



OPEN ACCESS

EDITED BY

Panagiotis Tzirakis,
Hume AI, United States

REVIEWED BY

Trasha Gupta,
Delhi Technological University, India
Gunjan Verma,
GL Bajaj Institute of Technology and
Management (GLBITM), India

*CORRESPONDENCE

Marciano Vargas-Treviño
✉ mvargas.cat@uabjo.mx
Horacio Rostro-Gonzalez
✉ horacio.rostro@iqs.url.edu

RECEIVED 01 January 2025
ACCEPTED 28 February 2025
PUBLISHED 19 March 2025

CITATION

Gonzalez-Acosta AMS, Vargas-Treviño M,
Batres-Mendoza P, Guerra-Hernandez EI,
Gutierrez-Gutierrez J, Cano-Perez JL,
Solis-Arrazola MA and Rostro-Gonzalez H
(2025) The first look: a biometric analysis of
emotion recognition using key facial features.
Front. Comput. Sci. 7:1554320.
doi: 10.3389/fcomp.2025.1554320

COPYRIGHT

© 2025 Gonzalez-Acosta, Vargas-Treviño,
Batres-Mendoza, Guerra-Hernandez,
Gutierrez-Gutierrez, Cano-Perez,
Solis-Arrazola and Rostro-Gonzalez. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

The first look: a biometric analysis of emotion recognition using key facial features

Ana M. S. Gonzalez-Acosta¹, Marciano Vargas-Treviño^{1*},
Patricia Batres-Mendoza¹, Erick I. Guerra-Hernandez¹,
Jaime Gutierrez-Gutierrez¹, Jose L. Cano-Perez¹,
Manuel A. Solis-Arrazola^{1,2} and Horacio Rostro-Gonzalez^{2,3*}

¹Laboratory of Artificial Intelligence, Robotics and Control, Faculty of Biological Systems and Technological Innovations, Benito Juárez Autonomous University of Oaxaca, Oaxaca, Mexico, ²Department of Electronics Engineering, División de Ingenierías del Campus Irapuato-Salamanca, University of Guanajuato, Salamanca, Mexico, ³Grupo de Ingeniería de Productos Industriales - Research Group, School of Engineering - Instituto Químico de Sarrià, Universitat Ramon Llull, Barcelona, Spain

Introduction: Facial expressions play a crucial role in human emotion recognition and social interaction. Prior research has highlighted the significance of the eyes and mouth in identifying emotions; however, limited studies have validated these claims using robust biometric evidence. This study investigates the prioritization of facial features during emotion recognition and introduces an optimized approach to landmark-based analysis, enhancing efficiency without compromising accuracy.

Methods: A total of 30 participants were recruited to evaluate images depicting six emotions: anger, disgust, fear, neutrality, sadness, and happiness. Eye-tracking technology was utilized to record gaze patterns, identifying the specific facial regions participants focused on during emotion recognition. The collected data informed the development of a streamlined facial landmark model, reducing the complexity of traditional approaches while preserving essential information.

Results: The findings confirmed a consistent prioritization of the eyes and mouth, with minimal attention allocated to other facial areas. Leveraging these insights, we designed a reduced landmark model that minimizes the conventional 68-point structure to just 24 critical points, maintaining recognition accuracy while significantly improving processing speed.

Discussion: The proposed model was evaluated using multiple classifiers, including Multi-Layer Perceptron (MLP), Random Decision Forest (RDF), and Support Vector Machine (SVM), demonstrating its robustness across various machine learning approaches. The optimized landmark selection reduces computational costs and enhances real-time emotion recognition applications. These results suggest that focusing on key facial features can improve the efficiency of biometric-based emotion recognition systems without sacrificing accuracy.

KEYWORDS

emotion recognition, eye-tracking analysis, facial landmarks, biometric validation, machine learning and AI

1 Introduction

Facial expressions are a universal and integral component of human communication, providing crucial insights into emotions, intentions, and psychological states (Bernhardt, 2022; Hwang and Matsumoto, 2015; Jack and Schyns, 2015). From subtle microexpressions to overt displays of feelings, the ability to accurately interpret these signals is essential for

effective social interaction (Iwasaki and Noguchi, 2016; Jain and Bhakta, 2024; Wahid et al., 2023). While humans naturally excel at recognizing emotions, the underlying mechanisms driving this process remain an area of significant interest within the fields of psychology, neuroscience, and computational modeling (Aday et al., 2017; Eppel, 2018; Malezieux et al., 2023; Shackman and Wager, 2019). One question that has garnered attention is the extent to which specific facial regions, such as the eyes and mouth, dominate emotion recognition and whether other parts of the face play a role (Hernandez-Matamoros et al., 2015; Lekdioui et al., 2017; Wegryn et al., 2017).

Previous research has consistently suggested that the eyes and mouth are the most salient features in emotion detection (De Carolis et al., 2023; Guarnera et al., 2015; Ko, 2018). Studies rooted in Ekman's theory of basic emotions have shown that these regions convey critical information about emotional states such as anger, happiness, sadness, and fear (Keltner et al., 2002). However, many of these findings are based on self-reported data or experimental designs that do not incorporate objective biometric measures (Coppini et al., 2023; López et al., 2022; McStay, 2020). This limitation raises questions about the validity of these claims and whether other facial features might contribute to emotion recognition in ways that have not yet been fully understood.

To address this gap, our study employs eye-tracking technology to provide a more precise and quantifiable analysis of how individuals process facial expressions (Lim et al., 2020; Tarnowski et al., 2020). Eye-tracking allows researchers to observe gaze patterns, identifying the specific facial regions that participants prioritize when recognizing emotions (Carter and Luke, 2020; Hickson et al., 2019). This approach offers a unique opportunity to validate whether the eyes and mouth are indeed the primary focus during emotion recognition and to explore whether attention is distributed across other areas of the face (Vehlen et al., 2021).

Beyond advancing our understanding of human behavior, this research also holds significant implications for practical applications. One key area is the development of computational models for facial emotion recognition, which rely on facial landmark detection to identify emotional states (Chitti et al., 2025; Mukhiddinov et al., 2023; Vajayanthi and Arunnehr, 2024). Traditional landmark models, such as the widely used 68-point configuration, encompass numerous reference points across the face (Wu and Ji, 2019). While comprehensive, these models can be computationally intensive, particularly in real-time scenarios. By identifying the most critical landmarks for emotion detection, we aim to propose a streamlined model that reduces complexity without sacrificing accuracy.

The potential applications of such a model are extensive, ranging from psychology and neuroscience to technology-driven fields such as artificial intelligence and robotics (Schmitz-Hübsch et al., 2024). For instance, emotion-aware systems could benefit from faster and more efficient facial expression analysis, enabling more responsive interactions in environments such as healthcare, education, and customer service.

This study, therefore, has two primary objectives. First, to validate, through biometric evidence, whether the eyes and mouth are the dominant regions observed during emotion recognition or if attention is distributed across other facial areas. Second, to leverage these findings to propose a reduced landmark model for

emotion detection, paving the way for more efficient computational approaches. The full and reduced model has been validated using MLP, RDF, and SVM, demonstrating its effectiveness in maintaining accuracy while optimizing processing efficiency.

The remainder of the paper is structured as follows: Section 2 presents the methodology, outlining key research questions, the dataset used, the experimental setup, and the classification methods applied. Section 3 details the obtained results, while Section 4 provides the conclusions of the study.

2 Materials and methods

Figure 1 presents a general overview of the proposed methodology, which consists of 5 stages. The first stage involves defining the dataset, for which we used FACES (Ebner et al., 2010), a widely utilized database that allows for reliable comparison of results. This is followed by a biometric analysis using eye-tracking technology applied to 30 participants. This step provides valuable insights into the areas of the face most frequently observed during emotion recognition. Based on this biometric data, the next stage focuses on reducing the number of facial landmarks. Finally, an artificial neural network (ANN) is employed to evaluate the accuracy of emotion recognition using the reduced set of landmarks. The results are then compared against the performance achieved with the full set of landmarks. Each of these stages is described in detail in the following sections.

2.1 Research purpose and questions

In Figure 2, we see an adult woman with a happy expression on his face. For us as humans, recognizing the expression in this image is effortless, as we have learned to interpret people's emotions throughout our development as a fundamental aspect of social interaction. However, from a human behavior perspective, an intriguing question arises:

Question 1: What is the first thing we look at to identify someone's expression?

At first glance, the answer seems obvious. When asking this question to a group of 30 participants, the unanimous response is the eyes, followed by the mouth. To validate this assumption, we conducted an initial experiment with a straightforward design: participants were shown images of individuals expressing various emotions and were asked to identify the emotion via a questionnaire.

Question 2: What are the primary facial regions we focus on during facial expression recognition?

As mentioned earlier, the obvious answer to the question of what we observe first when recognizing facial expressions is the eyes. However, it is crucial to biometrically validate this assumption while also examining which other regions, aside from the eyes, are observed during this process. To achieve this, we have defined four Areas of Interest (AOIs), as will be described in Section 2.5. These areas correspond to each eye, the nose, and the mouth. Although the nose is not typically mentioned by participants when asked, it plays an essential role in the transition from observing the eyes to the mouth, as this shift naturally occurs through the nasal region.

