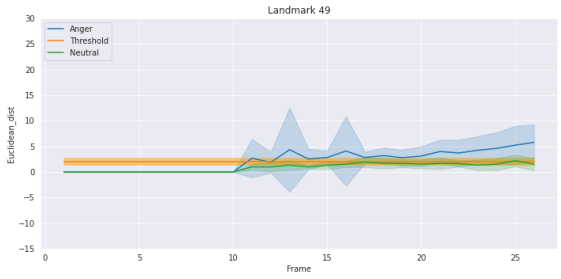

**(c)** Euclidean distances evolution of landmark 44



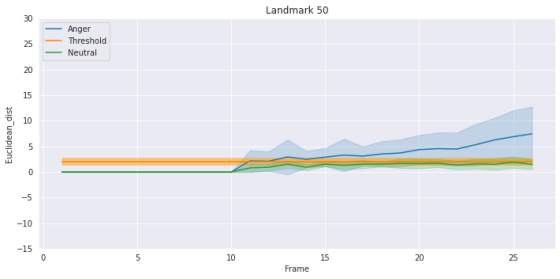

**(d)** Euclidean distances evolution of landmark 45

**Figure 3.12:** Plots representing the evolution of the Euclidean distances of the landmarks mapping the upper eyelids in the left and right eye in the neutral facial expression when subject F03 blinks their eyes.

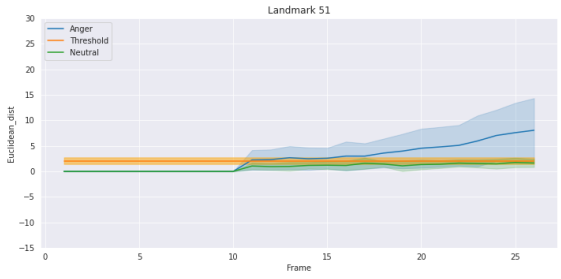
Now analyzing the charts representing the evolution of the Euclidean distances of the landmarks that map the mouth, depicted in Figure 3.13, it is possible to notice that, just as in the eye regions, all the landmarks have similar behavior. In this case, during the assessment of the video, it was impossible to identify any changes in specific landmarks that could be distinctive of this particular emotion. Therefore, and to not lose important information, all landmarks mapping the mouth are considered essential to characterize the facial expression of anger.