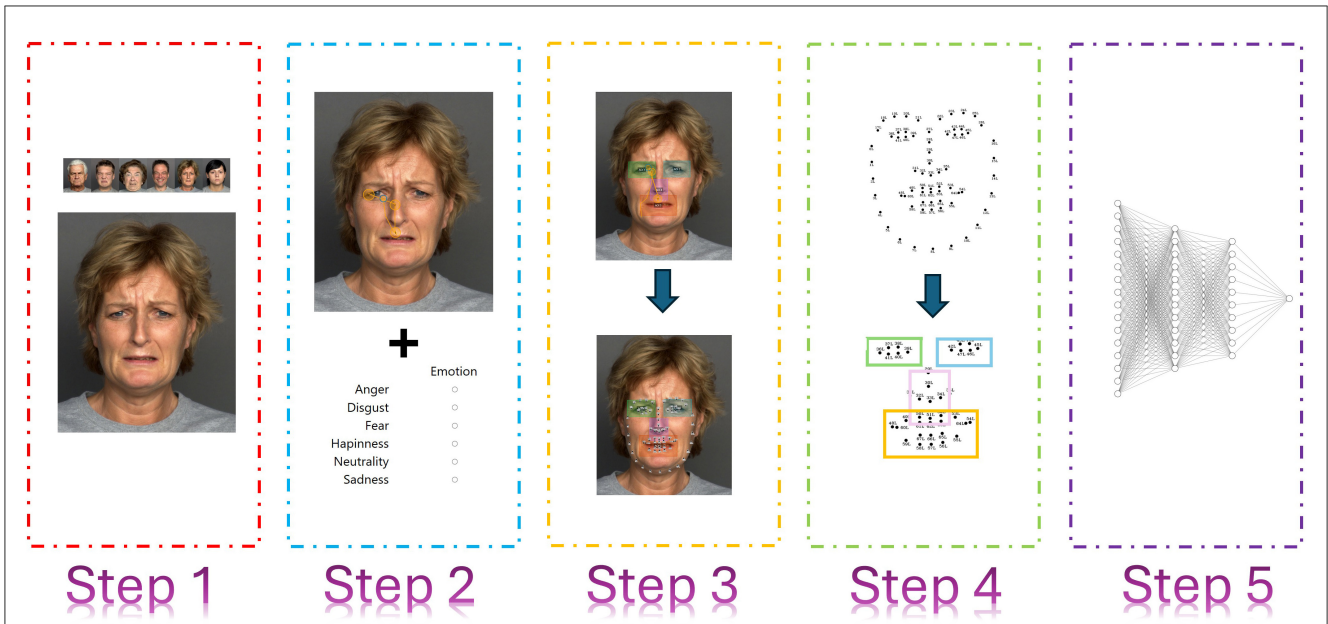
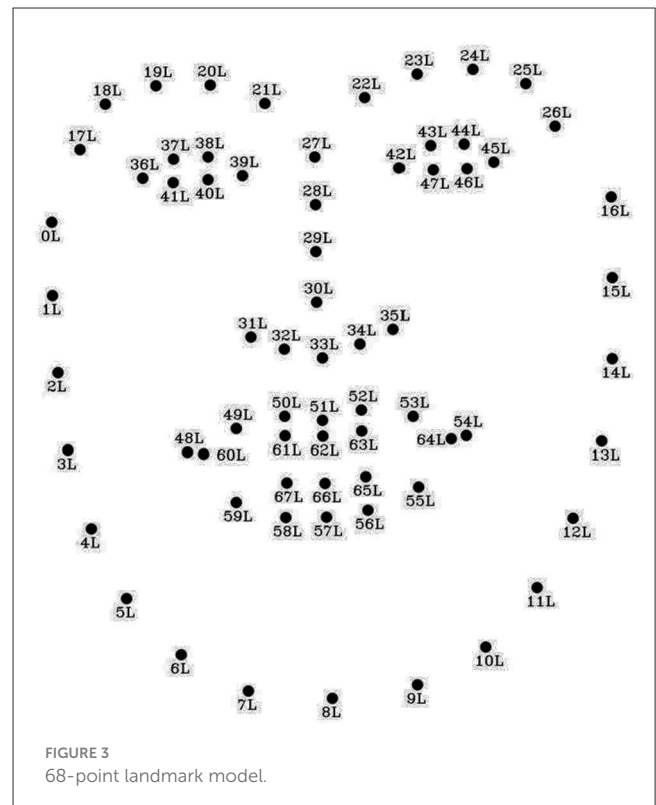


FIGURE 1

General workflow of the study. Step 1 involves defining the dataset. Steps 2, 3 focus on biometric analysis. Step 4 entails feature extraction, while Step 5 involves emotion recognition using three classification algorithms (MLP, RDF, and SVM), leveraging the features obtained from the biometric analysis. Original facial image taken from Ebner et al. (2010).



Question 3: Are there facial regions that we ignore during the recognition of expressions?

In the recognition of faces, emotions, or facial expressions using computational techniques, a commonly employed standard is the

68-point landmark analysis (Figure 3). This method maps 68 points onto the face, essentially creating a detailed outline of its structure. These points clearly define features such as the eyes, mouth, nose, and the facial contour. However, a key question arises: are all these



points truly analyzed when we recognize an expression on another person, or are there certain points that we tend to overlook during this process?

Question 4: Can this study lead to the definition of new analysis zones or modifications to existing ones in facial expression recognition?

The final and one of the most significant questions in this study is whether a biometric analysis of facial expression/emotion recognition can lead to a more efficient computational approach for processing facial information. By efficiency, we refer not only to a faster process but also to one that maintains the performance levels of current algorithms. This question also aims to explore the potential for defining new analysis strategies in human-computer interaction systems or service robots, ensuring a balance between computational efficiency and algorithmic accuracy.