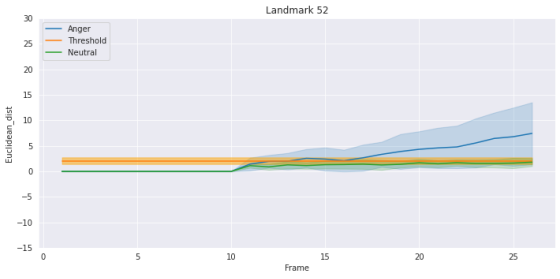

**(a)** Euclidean distances evolution of landmark 49



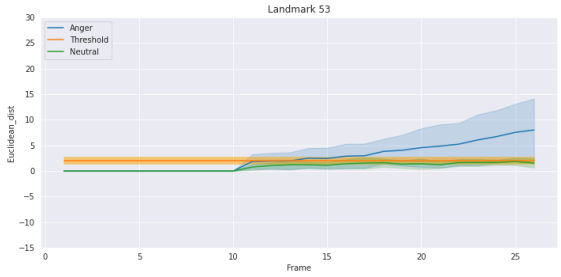

**(b)** Euclidean distances evolution of landmark 50



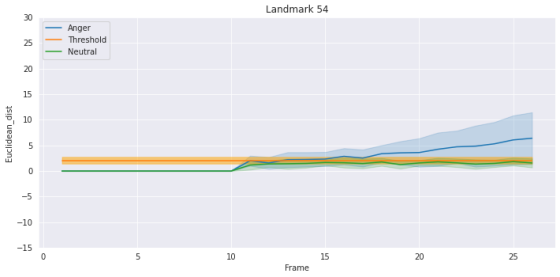

**(c)** Euclidean distances evolution of landmark 51



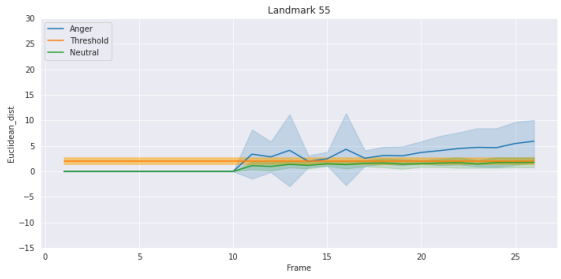

**(d)** Euclidean distances evolution of landmark 52



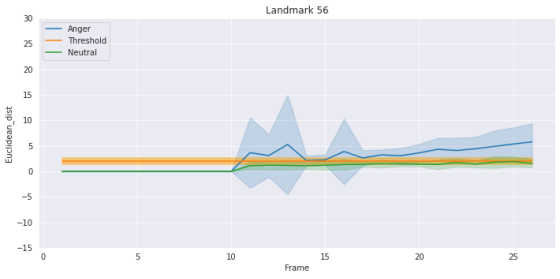

**(e)** Euclidean distances evolution of landmark 53



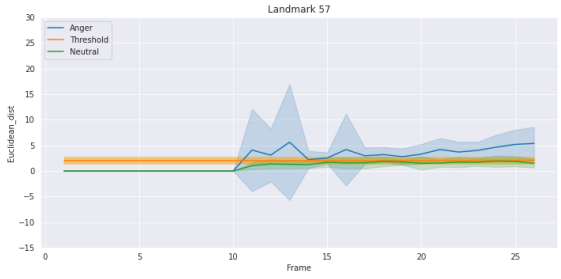

**(f)** Euclidean distances evolution of landmark 54



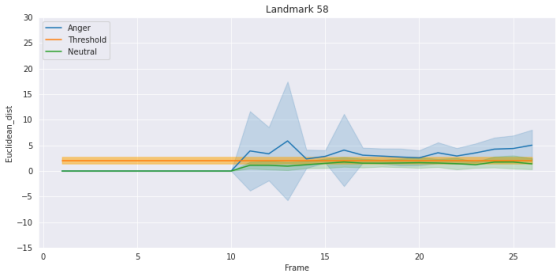

**(g)** Euclidean distances evolution of landmark 55



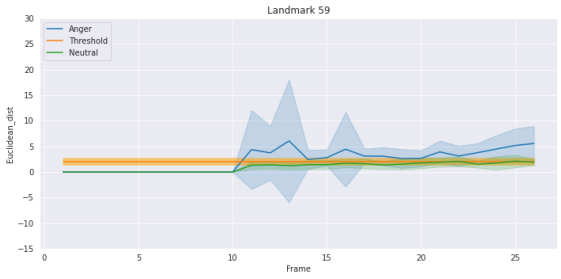

**(h)** Euclidean distances evolution of landmark 56



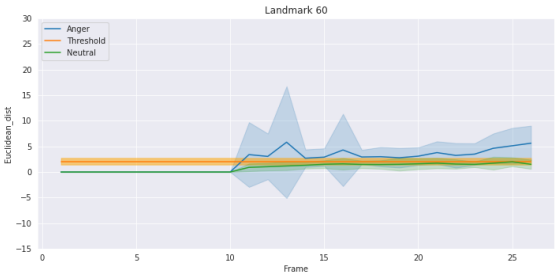

**(i)** Euclidean distances evolution of landmark 57



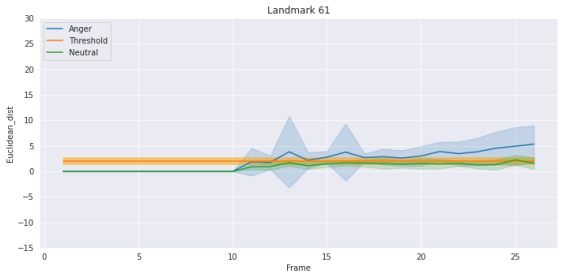

**(j)** Euclidean distances evolution of landmark 58



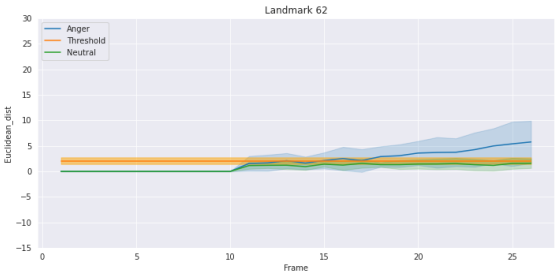

**(k)** Euclidean distances evolution of landmark 59



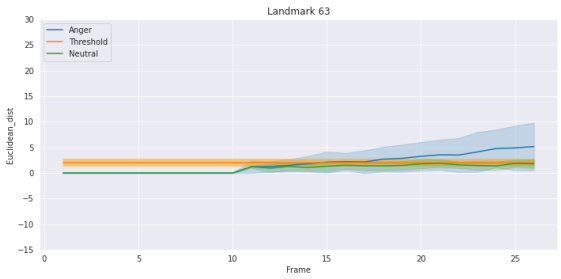

**(l)** Euclidean distances evolution of landmark 60



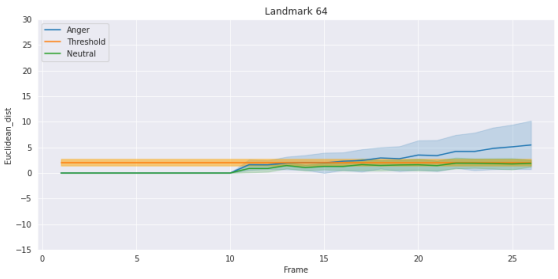

**(m)** Euclidean distances evolution of landmark 61



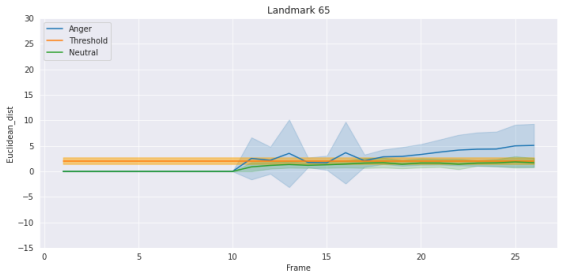

**(n)** Euclidean distances evolution of landmark 62



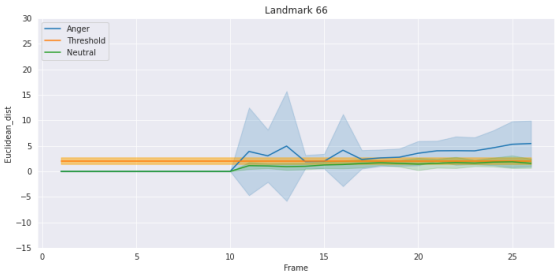

**(o)** Euclidean distances evolution of landmark 63



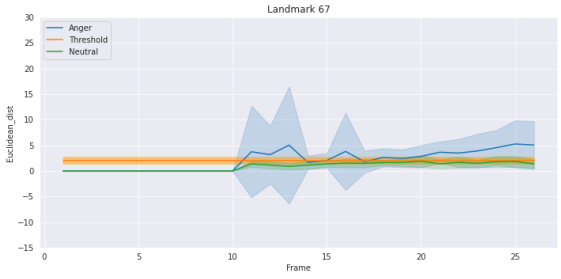

**(p)** Euclidean distances evolution of landmark 64



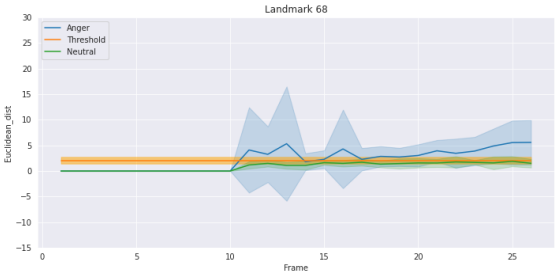

**(q)** Euclidean distances evolution of landmark 65




**(r)** Euclidean distances evolution of landmark 66




**(s)** Euclidean distances evolution of landmark 67




**(t)** Euclidean distances evolution of landmark 68

**Figure 3.13:** Plots representing the evolution of the Euclidean distances of the landmarks of the mouth region in anger condition.

At this point, we have selected the most informative landmarks and the most informative regions of the face for the angry facial expression.

The analysis methodology for the other emotions was done within the same criteria. Since this process requires extensive analysis and explanation of the visual data, it was decided not to display the charts related to the remaining emotions and sum up the information on Table 3.2 [2].

| **Emotions** | **Facial zones** | **Landmarks** | **Nr. of selected landmarks** |
|---|---|---|---|
| Anger | Left eyebrow | 18 - 22 | 34 |
| | Right eyebrow | 23 - 27 | |
| | Left eye | 38, 39 | |
| | Right eye | 44, 45 | |
| | Mouth | 49 - 68 | |
| Contempt | Jawline | 4 - 14 | 39 |
| | Left eyebrow | 19 - 22 | |
| | Right eyebrow | 23 - 26 | |
| | Mouth | 49 - 68 | |
| Disgust | Left eyebrow | 21, 22 | 23 |
| | Right eyebrow | 23, 24 | |
| | Nose | 32 - 36 | |
| | Mouth | 49 -56, 60-65 | |
| Fear | Jawline | 5 - 13 | 34 |
| | Left eyebrow | 18 - 22 | |
| | Right eyebrow | 23 - 27 | |
| | Left eye | 38, 39 | |
| | Right eye | 44, 45 | |
| | Mouth | 49, 50, 54 - 61, 65 - 68 | |
| Joy | Jawline | 2 - 16 | 35 |
| | Mouth | 49 - 68 | |
| Sadness | Left eyebrow | 18 - 22 | 34 |
| | Right eyebrow | 23 - 27 | |
| | Left eye | 38, 39 | |
| | Right eye | 44, 45 | |
| | Mouth | 49 - 68 | |
| Surprise | Jawline | 5 - 13 | 35 |
| | Left eyebrow | 18 - 22 | |
| | Right eyebrow | 23 - 27 | |
| | Nose | 28 - 30 | |
| | Left eye | 38, 39 | |
| | Right eye | 44, 45 | |
| | Mouth | 49, 55 - 61, 65 - 68 | |
| Neutral | None | 1 - 68 | 68 |

**Table 3.2:** Most significant regions and selected landmarks in each emotion.

[2]For better perception, you may find an illustrative image of the selected facial landmarks in the Appendix.

Since state-of-the-art solutions only studied the basic emotions, it was decided to leave aside the facial expression of embarrassment since it was the only complex emotion left.

A set of hypotheses were achieved regarding each one of the remaining emotions. These hypotheses are the selected landmarks that better characterize each emotion, i.e., the landmarks that underwent substantial changes during the evolution of each facial expression.

From the table, besides the most informative regions and landmarks in each emotion, it is also possible to deduce that the most significant region overall is the mouth because it is the only area that always shows activity, no matter the emotion expressed. Also, from the information in the last column, it was possible to reduce the total number of landmarks used to characterize each emotion.

Naturally, the neutral facial expression continues to be described by all 68 landmarks since it is the baseline to all other emotions and does not exhibit any facial alterations.

CHAPTER 4

# Facial expressions classification

From the previous study, we obtained a set of hypotheses describing the most informative landmarks in each emotion. The next stage was to validate these hypotheses. This chapter will present the features extracted and selected as relevant, the chosen model for this work, and the results obtained.

## 4.1 Feature extraction and selection

From the results of the previous chapter, we have a set of facial landmarks coordinates that characterize each emotion. However, these raw coordinates are insufficient and inadequate for a machine learning model to learn and detect patterns because the coordinates alone do not hold any meaning. Also, the same coordinates might occur in different facial expressions. Therefore, it is necessary to extract useful features from these coordinates.

In the study conducted earlier, the Euclidean distances between landmarks with the same *id* throughout the frames of each video were extracted, as shown in Figure 3.5. These distances were given as input features to the model.

Besides extracting features along the frames, extracting features within the same frame was also considered relevant. Thus, based on some parts of the methods described in [28] and [26], the following features were extracted:

- **Method 1:** Euclidean distances to one fixed reference point (nose);
- **Method 2:** Euclidean distances to several fixed reference points (left eye, right eye, nose, and mouth).

In the first method, there were calculated the Euclidean distances between a fixed reference point (tip of the nose mapped by landmark 31) and each one of the coordinates of the 68 landmarks. Figure 4.1 shows an example of this method.

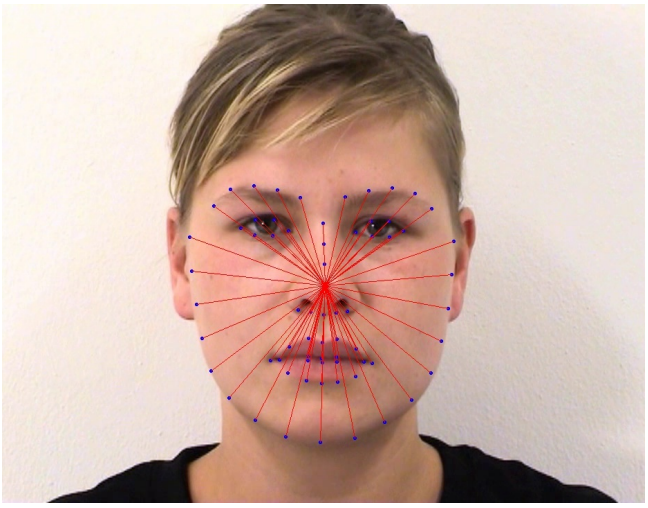


**Figure 4.1:** Euclidean distances to one fixed reference point (tip of the nose).

In method 2, the facial landmarks are divided into four regions:

1. Left eye area (landmarks 18 to 22 and 37 to 42);
2. Right eye area (landmarks 23 to 27 and 43 to 48);
3. Mouth area (landmarks 49 to 68);
4. remaining landmarks.

Each area has a reference point. The reference points coordinates of the left eye, right eye, and mouth areas were calculated as follows:

$$C_n = \frac{L_1 + L_2}{2} \tag{4.1}$$

where $C_n$ represents the coordinates of the reference point in the corresponding *nth* regions, and $L_1$ and $L_2$ represent the left and right corners of the eyes, respectively [26].

The fourth reference point is the tip of the nose (landmark 31), as in method 1.

In this case, the Euclidean distances are calculated between the landmarks and the reference point belonging to the same area, as depicted in Figure 4.2 (distances to the nose reference point are not represented to facilitate the visualization).

**Figure 4.2:** Euclidean distances to several fixed reference points.

After extracting these features, it was noticed that there are duplicated values in the landmarks where the reference point is the tip of the nose. This causes redundancy in the input features. Therefore, it was decided to drop the first method and proceed only with the Euclidean distances to several reference points and the Euclidean distances between landmarks with the same id throughout the frames of each video.