2.2 The data

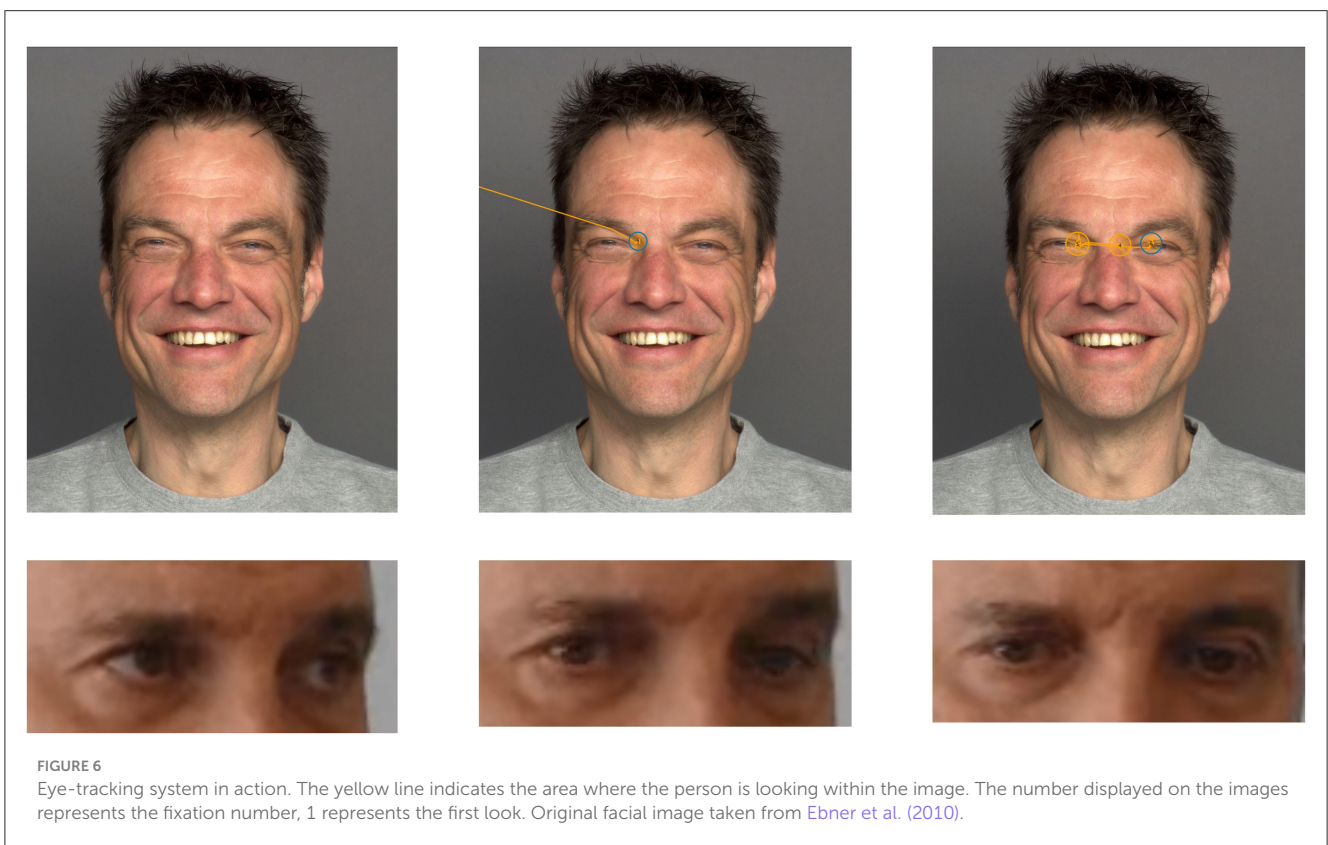
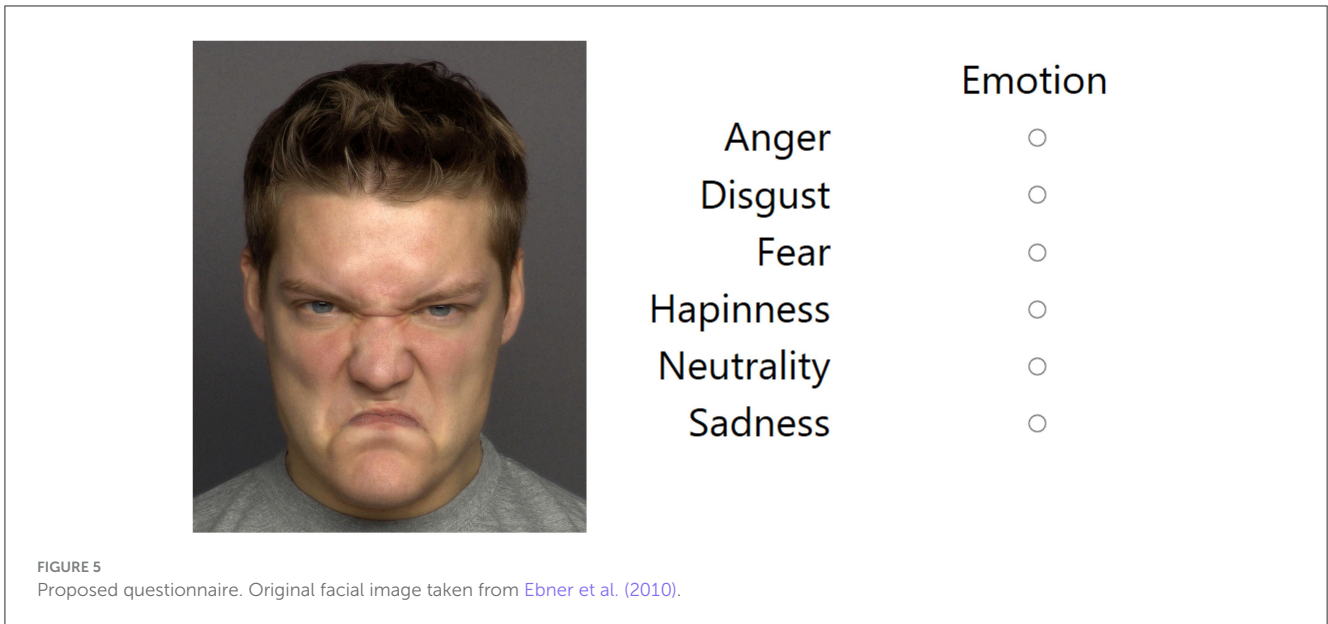
To ensure the validity of our study, we selected a dataset that is well-recognized and contains images with clearly identifiable

facial expressions (emotions). For this reason, we chose the FACES dataset ([Ebner et al., 2010](#)) ([Figure 4](#)). FACES is a comprehensive dataset consisting of naturalistic facial images of 171 individuals, including young ($n = 58$), middle-aged ($n = 56$), and older ($n = 57$) women and men. Each individual displays six distinct facial expressions: neutrality, sadness, disgust, fear, anger, and happiness.

From the FACES dataset, we selected only the 6 images shown in [Figure 4](#) and presented them to a group of 30 participants. These participants are divided into 20 individuals aged between 20 and 25 years, and the remaining 10 are between 45 and 55 years old. The gender distribution is 60% male and 40% female. Given that this research was conducted in an international setting, the participants represent diverse ethnic backgrounds. Following this, participants completed a questionnaire, which will be discussed in the next section, to identify the emotion depicted in each image.

2.3 The questionnaire

Once the images to be used were selected, an experiment was designed. It began with a set of instructions provided to the



participants, informing them that they would be shown a series of images followed by a questionnaire. In this questionnaire, they were required to identify the emotion conveyed in each of the presented images.

Figure 5 shows a screenshot of the questionnaire administered to the 30 participants. Each time an image was presented, participants were required to select one of the six options

provided, choosing the one that best matched the emotion they identified in the image. The exposure time for each image was set to 5 seconds.

While the accuracy of their responses regarding the emotion is of secondary importance in this study, the core focus lies in analyzing, through eye-tracking technology, which facial regions participants observe when trying to discern an expression. This

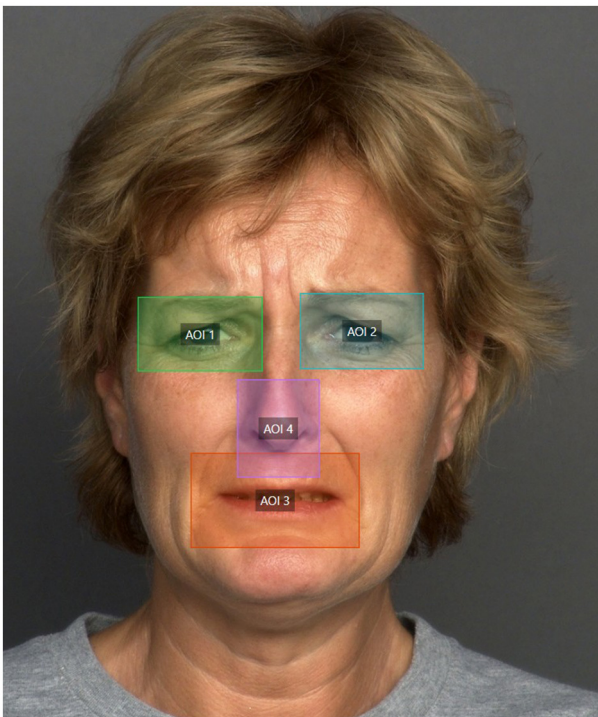


FIGURE 7
Areas of Interest (AOIs) to analyze. Original facial image taken from Ebner et al. (2010).

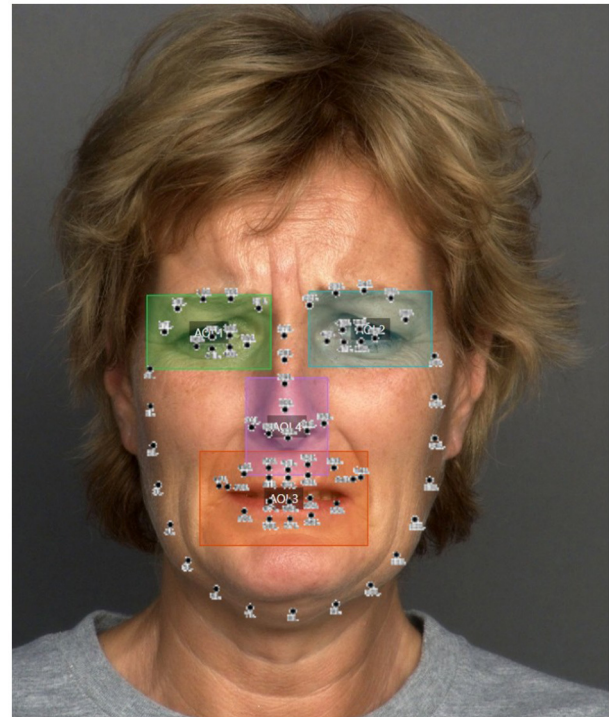


FIGURE 8
Overlay of landmarks on Areas of Interest. Original facial image taken from Ebner et al. (2010).

data provides valuable insights into the cognitive processes underlying emotion recognition and this is explained below.

2.4 Eye-tracking analysis

During the experiment, a Smart Eye AI-X eye-tracking system was used to record the participants' eye activity. The Smart Eye AI-X system is a well-validated eye-tracking technology used in various fields, including cognitive research, human-computer interaction, and biometric analysis (Hartnett et al., 2025; Castner et al., 2024; Lopez-Martinez et al., 2024; Kaliukhovich et al., 2020). Here, we used the Smart Eye AI-X system to extract key fixation-based metrics, particularly dwell time across different AOIs (Areas of Interest), which informed our biometric analysis of facial emotion recognition. The system's high-resolution gaze data enabled us to quantify visual attention patterns, which were later used to refine landmark-based feature extraction for our classification algorithms. Notably, the observed gaze patterns influenced the selection of the most relevant facial regions, helping us optimize the landmark reduction strategy without significant performance loss.