## 4.2 Machine learning model implementation

The data set used to train and test the machine learning algorithms was the same one used in the previous study, the ADFES-BIV data set. In this case, it was decided to use the videos displaying low, intermediate, and high intensities of emotion to increase the number of training samples. All the videos are labeled with the emotion expressed after the neutral expression phase.

Since the data set is labeled, the best approach is to use supervised learning to classify the emotions. Furthermore, since the goal is to categorize each emotion into anger, contempt, disgust, fear, joy, sadness, surprise, and neutral based only on specific landmarks, the best approach is to use the classification method.

In the previous study, we determined the activation frame for each subject in each emotion (Table 3.1). Therefore, it was considered relevant to use that information and label each frame of each video based on the activation frames defined earlier. For instance, the activation frame of participant F01 in the emotion of anger is frame 22, which means that from frame 1 to frame 21, the participant shows a neutral facial expression, and from frame 22 to frame 26, the participant expresses anger. Thus, each frame was labeled according to that principle, as illustrated in Figure 4.3.

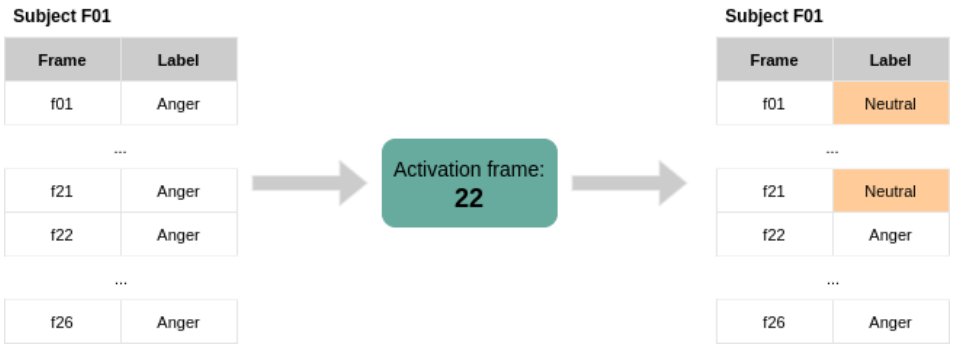



**Figure 4.3:** Illustrative process of frame labeling for participant F01 while expressing anger.

Furthermore, specific landmarks were selected that characterize each emotion, which implies that the model will be given different input features depending on the emotion expressed in each frame. For example, in the facial expression of anger, there were designated landmarks 18 to 27, 38, 39, 44, 45, and 49 to 68. From a total of 68 landmarks, only 34 were considered essential to describe the emotion of anger. Only the extracted features relative to these landmarks will be used when training the model. The same happens for the other emotions.

Thus, some columns will have missing values, which is a problem when trying to perform multiclass classification. Therefore, it was decided to break this multiclass classification problem into several binary classifiers.

An SVM algorithm was trained for every class, where the class distribution is the target class versus the rest of the classes. So, each frame was re-labeled accordingly, where the target class has label 1, and the rest of the classes have label 0, as exemplified in Figure 4.4.

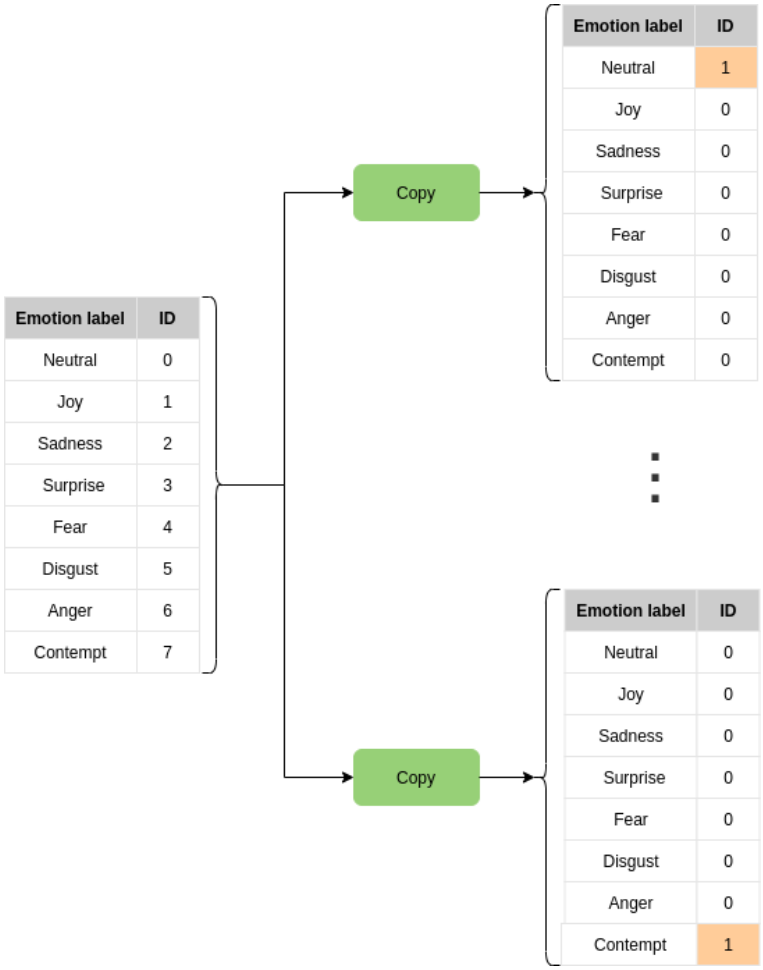



**Figure 4.4:** Transformation from multiclass labels to binary labels.

Since there are eight different classes, eight copies of the data set were created with the new binary labels. Each data set copy was split into training and test groups, each consisting of the same samples for training and testing. The data was divided into 80% for training and 20% for testing.

After labeling each frame according to the activation frames and then again later into binary labels, the data became highly imbalanced. Figure 4.5 shows the number of samples for each class. Typically, a balanced data set yields better machine learning implementations since it contains the same amount of representations of the different classes. In this case, the learning algorithm assigns the same weight to each class, avoiding potential biases.

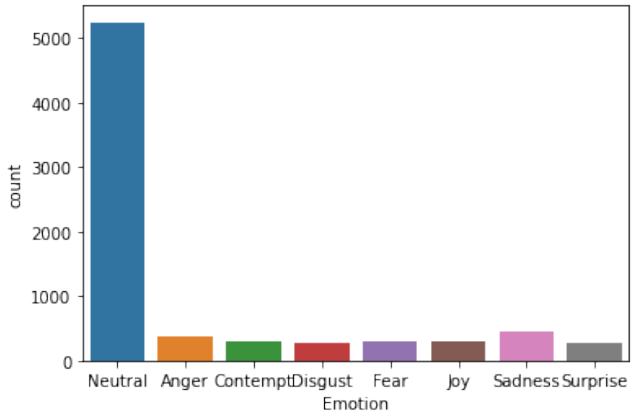



**Figure 4.5:** Class distribution after labeling each frame according to the activation frame.

To solve this problem of data imbalance, we can choose to act on the data using resampling techniques, such as under-sampling and/or oversampling, or act on the machine learning algorithm by opting for cost-sensitive learning algorithms. In this work, it was considered that oversampling would create undesirable representations of the data, and under-sampling would significantly reduce the number of samples. Therefore, cost-sensitive learning was the chosen method.

Cost-sensitive learning involves explicitly defining and using costs when training the machine learning algorithm. The idea is to increase the weight of the minority class and decrease the weight of the majority class so that the algorithm gives more importance to the class with a higher weight.

In order to implement an SVM algorithm that supports class weighting, it was used the *scikit-learn* Python machine learning library [1].

The SVM algorithm tries to find a line or a hyperplane that best separates the data, called the decision boundary. This hyperplane is defined by a margin that maximizes the distance between the decision boundary and the closest examples of each class. Most of the time, the data is not separable, making it necessary to soften the margin to allow some wrong samples to appear on the wrong side of the decision boundary. The softening of the margin is controlled by the regularization parameter $C$.

The Support Vector Classifier (SVC) class provides an argument (*class_weight*) that can be specified as a model hyperparameter. The *class_weight* is a dictionary defining each class label and the weighting to apply to the $C$ value in the calculation of the soft margin, defined as:

$$C_i = weight_i \times C \tag{4.2}$$

This allows us to weigh the C value in proportion to the importance of each class. For instance, a larger weight can be defined for the minority class, allowing the margin to be softer, and a smaller weight can be used for the majority class to harden the margin and prevent misclassification.

[1] https://scikit-learn.org/stable/

In this case, the *class_weight* parameter was set to *'balanced'* mode, which uses the training labels to adjust weights inversely proportional to class frequencies automatically.

Another important aspect to take into consideration is that the SVC algorithm requires the data to be normalized to guarantee its better functioning since this algorithm bases its learning process on distances. Therefore, the data normalization was done using the *Normalizer* method, also from the *scikit-learn* library.

The next step was the hyperparameter tunning to select the optimal parameters to train the model. For that, the *GridSearchCV* method from the *scikit-learn* library was used.

The possible parameters to select from were the following:

- *C*: [0.1, 1, 10, 100, 1000]
- *gamma*: [1, 0.1, 0.01, 0.001, 0.0001]
- *kernel*: ['rbf']
- *class_weight*: ['balanced']

### 4.3 Results

The performance evaluation metrics used in this project were accuracy, F1 score, precision, and recall. Accuracy is the most common metric to describe a model's performance and is defined as the ratio of correct predictions to the total number of predictions. The F1 score is the balance between precision and recall and is determined by the harmonic mean between these two metrics, where precision is the ratio of true positives and total positives predicted, and recall is the ratio of true positives and all the positives in the ground truth. The evaluation metrics obtained for each classifier are presented in Table 4.1.

| | Train set | | | | Test set | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | F1 Score | Precision | Recall | Accuracy | F1 Score | Precision | Recall |
| Clf Anger | 96.0 | 71.4 | 56.5 | 97.1 | 95.3 | 66.3 | 52.7 | 89.6 |
| Clf Contempt | 89.6 | 42.4 | 27.5 | 92.7 | 89.0 | 37.7 | 23.8 | 90.9 |
| Clf Disgust | 94.0 | 52.6 | 36.2 | 96.2 | 94.5 | 56.8 | 40.9 | 93.1 |
| Clf Fear | 93.7 | 51.6 | 36.8 | 86.3 | 73.0 | 17.9 | 10.0 | 81.5 |
| Clf Joy | 93.6 | 53.1 | 36.8 | 95.2 | 91.6 | 47.9 | 33.1 | 86.6 |
| Clf Sadness | 92.8 | 61.6 | 45.4 | 95.9 | 78.0 | 33.7 | 20.4 | 96.6 |
| Clf Surprise | 91.1 | 43.1 | 28.3 | 91.0 | 91.5 | 32.8 | 20.5 | 81.6 |
| Clf Neutral | 85.3 | 89.1 | 92.7 | 85.7 | 85.0 | 89.0 | 92.7 | 85.6 |

**Table 4.1:** Evaluation metrics results of each SVM classifier.

Analyzing the results, one can notice that the accuracy obtained is very satisfactory. However, the F1 score, and more precisely, the precision score, could be better. These results suggest that there is a considerable amount of false positives, confirmed by the confusion matrices [2] as well. In most cases, the algorithm predicts more false positives than true positives, meaning that the model classifies multiple samples from the negative class as belonging to the positive class.

[2] You may find the supporting confusion matrices in the appendix.

The imbalance in the data can be a motive for lower precision scores. As already mentioned, there are more samples of the neutral expression, which creates a skew in the data. Therefore, it was decided to remove the neutral expression from all the training and testing groups to verify if these data were affecting the performance of the algorithms.

| | Train set | | | | Test set | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | F1 Score | Precision | Recall | Accuracy | F1 Score | Precision | Recall |
| Clf Anger | 96.1 | 89.2 | 83.5 | 95.7 | 95.3 | 88.0 | 86.5 | 89.5 |
| Clf Contempt | 87.1 | 65.5 | 49.9 | 95.2 | 84.9 | 64.6 | 50.4 | 89.9 |
| Clf Disgust | 94.8 | 82.5 | 71.3 | 97.8 | 96.0 | 80.4 | 72.5 | 90.2 |
| Clf Fear | 90.7 | 71.0 | 58.7 | 90.0 | 88.4 | 63.4 | 54.2 | 76.3 |
| Clf Joy | 97.0 | 89.4 | 83.5 | 96.2 | 96.4 | 87.1 | 80.6 | 94.7 |
| Clf Sadness | 90.7 | 80.6 | 69.3 | 96.4 | 89.3 | 77.8 | 66.7 | 93.3 |
| Clf Surprise | 91.5 | 72.7 | 58.7 | 95.3 | 89.7 | 65.7 | 51.2 | 91.7 |

**Table 4.2:** Evaluation metrics results of each SVM classifier after removing the data corresponding to the neutral facial expression.

Table 4.2 shows the results obtained for each classifier without the data belonging to the neutral facial expression. As can be observed, the accuracy score did not suffer substantial alterations, maintaining high values, which means that the models made correct predictions in most cases. However, it is noticeable the increase in the precision and F1 scores. These results indicate that the neutral expression indeed affected the outcome of the classifiers.

Naturally, when we reduce the number of training samples, the performance evaluation metrics tend to improve since the algorithm is more restricted to that smaller set of data. By removing the neutral expression data, we also significantly reduced the data imbalance, which can explain the improvement in the precision score and, consequently, the F1 score since it is based on the precision and recall values.

Another reason why the neutral expression may be influencing the results can be related to the characteristics of this emotion per si. It is possible that some features extracted from the neutral facial expression may be transversal to the other expressions, and thus, the algorithm misclassifies the neutral samples. This theory was not explored in this work.

It is also important to note that the recall score presented high values that remained stable both with and without the neutral expression data, which means that in both cases, the algorithm correctly predicted the majority of the true positives, i.e., the positive class is correctly detected in most cases. The analysis of the confusion matrices also reinforces these claims.

Furthermore, it was considered relevant to portray the predictions from each classifier in the bi-dimensional model of emotions. The goal was to analyze the distribution of each emotion in the quadrants and understand if there was any connection between the correct and incorrect predictions obtained in each classifier. Since the dataset used does not have the valence and arousal values of each emotion, it was necessary to impose the values of each point in the plots. For that, Figure 4.6 was used to reference the placement of the emotions in the charts.

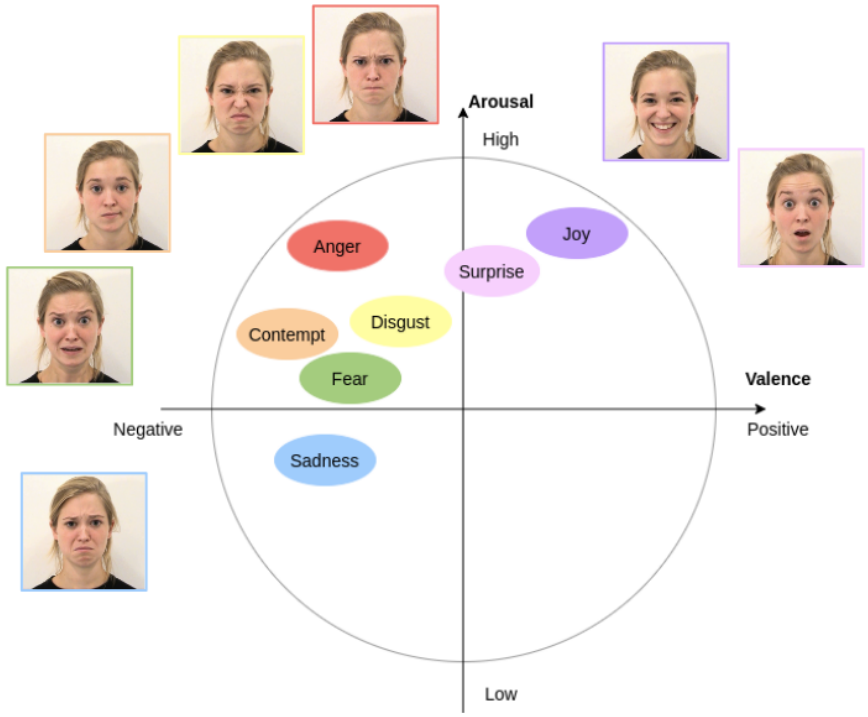
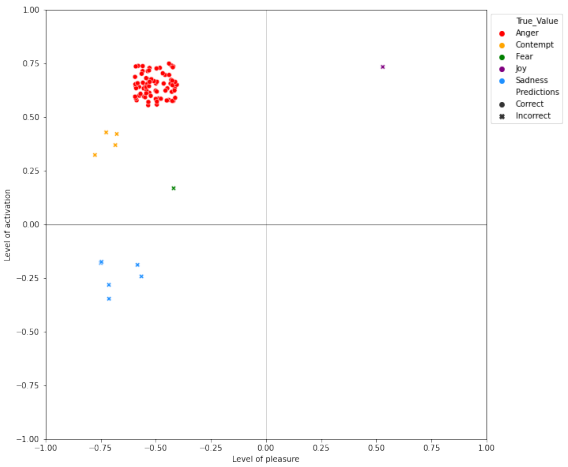


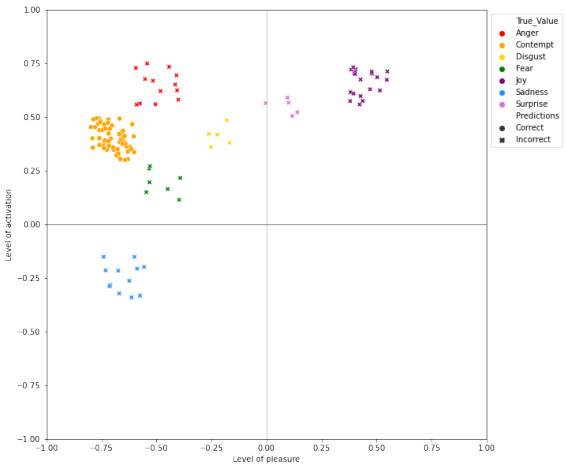

**Figure 4.6:** The basic emotions in the bi-dimensional model, adapted from [38]

A plot was created for each classifier's predictions in the testing data without the neutral facial expression (Figure 4.7). Although we transformed the problem into a binary classification, it is possible to confirm if the classifier made the correct prediction and infer all the classes predicted since we know the true value of each sample. Therefore, each plot shows the correct and incorrect predictions in each classifier and discriminates all the classes detected as well.