Figure 6 illustrates a segment of the experiment, highlighting an intriguing pattern observed in nearly all participants. In moment 1, we see the participant distracted or not yet engaged. However, in moment 2, upon noticing the appearance of the first image, the participant quickly begins analyzing it. Interestingly, their initial focus is on the left eye, a behavior consistent across

all participants except for one case, which will be explored later. In moment 3, the participant performs a scan between both eyes, further confirming our earlier observation: the eyes are the primary region participants focus on when identifying facial expressions.

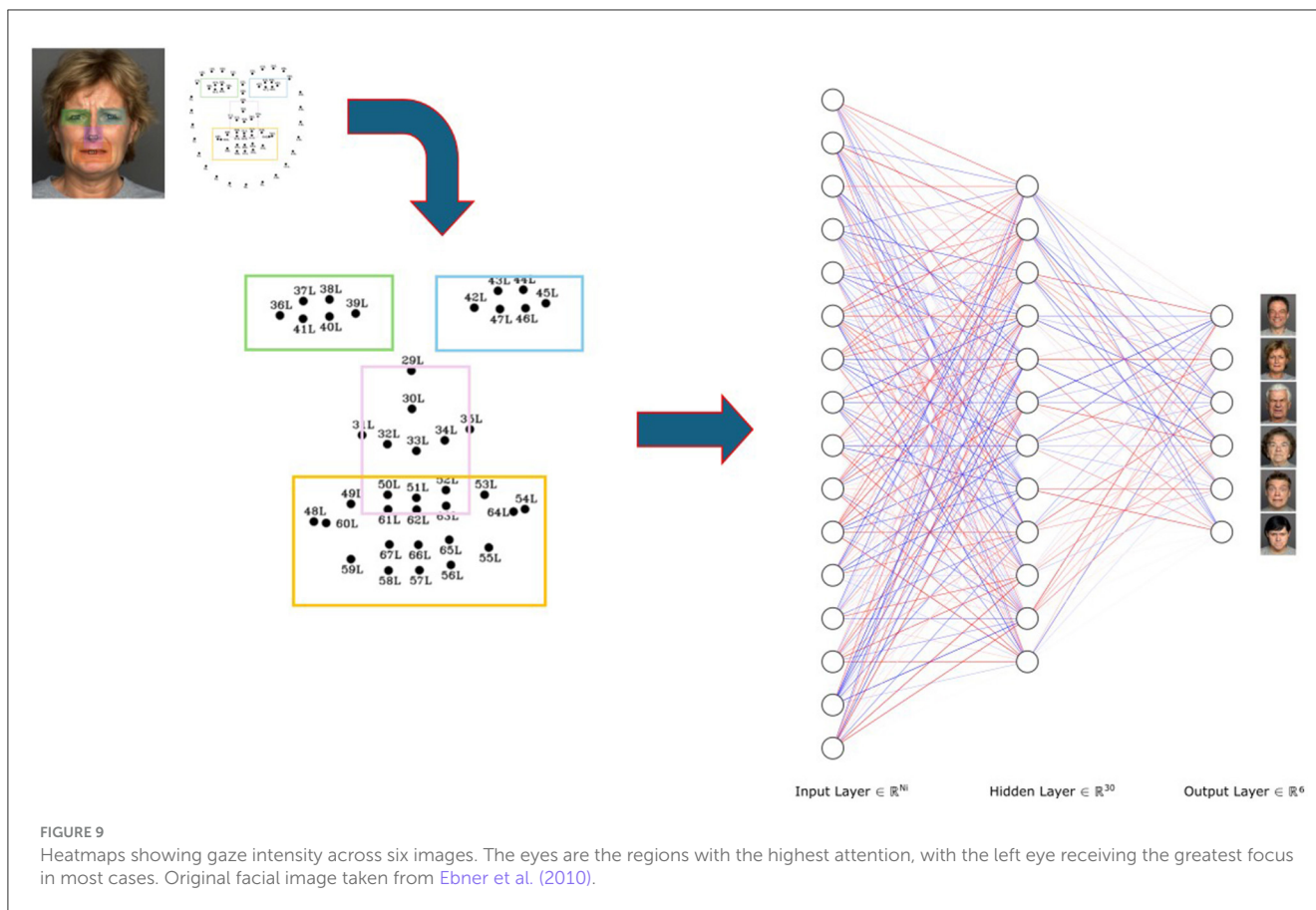
2.5 Defining Areas Of Interest (AOIs)

A preliminary analysis of the eye-tracking data led to the definition of four areas of interest (AOIs) to facilitate a more detailed examination of the results. Specifically, an AOI was assigned to each eye, one to the nose, and another to the mouth. These areas are illustrated in Figure 7.

The size of AOIs is 16.6 cm² for the eye region, 28.3 cm² for the mouth, and 14.4 cm² for the nose. Although the nose is not typically considered by participants as a key region for recognizing emotions or facial expressions, it plays an important role as a transitional area between the eyes and mouth. Therefore, even though it may not initially seem critical for analysis, it is a region with significant ocular activity. This aspect is discussed in detail in Section 3.

2.6 Facial landmarks

It is widely recognized that facial expression recognition relies on identifying key facial elements that enable the extraction of



features associated with specific emotional states. These features are often represented as reference points, commonly known as landmarks, which serve as a framework for analyzing facial expressions. Among the most commonly used standards in this field is the 68-point landmark model (King, 2009), which provides a comprehensive mapping of the face by marking critical areas such as the eyes, eyebrows, nose, and mouth. Figure 3 shows the indexes of the 68 landmark coordinates visualized on the image.

This standard has become a cornerstone in both academic research and practical applications, facilitating tasks ranging from emotion detection to facial recognition and animation. However, while the 68-point model offers detailed and robust facial feature mapping, its complexity and computational demands can pose challenges, particularly in real-time applications or scenarios involving large datasets. In this regard, we explored the use of a reduced set of landmarks derived from the information provided by the analyzed AOIs, aiming to enhance efficiency and response time in human-computer interaction applications.

This section is one of the most interesting aspects of this study. By identifying, through eye-tracking analysis, the areas most relevant to facial expression recognition, specifically, emotions, we aim to reduce the number of landmarks required for expression detection. This is particularly significant when rapid information processing is needed, such as in human-robot communication systems.

In such scenarios, if robots can quickly and accurately recognize human emotions, they will be better equipped to anticipate and deliver more appropriate responses. This capability not only enhances the efficiency of interaction but also improves the overall quality of communication, enabling a more seamless and intuitive exchange between humans and machines.

In Figure 8, we have overlaid the 68-point landmarks on the image with the defined Areas of Interest. This visualization reveals that when a person analyzes the facial expression of another, there are several landmarks that are not actively considered or contribute minimally to the recognition process. This process is biometrically validated through heatmaps generated from the eye-tracking data collected from participants (see Figure 9). These heatmaps provide a visual representation of gaze patterns, highlighting the areas of the face that participants focused on most during the emotion recognition task, thereby supporting the findings of this study.

This observation forms the foundation of our work. By reducing the number of points of interest, we aim to streamline the computational cost associated with facial expression recognition. Traditional landmark models, while comprehensive, often include points that add unnecessary complexity to the analysis without significantly improving accuracy.

Our approach seeks to identify and prioritize only the most critical landmarks for emotion detection, ensuring a more efficient and resource-effective process. This optimization is particularly relevant for real-time applications, where reducing computational