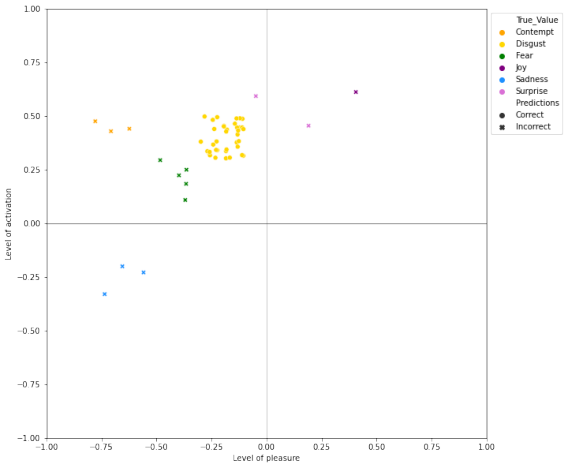
Each plot contains the correct (•) and incorrect (x) predictions of the classifiers. However, since the symbols in the images are barely perceptible, Table 4.3 includes the percentage of the labels predicted in each classifier.



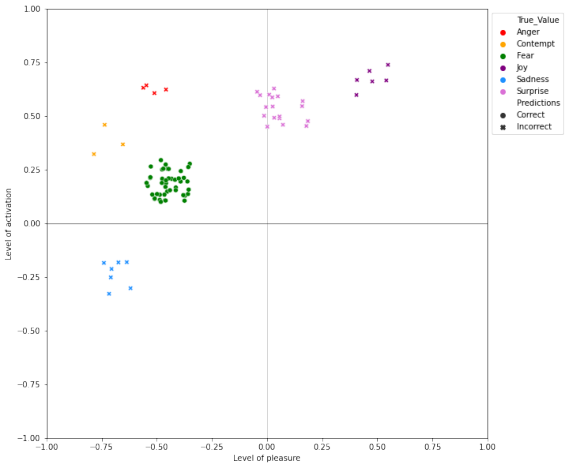

**(a)** Test predictions in the classifier of anger.



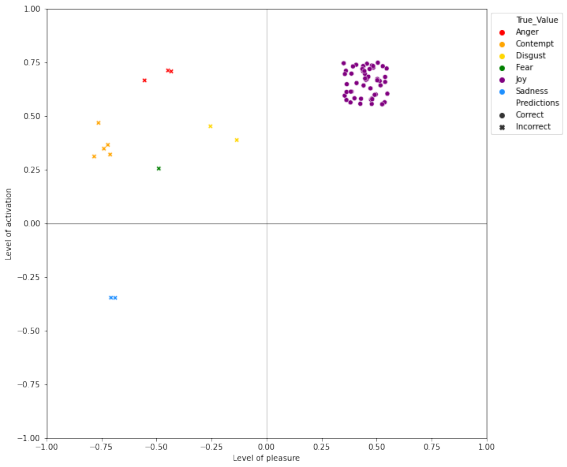

**(b)** Test predictions in the classifier of contempt.



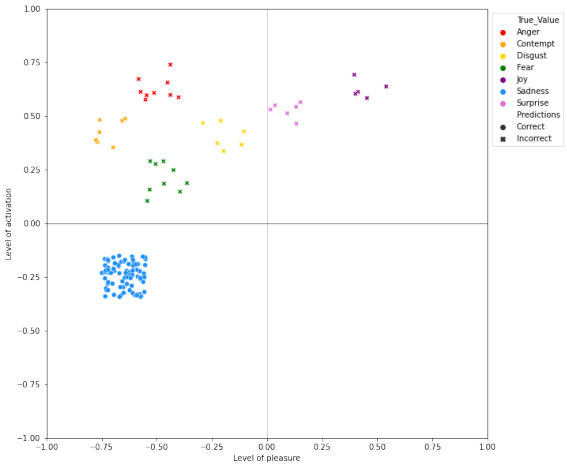

**(c)** Test predictions in the classifier of disgust.



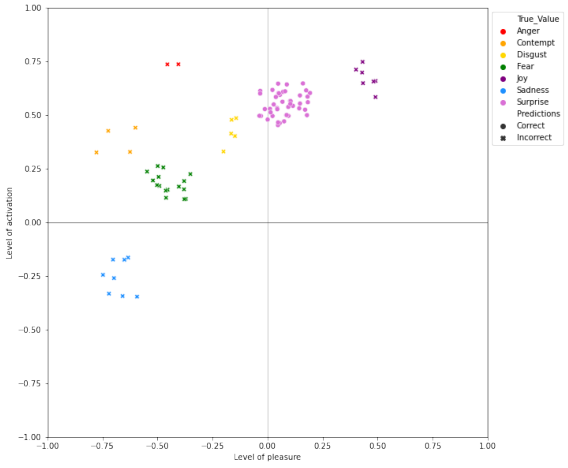

**(d)** Test predictions in the classifier of fear.




**(e)** Test predictions in the classifier of joy.




**(f)** Test predictions in the classifier of sadness.




**(g)** Test predictions in the classifier of surprise.

**Figure 4.7:** Test predictions of each classifier portrayed in the dimensional model of emotions.

| | Predicted labels (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Anger | Contempt | Disgust | Fear | Joy | Sadness | Surprise |
| Clf Anger | **86.52** | 4.49 | 0 | 1.12 | 1.12 | 6.74 | 0 |
| Clf Contempt | 10.48 | **50** | 4.03 | 5.65 | 16.13 | 9.68 | 4.03 |
| Clf Disgust | 0 | 5.88 | **72.55** | 9.8 | 1.96 | 5.88 | 3.92 |
| Clf Fear | 4.82 | 3.61 | 0 | **54.22** | 7.23 | 8.43 | 21.69 |
| Clf Joy | 4.48 | 7.46 | 2.99 | 1.49 | **80.6** | 2.99 | 0 |
| Clf Sadness | 7.14 | 5.56 | 4.76 | 7.14 | 3.97 | **66.67** | 4.76 |
| Clf Surprise | 2.33 | 4.65 | 5.81 | 18.6 | 8.14 | 9.3 | **51.16** |

**Table 4.3:** Percentages of classifier identification to each of the described classes.

Observing the plots, one can notice that the majority of predictions are located in quadrant 2, indicating a high level of activation and negative valence. However, this may be due to the fact that most of the basic emotions belong to this quadrant of the dimensional model, except for the joy and surprise that are located in quadrant 1 and sadness in quadrant 3. This characteristic makes it more difficult to find correlations between the correct class and the other predicted classes.

Nevertheless, it is still possible to identify that in the case of anger, disgust, and sadness classifiers, most misclassifications are encountered on the negative side of valence, meaning that these classifiers are misinterpreting emotions associated mostly with unpleasantness.

Another interesting characteristic is that facial expressions of fear and surprise are frequently mistaken for each other, which is an aspect that also occurs in multi-class classification approaches. In fact, these two facial expressions reveal similar characteristics, and proof of that can be found in the selected landmarks for both of these emotions. As shown in Table 3.2, the landmarks that represent fear are almost the same as those that represent surprise, except that in the case of surprise, there are additional landmarks mapping part of the nose and the mouth has two fewer landmarks than in fear.

The optimal hyperparameters found for the SVM classifiers were:

- $C = 1000$;
- $gamma = 1$;

Besides the SVM classifier, the data was also trained and tested with a Logistic Regression (LR) classifier. However, the obtained results were worse than the ones found with the SVM. Thus, it was not considered relevant to show and analyze the outcomes of this algorithm in the scope of this work.

CHAPTER 5

# Discussion and conclusions

This dissertation aimed to study human facial expressions with the intent of determining which facial areas are most used to express a particular emotion. For that purpose, there were defined three main objectives: (1) determine the facial regions that better characterize each emotion; (2) evaluate which are the most significant landmarks in each condition; (3) find out which is the most expressive area of the face.

Finding an appropriate data set for this project was a challenging task. Firstly, the majority of the existing data sets are made of static images. Since the goal was to analyze the evolution of a facial expression to discern the areas that underwent alterations, it was required a video data set or an equivalent with sequential images of the same participant in all conditions. Also, the video must have only one subject per frame, preferably facing the camera and always at the same distance to avoid inconsistencies in the data.

A facial landmarking method that maps the salient regions of the human face with 68 landmarks was used to study and characterize the emotions based on the amount of movement demonstrated in each facial area described by the evolution of the Euclidean distances throughout the frames. In the end, it was achieved a set of selected landmarks (hypotheses) considered essential to describe each particular emotion.

The next step was to verify the viability of the hypotheses obtained. For that, the input features used to train the model were:

- The Euclidean distances between the landmarks with the same index throughout the different frames of the video;
- The Euclidean distances between each landmark and several origin points within the same frame.

Considering that each video evolves from a neutral expression to an apex of emotion, then, consequently, there is an evolution of the facial landmarks. Identifying the first frame containing emotional information is an important step to correctly train the model to recognize facial alterations associated with the emotion. A blind evaluation that uses all the frames without validating the frame where the activation occurred induces the classifier in error since the method will learn neutral expressions labeled as other emotional conditions. However,

assigning these frames to the neutral class caused a high imbalance in the data. For this reason, a cost-sensitive approach was adopted by assigning different weights to each class representation to prioritize the class with fewer samples during model training.

Multiclass problems are prone to misclassification and, in the particular case of emotions, this is emphasized since there are emotions that share facial expression characteristics. So, to simplify the approach, the multiclass problem was split into seven binary classifiers.

In conclusion, this work allowed a better understanding of facial expressions, proving that each basic emotion can be characterized by specific facial areas. The classification results showed that it is possible to discern an emotion with much less information if we focus on the appropriate regions of the face. Besides determining the facial areas and landmarks that better characterize each emotion, it was also established that the most expressive area of the human face is the mouth, which is the only facial part that always displays movement independently of emotions being expressed.