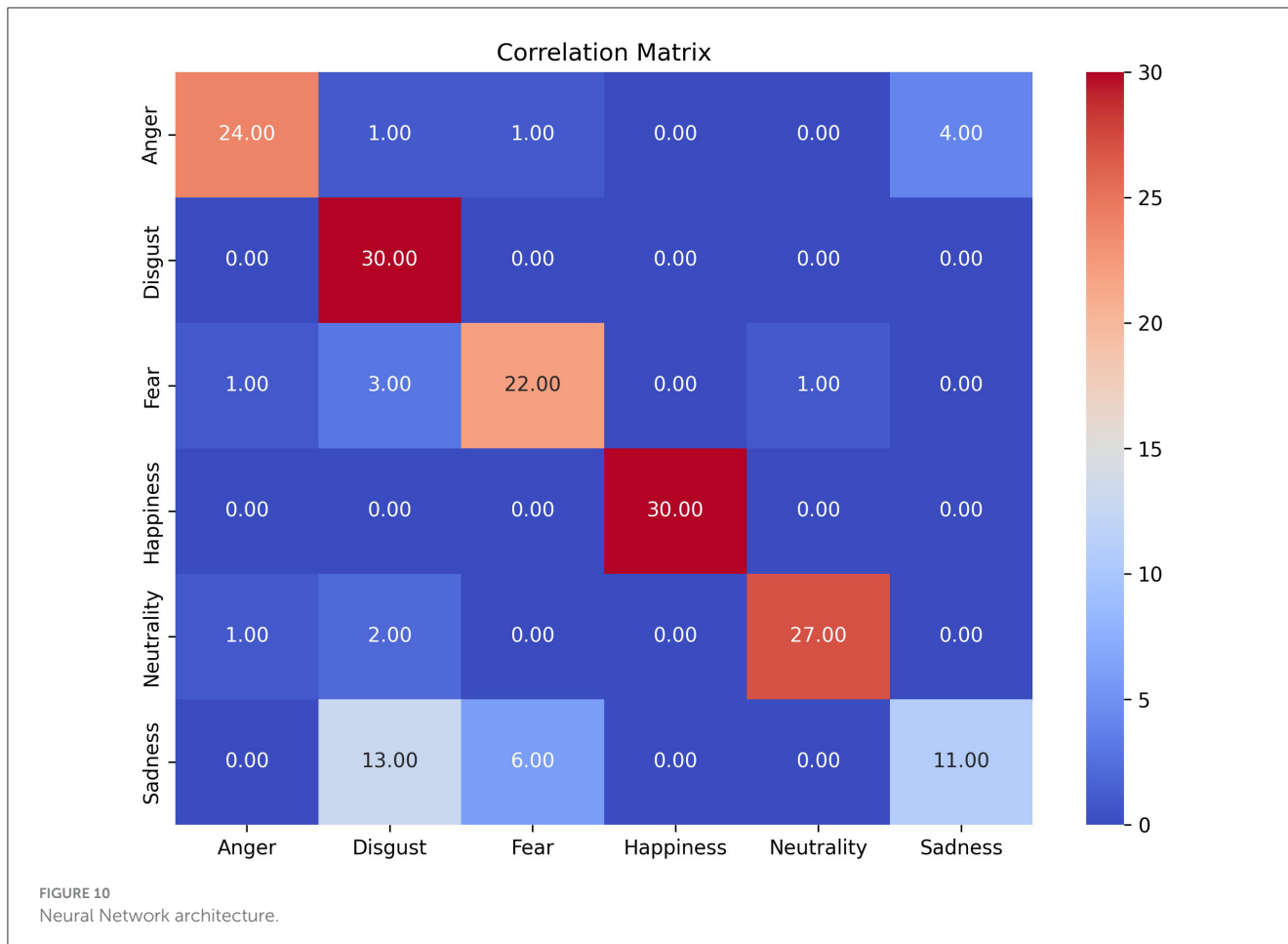


TABLE 1 Parameters of the algorithms used in WEKA.

Algorithm	Parameters	Weka function
RDF	Number of trees ()	<i>RandomForest</i>
	Number of randomly selected attributes per split ($K = \sqrt{\text{totalFeatures}}$)	
	Minimum number of instances per leaf ($M = 1.0$)	
SVM	Cost ($C = 1.0$)	<i>LibSVM</i>
	Gamma ($G = 0.0$)	
	Kernel type ($k = \text{Lineal}$)	
MLP	Learning rate ($L = 0.3$)	<i>MLPClassifier</i>
	Momentum ($M = 0.2$)	
	Training time ($N = 500$)	
	Neurons in the hidden layers ($H = \frac{\text{Features} + \text{Classes}}{2}$)	
	Seed ($S = 0$)	

demands can lead to faster and more responsive systems, paving the way for advancements in emotion recognition technologies across various fields.

2.7 Emotion recognition

In the final stage of the process, emotion recognition was performed using the reduced set of landmarks to validate the performance of our proposed approach. For this, we employed the methodology outlined in Solis-Arrazola et al. (2024), which involves identifying the relationships between the various landmarks by incorporating centroids and drawing segments between them. The core idea is that these segments expand or contract based on the muscle activity associated with specific facial expressions. This dynamic variation allows an artificial intelligence system, such as a Multi-Layer Perceptron, a Support Vector Machine or a Random Decision Forest to capture these relationships and accurately determine the emotion being expressed.

Figure 10 illustrates this process in detail, showcasing how the reduced landmark approach enables efficient and accurate emotion recognition through the interpretation of these relational patterns. In this figure, we present an MLP as the primary classifier, as it outperforms SVM and RDF. However, both SVM and RDF were used in this study for comparative analysis.

The neural network shown in Figure 10 is a Multi-Layer Perceptron (Alpaydin, 2014) and consists of three layers. In the input layer, three different configurations were used: 68, 44, and 24. These configurations correspond to the number of landmarks



used for training. 68 using the full set of landmarks, 44 focusing on the nose, eyes, and mouth, and 24 considering only the eyes and mouth. It is important to note that the landmarks are not directly fed into the network; instead, the segments between these points are used, representing the muscular activation occurring between them (see Figure 14). This captures the expansion and contraction of facial muscles (see Solis-Arrazola et al., 2024 for a more detailed description of this methodology).

In the hidden layer (Layer 2), 30 neurons were utilized, while the output layer consists of 6 only neurons, corresponding to the 6 emotions the network is capable of recognizing. The implementation was carried out using the WEKA programming environment (Hall et al., 2009).

In addition to the MLP neural network, two other classification algorithms were employed for comparison: Support Vector Machine (SVM) and Random Decision Forest (RDF). The SVM is a supervised learning algorithm that finds the optimal hyperplane to separate different classes in the feature space, making it particularly effective for high-dimensional data. RDF, on the other hand, is an ensemble learning method that constructs multiple decision trees and combines their outputs to improve classification accuracy and reduce overfitting. These algorithms were included to benchmark

the performance of the MLP and assess its advantages in emotion recognition. The parameters used for each algorithm are shown in Table 1.

3 Results

To validate our approach, we designed an experiment in which a group of 30 participants was presented with images of individuals expressing the six previously mentioned emotions (anger, sadness, happiness, neutrality, disgust, and fear). Each image was accompanied by a questionnaire (see Figure 5), where participants were asked to indicate, based on their perception, the emotion represented in the displayed image. The images were shown for a duration of 5 seconds. The results obtained are presented below:

3.1 Questionnaire analysis

Although the questionnaire is not particularly relevant to this study, we conducted an analysis using a correlation

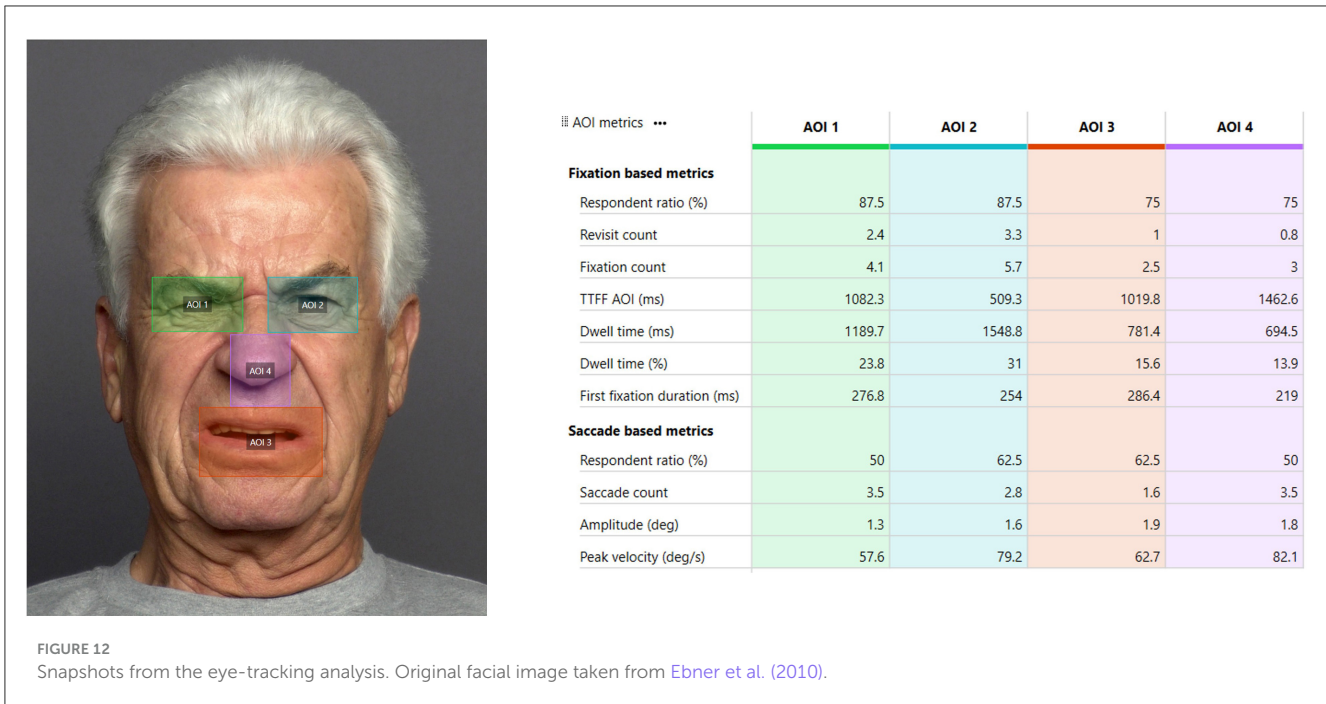


FIGURE 12
Snapshots from the eye-tracking analysis. Original facial image taken from Ebner et al. (2010).

matrix. This analysis allows us to assess the extent to which certain emotions can be confused with others. In this context, Figure 11 presents a correlation matrix illustrating the relationship between the emotion presented and the participants' responses.