Using a minimal number of facial landmarks in facial expressions recognition makes it possible to increase computational efficiency by using fewer input features to train the model and, consequently, promotes the usage of simpler algorithms. Also, if there is a need to store information, this method allows for preserving people's privacy because it diminishes the probability of re-identifying someone based only on these landmarks.

## 5.1 Future work

As a future work proposal, it would be useful to create a normalization layer to deal with some challenges of real-world facial images. Unlike most data sets created under specific and controlled environments, real-world images present characteristics that can negatively influence systems' learning process. For instance, research a method to deal with rotated faces or partial representations of the face. Also, explore a way to adjust the scale of the faces uniformly.

It would also be interesting to conduct a study on the neutral facial expression to understand which variables are affecting the results and if any characteristics are transversal to the other emotions.

Furthermore, it would be of enormous interest to create a decision layer to connect all the binary classifiers and have a single final result. At the moment, each classifier produces results independently, which leads to, in some cases, having two classifiers returning true for the same file. However, in a real scenario, each file should have only one result. A suggestion to solve this problem would be to, instead of returning categorical predictions (0 or 1), investigate a way to compute the confidence percentage of the classifier about the predicted label and return that value instead.

# References


[1] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors (Switzerland)*, vol. 18, 2 Feb. 2018, ISSN: 14248220. DOI: 10.3390/s18020401.

[2] N. Samadiani, G. Huang, B. Cai, *et al.*, *A review on automatic facial expression recognition systems assisted by multimodal sensor data*, Apr. 2019. DOI: 10.3390/s19081863.

[3] W. H. Abdulsalam, R. S. Alhamdani, and M. N. Abdullah, "Facial emotion recognition from videos using deep convolutional neural networks," *International Journal of Machine Learning and Computing*, vol. 9, pp. 14–19, 1 Feb. 2019, ISSN: 20103700. DOI: 10.18178/ijmlc.2019.9.1.759.

[4] M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, and J. Kissler, "Mapping the emotional face. how individual face parts contribute to successful emotion recognition," 2017. DOI: 10.1371/journal.pone.0177239. [Online]. Available: https://doi.org/10.1371/journal.pone.0177239.

[5] R. Arya, J. Singh, and A. Kumar, *A survey of multidisciplinary domains contributing to affective computing*, May 2021. DOI: 10.1016/j.cosrev.2021.100399.

[6] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, Sep. 2017, ISSN: 15662535. DOI: 10.1016/j.inffus.2017.02.003.

[7] T. S. Ashwin and R. M. R. Guddeti, "Affective database for e-learning and classroom environments using indian students' faces, hand gestures and body postures," *Future Generation Computer Systems*, vol. 108, pp. 334–348, Jul. 2020, ISSN: 0167739X. DOI: 10.1016/j.future.2020.02.075.

[8] J. C. Torrado, J. Gómez, and G. Montoro, "Emotional self-regulation of individuals with autism spectrum disorders: Smartwatches for monitoring and interaction," *Sensors*, vol. 17, p. 1359, Jun. 2017. DOI: 10.3390/s17061359.

[9] C. Zucco, B. Calabrese, and M. Cannataro, "Sentiment analysis and affective computing for depression monitoring," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 1988–1995. DOI: 10.1109/BIBM.2017.8217966.

[10] L. Cao, H. Zhang, X. Wang, and L. Feng, "Learning users inner thoughts and emotion changes for social media based suicide risk detection," *IEEE Transactions on Affective Computing*, pp. 1–1, 2021, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2021.3116026.

[11] S. Greene, H. Thapliyal, and A. Caban-Holt, "A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health," *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 44–56, 2016. DOI: 10.1109/MCE.2016.2590178.

[12] D. Caruelle, P. Shams, · . A. Gustafsson, and · . L. Lervik-Olsen, "Affective computing in marketing: Practical implications and research opportunities afforded by emotionally intelligent machines," 123. DOI: 10.1007/s11002-021-09609-0. [Online]. Available: https://doi.org/10.1007/s11002-021-09609-0.

[13] M. F. Alsharekh, "Facial emotion recognition in verbal communication based on deep learning," *Sensors*, vol. 22, 16 Aug. 2022, ISSN: 14248220. DOI: 10.3390/s22166105.

[14] I. Horkovska. "6 types of basic emotions and their effect on human behavior." (2022), [Online]. Available: https://us.calmerry.com/blog/psychology/6-types-of-basic-emotion/#Robert_Plutchiks_wheel_of_emotions (visited on 10/19/2022).

[15] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Corrigendum: Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 3, pp. 165–166, 2019, PMID: 31729935. DOI: 10.1177/1529100619889954. eprint: https://doi.org/10.1177/1529100619889954. [Online]. Available: https://doi.org/10.1177/1529100619889954.

[16] P. Ekman and W. Friesen, *Facial Action Coding System: Investigator's Guide*, ser. Facial Action Coding System: Investigator's Guide vol. 2. Consulting Psychologists Press, 1978. [Online]. Available: https://books.google.pt/books?id=7pqFtQAACAAJ.

[17] M. Codispoti, G. Mirabella, E. A. Clark, *et al.*, "The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review," *Frontiers in Psychology | www.frontiersin.org*, vol. 1, p. 920, 2020. DOI: 10.3389/fpsyg.2020.00920. [Online]. Available: www.frontiersin.org.

[18] P. E. Group. "Universal emotions." (), [Online]. Available: https://www.paulekman.com/universal-emotions/ (visited on 10/19/2022).

[19] P. Ekman, "Universal facial expressions of emotion," *California Mental Health*, vol. 8 (4), pp. 151–158, 1970.

[20] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, no. 6 (3/4), pp. 169–200, 1992.

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

[22] D. Preotiuc. "Sentiment, intensity and user attributes." (2015), [Online]. Available: https://wwbp.org/blog/sentiment-intensity-and-user-attributes/ (visited on 10/19/2022).

[23] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273–294, 3 Sep. 1977, ISSN: 0092-6566. DOI: 10.1016/0092-6566(77)90037-X.

[24] A. Mehrabian, "Pleasure-arousal.dominance: A general framework for describing and measuring individual differences in temperament," 1996, pp. 261–292.

[25] O. Bălan, G. Moise, L. Petrescu, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, "Emotion classification based on biophysical signals and machine learning techniques," *Symmetry*, vol. 12, 1 2020, ISSN: 20738994. DOI: 10.3390/sym12010021.

[26] Y. Qiu and Y. Wan, "Facial expression recognition based on landmarks," 2019. DOI: 10.1109/IAEAC47372.2019.8997580.

[27] M. I. Munasinghe, "Facial expression recognition using facial landmarks and random forest classifier," 2018. DOI: 10.1109/ICIS.2018.8466510.

[28] R. S. Raj, D. Pratiba, and R. P. Kumar, "Facial expression recognition using facial landmarks: A novel approach," *Advances in Science, Technology and Engineering Systems*, vol. 5, 5 2020, ISSN: 24156698. DOI: 10.25046/aj050504.

[29] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines," *Sensors*, vol. 13, pp. 7714–7734, 2013, ISSN: 1424-8220. DOI: 10.3390/s130607714. [Online]. Available: www.mdpi.com/journal/sensorsArticle.

[30] S. L. Happy, A. George, and A. Routray, "A real time facial expression classification system using local binary patterns," in *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, IEEE, Dec. 2012. DOI: 10.1109/ihci.2012.6481802. [Online]. Available: https://doi.org/10.1109%2Fihci.2012.6481802.

[31] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," vol. 76, pp. 7803–7821, 2017. DOI: 10.1007/s11042-016-3418-y.

[32] M. Mishra. "Convolutional neural networks, explained." (2020), [Online]. Available: https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939 (visited on 10/20/2022).

[33] J. Li, Y. Wang, J. See, and W. Liu, "Micro-expression recognition based on 3d flow convolutional neural network," *Pattern Analysis and Applications*, vol. 22, pp. 1331–1339, 2019. DOI: 10.1007/s10044-018-0757-5. [Online]. Available: https://doi.org/10.1007/s10044-018-0757-5.

[34] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017, ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2016.07.026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320316301753.

[35] R. Breuer and R. Kimmel, "A deep learning perspective on the origin of facial expressions," 2017.

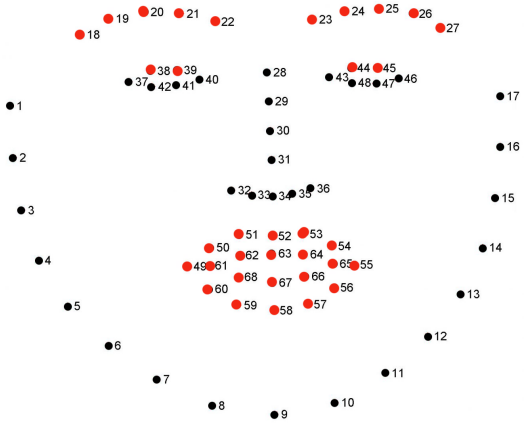
[36] T. S. Wingenbach, C. Ashwin, and M. Brosnan, "Validation of the amsterdam dynamic facial expression set ' bath intensity variations (adfes-biv): A set of videos expressing low, intermediate, and high intensity emotions," *PLoS ONE*, vol. 11, 1 Jan. 2016, ISSN: 19326203. DOI: 10.1371/journal.pone.0147112.