The matrix reveals that the emotion "sadness" was the most challenging for participants to recognize, as it was often confused with "fear" and "disgust." This confusion may stem from two factors. The first is that, as shown in Figure 4A, the expression doesn't clearly appear as sadness; it is quite ambiguous and depends largely on the participant's perception. For example, within the research team, some found it more similar to fear, while others, according to the correlation matrix, mostly confused it with disgust.

This confusion is also linked to the proximity of these emotions in the Circumplex Model of Affect, also known as the Valence-Arousal Model, proposed by Russell (1980). In this model, emotions are organized in a two-dimensional space based on their valence (positive or negative) and arousal (high or low). Sadness, fear, and disgust cluster closely together in this diagram due to their shared low valence and moderate-to-high arousal levels. This finding highlights the inherent complexity of recognizing and distinguishing between emotions with overlapping characteristics.

3.2 AOIs + Eye-tracking results

In steps 2 and 3, as defined in the diagram in Figure 1, a biometric analysis of the participants is conducted to evaluate the areas they focus on most when recognizing emotions in the images. For this purpose, four specific Areas of Interest are defined. In this context, Figure 12 presents snapshots captured during

the experiment, illustrating the eye-tracking behavior of various participants at different time intervals. These images provide a detailed depiction of the participants' gaze patterns as they analyzed the displayed facial expressions.

To enhance the analysis, the Areas of Interest have been overlaid onto the images, enabling a more precise validation of the regions that participants focused on most frequently. This visualization not only highlights the dynamic nature of gaze behavior across individuals but also reinforces the consistency of attention directed toward specific facial regions, such as the eyes and mouth. The snapshots also reveal significant eye activity in the nose region, as previously mentioned. This occurs because the transition of gaze between the eyes and the mouth naturally passes through the nose, making it unavoidable. However, this area is excluded in subsequent analyses for the purpose of automatic emotion recognition by the classification algorithm. These insights are crucial for understanding how humans prioritize certain facial features during emotion recognition, serving as a foundation for the development of more efficient landmark-based recognition models.

Figure 13 presents a table highlighting some of the measurements derived from eye-tracking data for the image of the elderly person in the dataset. The table highlights the section on fixation-based metrics, where the Dwell Time for the selected Areas of Interest is displayed. Dwell Time refers to the total time participants spend observing these regions. In this table, the percentage of observation for these AOIs accounts for 84.3% (~4.2 seconds) of the total gaze time on the image. This indicates the significant relevance of these regions for the current analysis, emphasizing their importance in understanding participants' visual attention patterns.

To perform a more comprehensive analysis, the Dwell Time metrics were calculated for all participants and images in the



experiment. These results are presented in Table 2, where the total exploration percentage of the AOIs is 78.44%, with a total time of 3.92 seconds. While other metrics could be explored, this is the one of greatest interest for the current study.

3.3 Heatmaps results

In this experiment, we analyzed the gaze data of the 30 participants through heatmaps across the six different emotions presented in Figure 4, focusing on identifying the areas that drew the most attention.

The results revealed a consistent pattern: the participants primarily focused on the left eye in most images. However, an exception was observed in the case of the image featuring the elderly man. In this instance, the left eye received less attention due to it being slightly closed as part of the person’s facial expression.

This deviation highlights how specific features of facial expressions can influence gaze behavior, altering the expected patterns of attention distribution. The heatmaps effectively demonstrate these dynamics, offering insights into the collective focus areas and outliers in the dataset.

TABLE 2 Dwell time metrics.

	AOI 1	AOI 2	AOI 3	AOI 4	Total
Young woman	1,235.32	1,329.02	431.71	1,298.04	
Young man	1,262.60	851.16	453.24	921.86	
Adult woman	1,358.78	819.08	625.21	868.38	
Adult man	1,061.46	951.05	851.85	799.90	
Old woman	1,421.83	1,127.47	633.96	1,017,04	
Old man	1,189.65	1,548.77	781.42	694.47	
Dwell Time Mean (ms)	1,254.94	1,104.43	629.57	933.28	3922.21
Dwell Time Percentage (%)	25.10	22.09	12.59	18.67	78.44

3.4 Emotion recognition

In the final stage of this research, the three classification algorithms (MLP, SVM, and RDF) were implemented to recognize emotions based on facial muscle activation. Using the biometric analysis conducted,

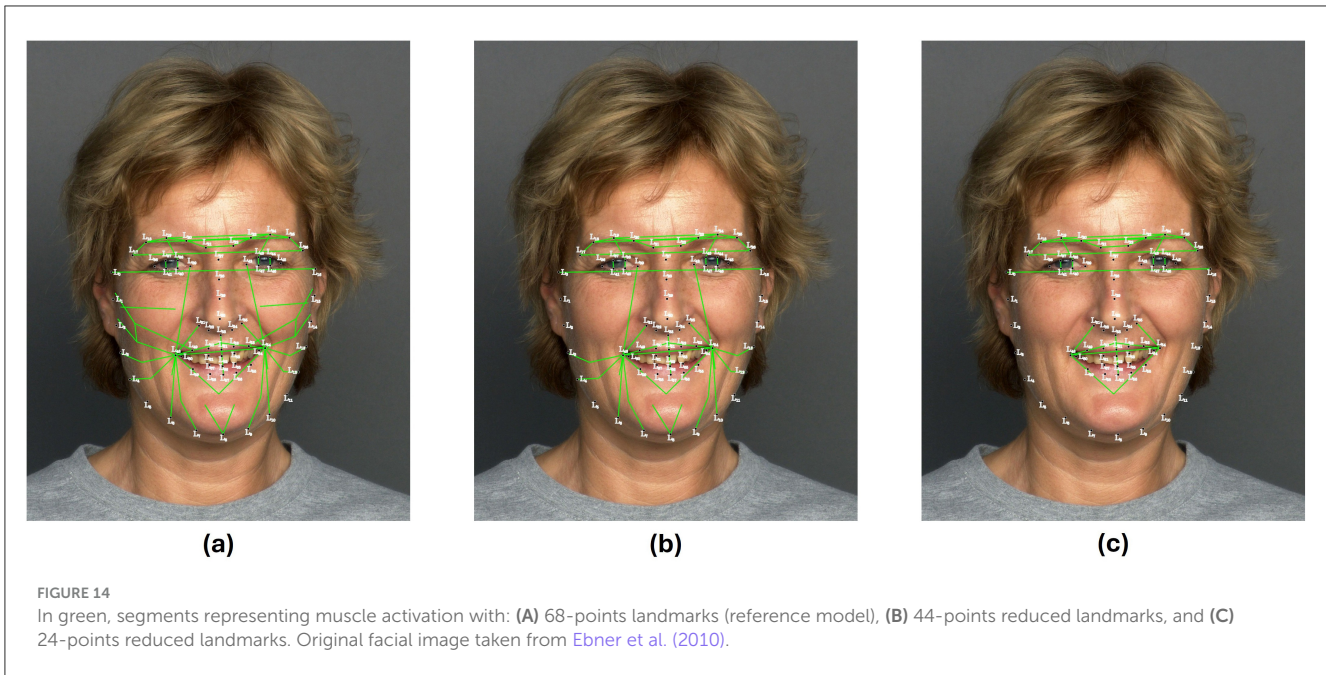


TABLE 3 Performance of the algorithms across different demographic groups using 68 facial landmarks.

Algorithm	Demographic group	Accuracy (%)	Precision	Recall	F1-score
MLP	Old male	76.50	0.763	0.765	0.763
	Old female	80.20	0.805	0.802	0.801
	Middle age male	87.40	0.873	0.874	0.873
	Middle age female	91.40	0.915	0.914	0.913
	Young male	93.40	0.935	0.934	0.934
	Young female	92.50	0.926	0.925	0.925
SVM	Old male	75.00	0.748	0.75	0.748
	Old female	75.90	0.757	0.759	0.757
	Middle age male	82.50	0.827	0.825	0.823
	Middle age female	88.30	0.883	0.883	0.882
	Young male	89.90	0.904	0.899	0.881
	Young female	89.40	0.90	0.894	0.895
RDF	Old male	72.60	0.72	0.726	0.719
	Old female	76.40	0.754	0.764	0.757
	Middle age male	83.90	0.834	0.839	0.835
	Middle age female	87.00	0.869	0.87	0.866
	Young male	86.50	0.867	0.865	0.863
	Young female	91.40	0.917	0.914	0.914

the areas most frequently observed during emotion recognition were identified. These areas allowed for a redefinition of the classic 68-point standard used by most algorithms.

Figure 14 illustrates the three different input configurations used to train the algorithms. In Figure 14A, the full set of 68 facial landmarks is shown, from which 54 segments were established to simulate the muscular connections between these points. In

Figure 14B, only the landmarks within the 4 defined AOIs were considered. Finally, in Figure 14C, the configuration was further reduced to include just 24 landmarks, considering only facial landmarks in mouth and eyes.

Tables 3-5 present the performance of the algorithms across the three study cases. It is evident that the group of elderly individuals poses the greatest challenge for the algorithms to learn. We attribute this to the increased presence of wrinkles in older

TABLE 4 Performance of the algorithms across different demographic groups using 44 facial landmarks.

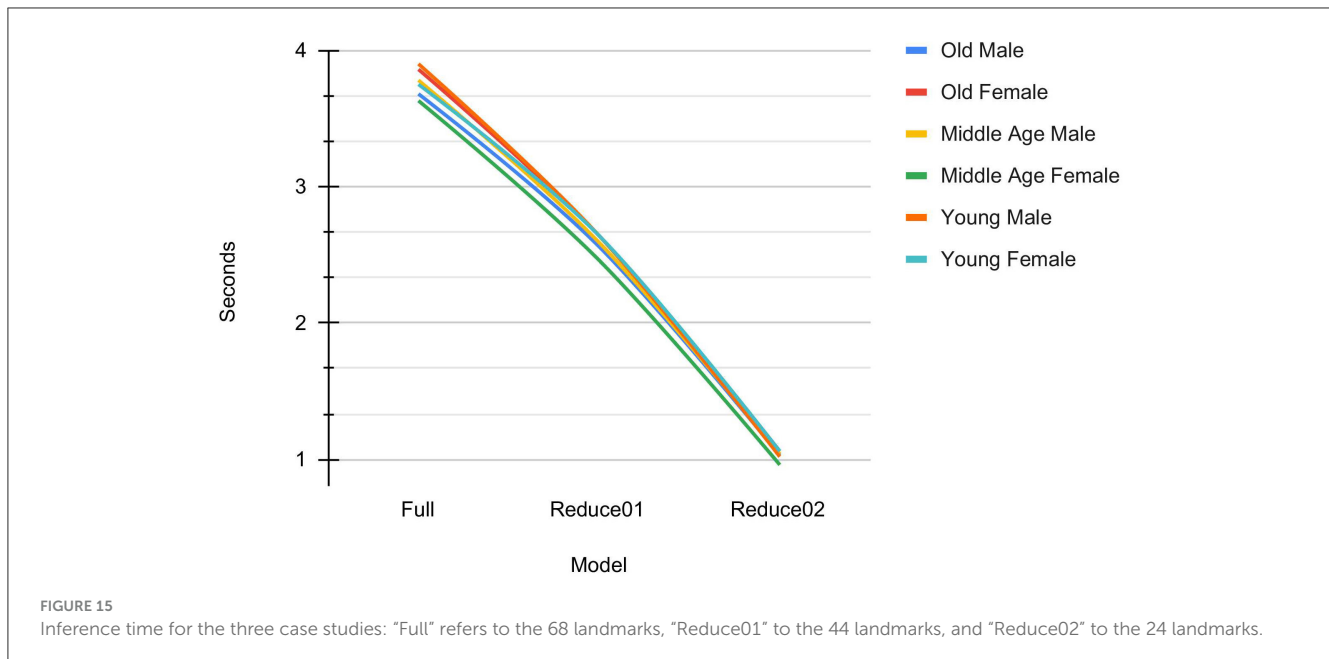
Algorithm	Demographic group	Accuracy (%)	Precision	Recall	F1-score
MLP	Old Male	73.20	0.729	0.732	0.729
	Old female	80.50	0.803	0.805	0.802
	Middle age male	87.10	0.871	0.871	0.87
	Middle age female	90.70	0.907	0.907	0.907
	Young male	93.10	0.932	0.931	0.931
	Young female	92.20	0.923	0.922	0.922
SVM	Old male	76.80	0.768	0.768	0.766
	Old female	77.10	0.771	0.773	0.771
	Middle age male	82.20	0.821	0.822	0.819
	Middle age female	86.70	0.866	0.867	0.865
	Young male	90.50	0.911	0.905	0.904
	Young female	89.40	0.899	0.894	0.895
RDF	Old Male	72.60	0.723	0.726	0.72
	Old female	75.60	0.745	0.756	0.748
	Middle age male	83.30	0.827	0.833	0.828
	Middle age female	85.20	0.852	0.852	0.851
	Young male	89.40	0.895	0.894	0.893
	Young female	89.40	0.898	0.894	0.894

TABLE 5 Performance of the algorithms across different demographic groups using 24 facial landmarks.

Algorithm	Demographic group	Accuracy (%)	Precision	Recall	F1-score
MLP	Old male	72.90	0.725	0.729	0.727
	Old female	78.20	0.785	0.782	0.782
	Middle age male	86.80	0.866	0.868	0.866
	Middle age female	87.30	0.875	0.873	0.874
	Young male	90.20	0.903	0.902	0.902
	Young female	90.50	0.907	0.905	0.904
SVM	Old male	72.00	0.722	0.72	0.719
	Old female	75.90	0.754	0.759	0.755
	Middle age male	83.90	0.839	0.839	0.834
	Middle age female	85.20	0.847	0.852	0.845
	Young male	86.50	0.866	0.865	0.864
	Young female	87.60	0.885	0.876	0.879
RDF	Old male	73.80	0.732	0.738	0.733
	Old female	75.00	0.739	0.75	0.743
	Middle age male	70.10	0.696	0.701	0.698
	Middle age female	68.50	0.691	0.685	0.687
	Young male	86.80	0.867	0.868	0.841
	Young female	87.60	0.879	0.876	0.877

adults, which complicates the definition of facial muscle activation. Further studies on this population could be conducted in the future by incorporating neural information through EEG signals,

allowing us to correlate these signals with facial muscle activation to enhance classification performance. However, for the rest of the images, the results are quite satisfactory. Additionally, it can



be observed that there is no significant loss in performance when reducing the number of points, validating the effectiveness of our proposed approach.

From [Tables 3-5](#), it can be observed that, in general, the algorithm that achieves the best performance in emotion recognition based on muscle activation across the three cases is the MLP. Additionally, the results indicate that younger demographic groups tend to achieve higher accuracy. This is primarily because facial muscle activity is more distinguishable in younger individuals, whereas wrinkles in older adults can obscure these facial features. Finally, it is observed that, in most cases, women perform better in emotion recognition. This is to some extent expected, as women are generally more expressive.

[Figure 15](#) presents a graph depicting the inference time for the three case studies. The results show that reducing the number of landmarks improves the inference time of the algorithms. This finding is particularly relevant for applications in human-robot interaction and embedded systems, where real-time processing is crucial.

4 Conclusion

This study highlights the critical role of the eyes and mouth in emotion recognition, as consistently demonstrated by gaze patterns observed through eye-tracking analysis. These findings provide robust biometric evidence that not all facial landmarks contribute equally to the recognition process, challenging the necessity of traditional 68-point models. By leveraging the insights gained from eye-tracking data, we proposed a reduced facial landmark model that focuses on the most significant regions of the face. Specifically, the model reduces the number of landmarks from 68 to 24, achieving this optimization with only a minimal decrease in accuracy.

To validate the effectiveness of this reduced landmark model, a neural network was employed to classify emotions based on the proposed set of landmarks. The results confirmed that the streamlined model retains a high level of performance while significantly reducing computational demands. The main advantage of this reduction is the lower computational cost, making it particularly suitable for real-time applications in human-computer interaction (HCI), virtual reality (VR), and mobile or embedded systems, where processing power is limited. It is true that this reduction in computational cost may not be noticeable on conventional computers, where the decrease is almost imperceptible. However, when it comes to robotic platforms or edge computing systems, reducing the number of data or features to process can save significant computational resources in terms of both area and processing time. Future research could involve embedding this algorithm in low-cost devices or robotic platforms with limited resources and measuring performance in terms of energy consumption and processing speed.

The implications of this work extend to diverse fields, including robotics, behavioral research, and emotion-aware technologies, where rapid and reliable emotion recognition is crucial. Furthermore, the use of a neural network to validate the reduced model highlights the potential for integrating such approaches into advanced machine learning systems. For future research, we plan to extend the validation of our reduced facial landmark model by testing it on more diverse datasets that include a broader range of ages, ethnic backgrounds, and cultural differences. This will help assess the model's generalizability and its robustness in different populations. Additionally, we aim to explore its applicability in various domains beyond emotion recognition. In healthcare, for instance, this reduced model could be valuable for detecting facial muscle impairments in patients with neurological conditions such as facial paralysis or Parkinson's disease. By reducing

computational costs, it could facilitate real-time monitoring in clinical settings or even be integrated into assistive technologies for non-invasive patient assessment. In robotics, the model could enhance human-robot interaction by improving the ability of robotic systems to interpret human emotions and respond accordingly. Given its low computational requirements, it would be particularly suitable for deployment in embedded systems and edge computing platforms, where processing power is limited.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies involving humans because this study did not receive formal ethics approval from an Institutional Review Board (IRB), as it was conducted in accordance with national and institutional guidelines that do not require IRB approval for non-invasive behavioral studies involving minimal risk. However, informed consent was obtained from all participants before their participation in the study. Each participant was provided with detailed information regarding the purpose, procedures, and potential implications of the research, and they voluntarily agreed to participate. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AG-A: Writing – original draft. MV-T: Writing – review & editing. PB-M: Writing – original draft, Writing – review & editing. EG-H: Writing – original draft. JG-G: Writing – review & editing. JC-P: Writing – original draft. MS-A: Writing – original draft,

Writing – review & editing. HR-G: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research has been supported by the National Council of Humanities, Science and Technology of Mexico (CONAHCYT) through the scholarship 413813.