[37] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[38] C. Turan. "A new era in human computer interaction: Facial expression recognition." (2017), [Online]. Available: https://signalprocessingsociety.org/publications-resources/blog/new-era-human-computer-interaction-facial-expression-recognition (visited on 10/25/2022).

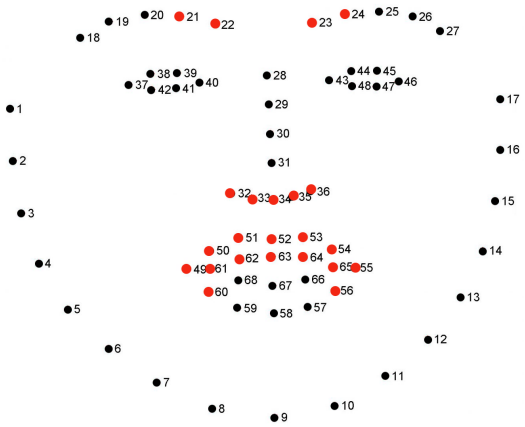
# Appendix

## Selected facial landmarks for each emotion



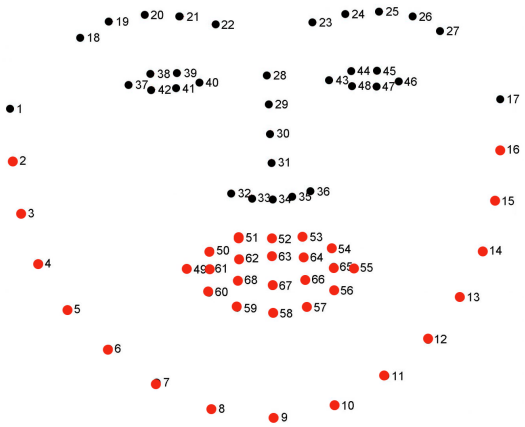
**(a)** Selected landmarks for the facial expression of anger.

**(b)** Selected landmarks for the facial expression of contempt.



**(c)** Selected landmarks for the facial expression of disgust.

**(d)** Selected landmarks for the facial expression of fear.

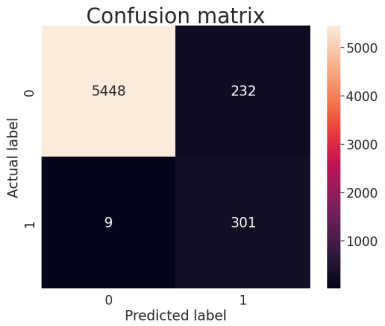


**(e)** Selected landmarks for the facial expression of joy.

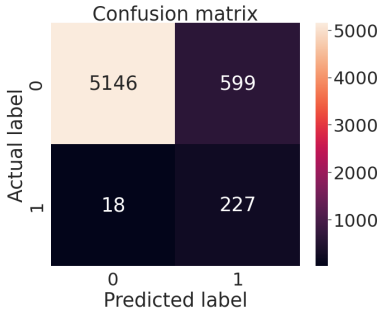
**(f)** Selected landmarks for the facial expression of sadness.


**(g)** Selected landmarks for the facial expression of surprise.

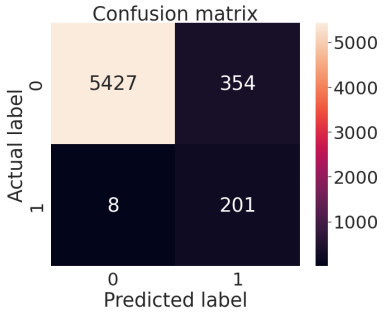
**Figure 1:** Selected landmarks for each emotion after the evaluation of the most informative landmarks.



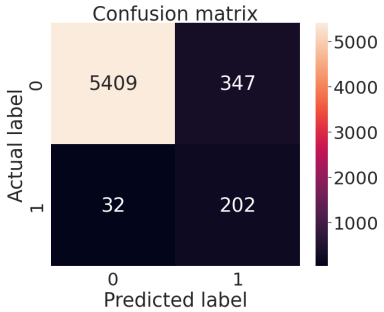

**(a)** Training confusion matrix of the Anger classifier.



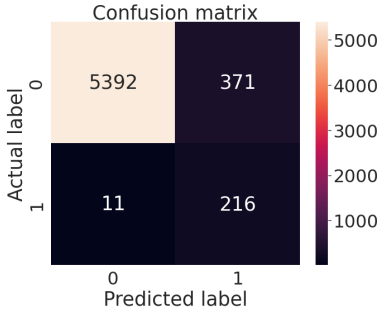

**(b)** Training confusion matrix of the Contempt classifier.



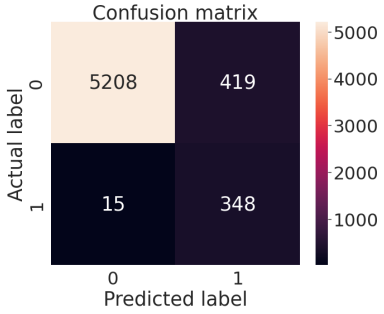

**(c)** Training confusion matrix of the Disgust classifier.



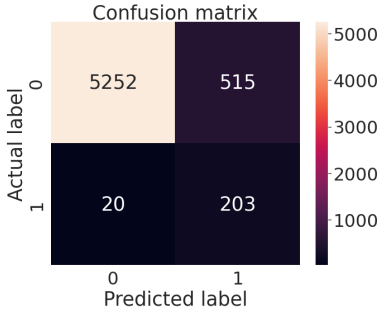

**(d)** Training confusion matrix of the Fear classifier.



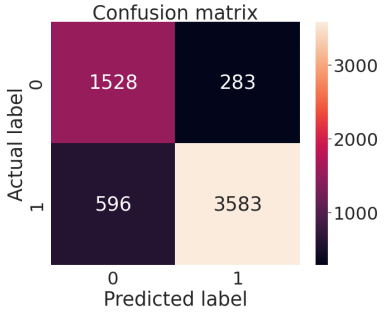

**(e)** Training confusion matrix of the Joy classifier.




**(f)** Training confusion matrix of the Sadness classifier.

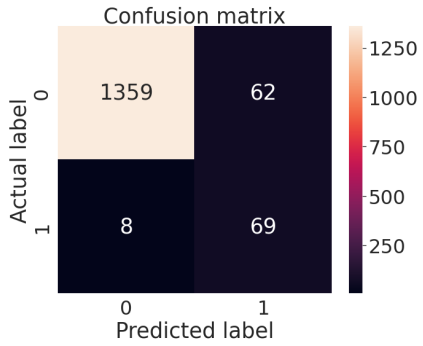



**(g)** Training confusion matrix of the Surprise classifier.

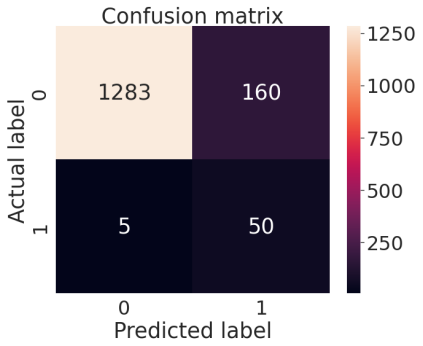



**(h)** Training confusion matrix of the Neutral classifier.

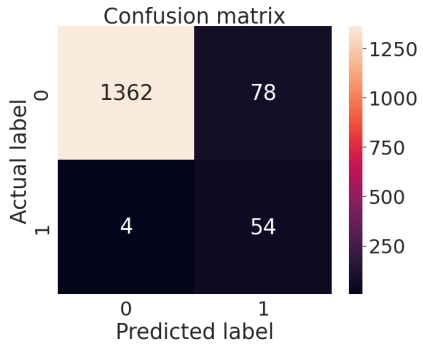
**Figure 2:** Training confusion matrices of the SVM classifiers.