Acknowledgments

The authors would like to acknowledge Dr. Jan-Hinrich Meyer and Dr. Jorge Matute from the Neuromarketing Laboratory at IQS for their invaluable support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aday, J., Rizer, W., and Carlson, J. M. (2017). "Chapter 2 - neural mechanisms of emotions and affect,?" in *Emotions and Affect in Human Factors and Human-Computer Interaction*, ed. M. Jeon (San Diego: Academic Press), 27–87. doi: 10.1016/B978-0-12-801851-4.00002-1
- Alpaydm, E. (2014). *Introduction to Machine Learning, 3rd Edn.* Cambridge, MA: The MIT Press, 640.
- Bernhardt, C. (2022). *Facial Expressions and Emotions.* Wiesbaden: Springer Fachmedien Wiesbaden, 221–259. doi: 10.1007/978-3-658-36929-3_10
- Carter, B. T., and Luke, S. G. (2020). Best practices in eye tracking research. *Int. J. Psychophysiol.* 155, 49–62. doi: 10.1016/j.ijpsycho.2020.05.010
- Castner, N., Arsiwala-Scheppach, L., Mertens, S., Krois, J., Thaqi, E., Kasneci, E., et al. (2024). Expert gaze as a usability indicator of medical AI decision support systems: a preliminary study. *NPJ Dig. Med.* 7:199. doi: 10.1038/s41746-024-01192-8
- Chitti, E., Actis-Grosso, R., Ricciardelli, P., Olivari, B., Carenzi, C., Tedoldi, M., et al. (2025). Miemo: a multi-modal platform on emotion recognition for children with autism spectrum condition. *Comput. Hum. Behav. Rep.* 17:100549. doi: 10.1016/j.chbr.2024.100549
- Coppini, S., Lucifora, C., Vicario, C. M., and Gangemi, A. (2023). Experiments on real-life emotions challenge Ekman's model. *Sci. Rep.* 13:9511. doi: 10.1038/s41598-023-36201-5
- De Carolis, B., Macchiarulo, N., Palestra, G., De Matteis, A. P., and Lippolis, A. (2023). "Fermouth: facial emotion recognition from the mouth region,?" in *Image Analysis and Processing-ICIAP 2023*, eds. G. L. Foresti, A. Fusiello, E. Hancock (Cham: Springer Nature Switzerland), 147–158. doi: 10.1007/978-3-031-43148-7_13
- Ebner, N. C., Riediger, M., and Lindenberger, U. (2010). FACES-A database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods* 42, 351–362. doi: 10.3758/BRM.42.1.351

- Eppel, A. (2018). *Understanding and Recognizing Emotion*. Cham: Springer International Publishing, 47–70. doi: 10.1007/978-3-319-74995-2_3
- Guarnera, M., Hichy, Z., Cascio, M. I., and Carrubba, S. (2015). Facial expressions and ability to recognize emotions from eyes or mouth in children. *Eur. J. Psychol.* 11, 183–196. doi: 10.5964/ejop.v11i2.890
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *SIGKDD Explor. Newsl.* 11, 10–18. doi: 10.1145/1656274.1656278
- Hartnett, N., Bellman, S., Beal, V., Kennedy, R., Charron, C., and Varan, D. (2025). How to accurately measure attention to video advertising. *Int. J. Advert.* 44, 184–207. doi: 10.1080/02650487.2024.2435164
- Hernandez-Matamoros, A., Bonarini, A., Escamilla-Hernandez, E., Nakano-Miyatake, M., and Perez-Meana, H. (2015). “A facial expression recognition with automatic segmentation of face regions,” in *Intelligent Software Methodologies, Tools and Techniques*, eds. H. Fujita, and G. Guizzi (Cham: Springer International Publishing), 529–540. doi: 10.1007/978-3-319-22689-7_41
- Hickson, S., Dufour, N., Sud, A., Kwatra, V., and Essa, I. (2019). “Eyemotion: classifying facial expressions in VR using eye-tracking cameras,” in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1626–1635. doi: 10.1109/WACV.2019.00178
- Hwang, H., and Matsumoto, D. (2015). *Evidence for the Universality of Facial Expressions of Emotion*. New Delhi: Springer India, 41–56. doi: 10.1007/978-81-322-1934-7_3
- Iwasaki, M., and Noguchi, Y. (2016). Hiding true emotions: micro-expressions in eyes retrospectively concealed by mouth movements. *Sci. Rep.* 6:22049. doi: 10.1038/srep22049
- Jack, R., and Schyns, P. (2015). The human face as a dynamic tool for social communication. *Curr. Biol.* 25, R621–R634. doi: 10.1016/j.cub.2015.05.052
- Jain, A., and Bhakta, D. (2024). Micro-expressions: a survey. *Multimed. Tools Appl.* 83, 53165–53200. doi: 10.1007/s11042-023-17313-6
- Kaliukhovich, D. A., Manyakov, N. V., Bangerter, A., Ness, S., Skalkin, A., Goodwin, M. S., et al. (2020). Social attention to activities in children and adults with autism spectrum disorder: effects of context and age. *Mol. Autism* 11:79. doi: 10.1186/s13229-020-00388-5
- Keltner, D., Ekman, P., Gonzaga, G. C., and Beer, J. (2002). “Facial expression of emotion,” in *Handbook of Affective Sciences* (Oxford University Press). doi: 10.1093/oso/9780195126013.003.0022
- King, D. E. (2009). Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* 10, 1755–1758. doi: 10.5555/1577069.1755843
- Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors* 18:401. doi: 10.3390/s18020401
- Lekdioui, K., Ruichek, Y., Messoussi, R., Chaabi, Y., and Touahni, R. (2017). “Facial expression recognition using face-regions,” in *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 1–6. doi: 10.1109/ATSIP.2017.8075517
- Lim, J. Z., Mountstephens, J., and Teo, J. (2020). Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors* 20:2384. doi: 10.3390/s20082384
- López, A. M., León, E., and Quintero Montoya, O. L. (2022). Editorial: Biometric monitoring of emotions and behaviors. *Front. Neurobot.* 16:1079761. doi: 10.3389/fnbot.2022.1079761
- Lopez-Martinez, J., Checa, P., Soto-Hidalgo, J. M., Triguero, I., and Fernández, A. (2024). “A wearable eye-tracking approach for early autism detection with machine learning: unravelling challenges and opportunities,” in *2024 International Joint Conference on Neural Networks (IJCNN)*, 1–8. doi: 10.1109/IJCNN60899.2024.10650749
- Malezieux, M., Klein, A. S., and Gogolla, N. (2023). Neural circuits for emotion. *Annu. Rev. Neurosci.* 46, 211–231. doi: 10.1146/annurev-neuro-111020-103314
- McStay, A. (2020). Emotional ai, soft biometrics and the surveillance of emotional life: an unusual consensus on privacy. *Big Data Soc.* 7:2053951720904386. doi: 10.1177/2053951720904386
- Mukhiddinov, M., Djuraev, O., Akhmedov, F., Mukhamadiyev, A., and Cho, J. (2023). Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people. *Sensors* 23:1080. doi: 10.3390/s23031080
- Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714
- Schmitz-Hübsch, A., Becker, R., and Wirzberger, M. (2024). Emotion-performance relationship in safety-critical human-machine systems. *Comput. Hum. Behav. Rep.* 13:100364. doi: 10.1016/j.chbr.2023.100364
- Shackman, A. J., and Wager, T. D. (2019). The emotional brain: fundamental questions and strategies for future research. *Neurosci. Lett.* 693, 68–74. doi: 10.1016/j.neulet.2018.10.012
- Solis-Arrazola, M. A., Sanchez-Yañez, R. E., Garcia-Capulin, C. H., and Rostro-Gonzalez, H. (2024). Enhancing image-based facial expression recognition through muscle activation-based facial feature extraction. *Comput. Vis. Image Underst.* 240:103927. doi: 10.1016/j.cviu.2024.103927
- Tarnowski, P., Kołodziej, M., Majkowski, A., and Rak, R. J. (2020). Eye-tracking analysis for emotion recognition. *Comput. Intell. Neurosci.* 2020:2909267. doi: 10.1155/2020/2909267
- Vaijayanthi, S., and Arunnehr, J. (2024). Deep neural network-based emotion recognition using facial landmark features and particle swarm optimization. *Automatika* 65, 1088–1099. doi: 10.1080/00051144.2024.2343964
- Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., and Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Sci. Rep.* 11:2661. doi: 10.1038/s41598-021-81987-x
- Wahid, Z., Bari, A. S. M. H., and Gavrilova, M. (2023). Human micro-expressions in multimodal social behavioral biometrics. *Sensors* 23:8197. doi: 10.3390/s23198197
- Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., and Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PLoS ONE* 12:e0177239. doi: 10.1371/journal.pone.0177239
- Wu, Y., and Ji, Q. (2019). Facial landmark detection: a literature survey. *Int. J. Comput. Vis.* 127, 115–142. doi: 10.1007/s11263-018-1097-z