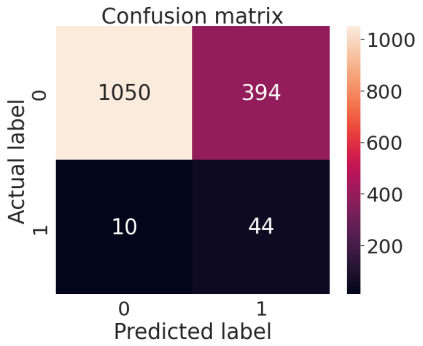

**(a)** Testing confusion matrix of the Anger classifier.



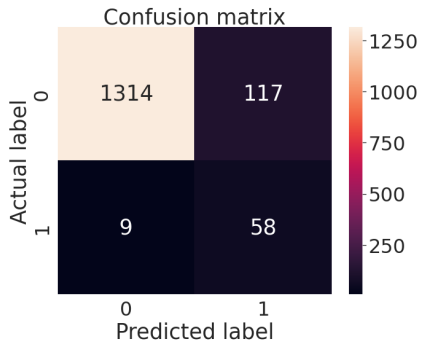

**(b)** Testing confusion matrix of the Contempt classifier.



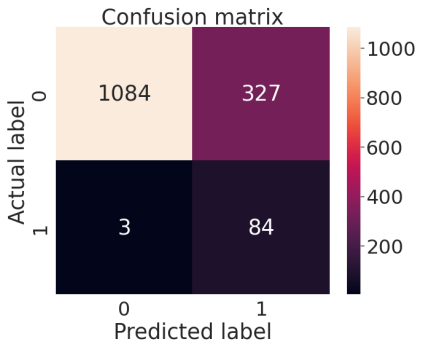

**(c)** Testing confusion matrix of the Disgust classifier.



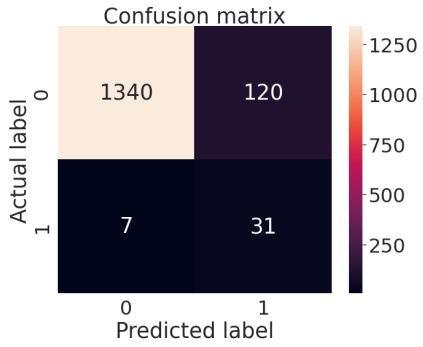

**(d)** Testing confusion matrix of the Fear classifier.



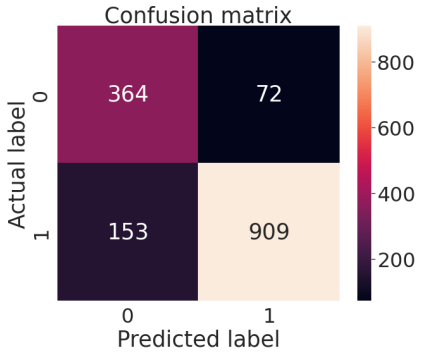

**(e)** Testing confusion matrix of the Joy classifier.




**(f)** Testing confusion matrix of the Sadness classifier.

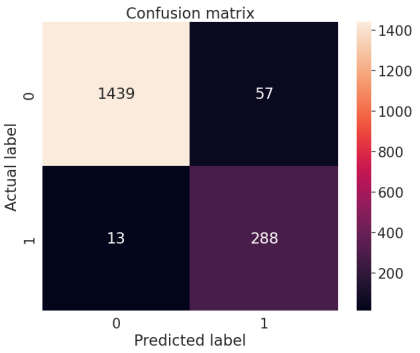



**(g)** Testing confusion matrix of the Surprise classifier.

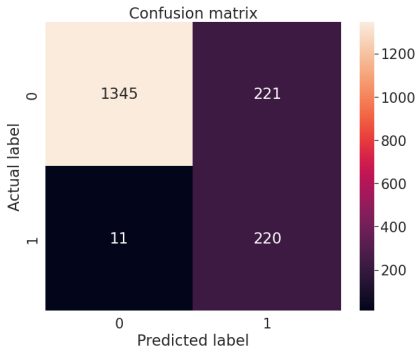



**(h)** Testing confusion matrix of the Neutral classifier.

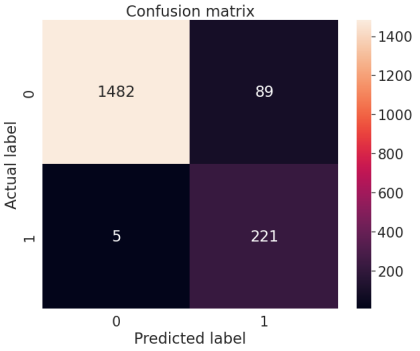
**Figure 3:** Testing confusion matrices of the SVM classifiers.



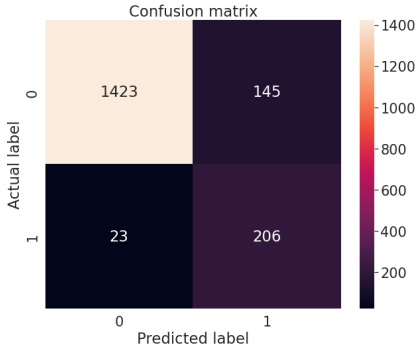

**(a)** Training confusion matrix of the Anger classifier.



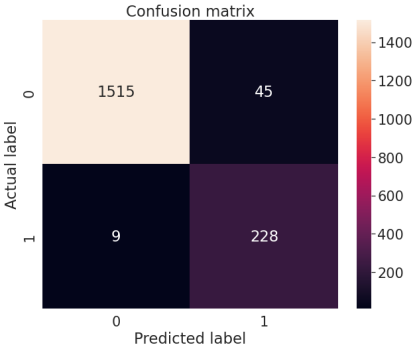

**(b)** Training confusion matrix of the Contempt classifier.



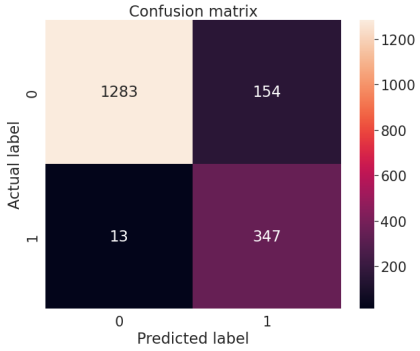

**(c)** Training confusion matrix of the Disgust classifier.



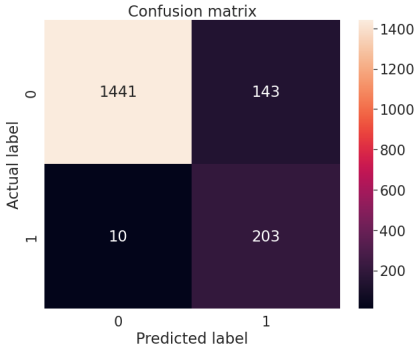

**(d)** Training confusion matrix of the Fear classifier.




**(e)** Training confusion matrix of the Joy classifier.

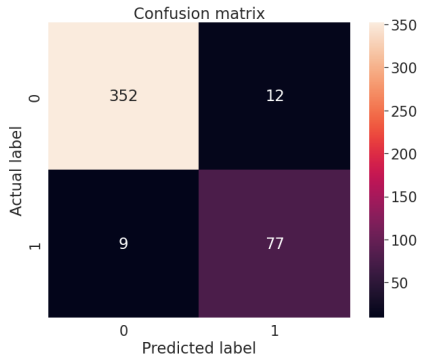



**(f)** Training confusion matrix of the Sadness classifier.

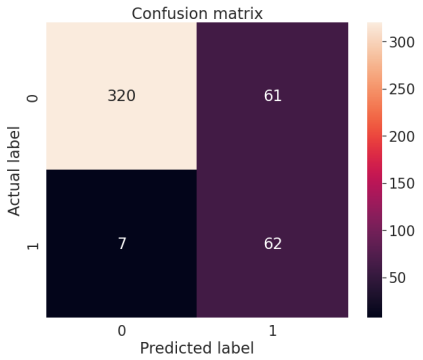



**(g)** Training confusion matrix of the Surprise classifier.

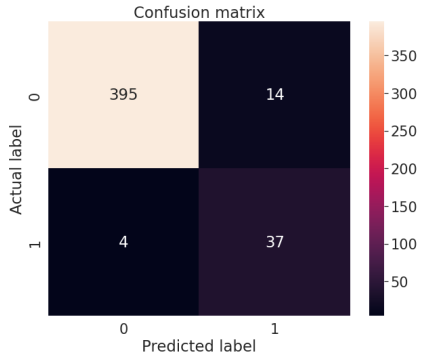
**Figure 4:** Training confusion matrices of the SVM classifiers without the data corresponding to the neutral facial expression.



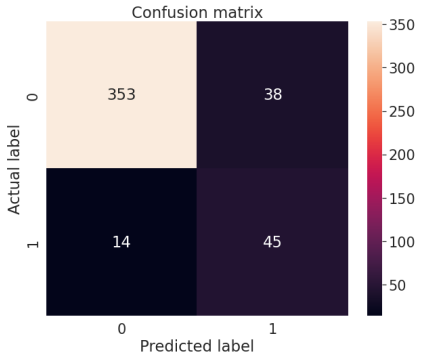

**(a)** Testing confusion matrix of the Anger classifier.



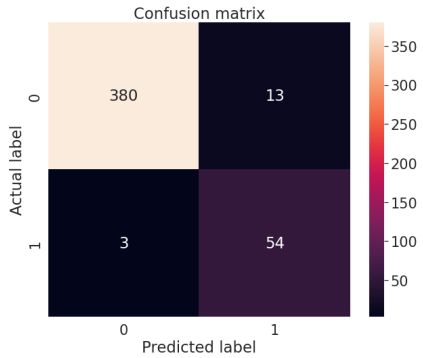

**(b)** Testing confusion matrix of the Contempt classifier.



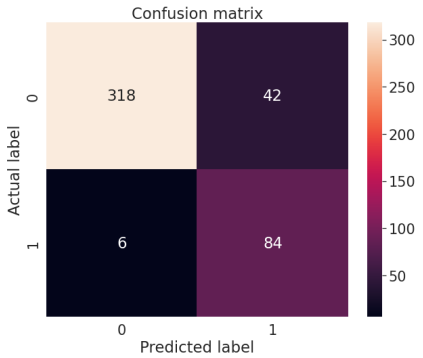

**(c)** Testing confusion matrix of the Disgust classifier.



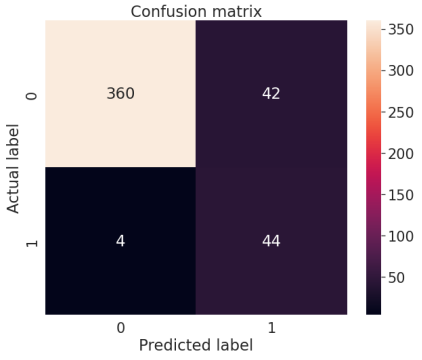

**(d)** Testing confusion matrix of the Fear classifier.




**(e)** Testing confusion matrix of the Joy classifier.




**(f)** Testing confusion matrix of the Sadness classifier.




**(g)** Testing confusion matrix of the Surprise classifier.

**Figure 5:** Testing confusion matrices of the SVM classifiers without the data corresponding to the neutral facial expression.