

Article

# Facial Emotion Recognition for Photo and Video Surveillance Based on Machine Learning and Visual Analytics

Oleg Kalyta <sup>1</sup>, Olexander Barmak <sup>1</sup> , Pavlo Radiuk <sup>1,\*</sup>  and Iurii Krak <sup>2,3</sup> 

<sup>1</sup> Department of Computer Science, Khmelnytskyi National University, 11 Instytut's'ka Str., 29016 Khmelnytskyi, Ukraine; oleg.kalyta@gmail.com (O.K.); alexander.barmak@gmail.com (O.B.)

<sup>2</sup> Department of Theoretical Cybernetics, Taras Shevchenko National University of Kyiv, 4d Akademika Hlushkova Ave., 03680 Kyiv, Ukraine; yuri.krak@gmail.com

<sup>3</sup> Laboratory of Communicative Information Technologies, V.M. Glushkov Institute of Cybernetics, 40 Akademika Hlushkova Ave., 03187 Kyiv, Ukraine

\* Correspondence: radiukp@khnmu.edu.ua; Tel.: +380-(097)-854-9148

**Featured Application:** Can be used in video surveillance systems for large groups of people.

**Abstract:** Modern video surveillance systems mainly rely on human operators to monitor and interpret the behavior of individuals in real time, which may lead to severe delays in responding to an emergency. Therefore, there is a need for continued research into the designing of interpretable and more transparent emotion recognition models that can effectively detect emotions in safety video surveillance systems. This study proposes a novel technique incorporating a straightforward model for detecting sudden changes in a person's emotional state using low-resolution photos and video frames from surveillance cameras. The proposed technique includes a method of the geometric interpretation of facial areas to extract features of facial expression, the method of hyperplane classification for identifying emotional states in the feature vector space, and the principles of visual analytics and "human in the loop" to obtain transparent and interpretable classifiers. The experimental testing using the developed software prototype validates the scientific claims of the proposed technique. Its implementation improves the reliability of abnormal behavior detection via facial expressions by 0.91–2.20%, depending on different emotions and environmental conditions. Moreover, it decreases the error probability in identifying sudden emotional shifts by 0.23–2.21% compared to existing counterparts. Future research will aim to improve the approach quantitatively and address the limitations discussed in this paper.

**Keywords:** emotion recognition; facial feature extraction; video surveillance; machine learning; visual analytics; hyperplane classification



**Citation:** Kalyta, O.; Barmak, O.; Radiuk, P.; Krak, I. Facial Emotion Recognition for Photo and Video Surveillance Based on Machine Learning and Visual Analytics. *Appl. Sci.* **2023**, *13*, 9890. <https://doi.org/10.3390/app13179890>

Academic Editors: Douglas O'Shaughnessy, Xavier Baró Solé, Xiaoming Zhang and Sergio Escalera

Received: 28 April 2023

Revised: 7 July 2023

Accepted: 30 August 2023

Published: 31 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Video surveillance systems are commonly used in various settings to ensure safety and security, particularly in places with large groups of people, such as public spaces, transportation systems, and workplaces. However, these systems face several challenges [1–3], including the need to monitor and respond to potential threats in real time effectively [4]. One of the main problems with traditional video surveillance systems is that they rely on human operators to monitor [5] and interpret the behavior of individuals in real time [6]. Such an outcome can be challenging, particularly in crowded environments, where it can be difficult to distinguish between normal and suspicious behavior [7]. Additionally, operators may suffer from fatigue or inattention [8], leading to missed or delayed responses [9].

Emotion recognition technologies can help address the above-mentioned challenges by automatically analyzing individuals' facial expressions [10] and body language in real-time to identify potential emotional states that may indicate a threat or suspicious behavior. Such technologies employ machine learning (ML) algorithms [11–13] to detect

changes in facial expressions, such as frowns, raised eyebrows, or narrowed eyes, which may indicate negative emotions like Anger, Fear, or Sadness [14,15]. By incorporating emotion recognition technology into video surveillance systems [16], security personnel can receive alerts or notifications when potentially suspicious behavior is detected. This can enable a more targeted response and improve the effectiveness of security measures. For example, security personnel could be alerted to potential threats before they escalate [17,18], allowing for a more rapid and effective response. Technologies with an emotion recognition approach can also help improve public safety by identifying individuals who may require additional support or intervention, such as those experiencing distress [19] or mental health issues [20]. This could be instrumental in high-stress environments like airports, train stations, or crowded public events, where individuals may be more prone to emotional distress.

Previously, some studies have investigated the use of emotion recognition techniques based on ML [21–23] in video surveillance systems for large groups of people. For instance, a study [24] explored using facial expression recognition technology to detect aggression and violence in crowded public spaces. The system used a traditional ML algorithm to analyze facial expressions and identify potential threats, such as angry or aggressive behavior. The system was tested in a simulated environment and achieved an accuracy rate of over 90%. In another study [25], researchers developed a real-time emotion recognition system that could analyze the emotional states of individuals in a crowded environment. The system used a combination of facial expression recognition and body language analysis to identify potential threats or emotional distress. It was tested in a shopping mall and achieved an accuracy rate of over 80%. Another work [26] explored the use of emotion recognition technology in video surveillance systems for workplace safety. The system used facial expression recognition and physiological signal analysis to detect potential safety hazards and alert workers in real time. The system was tested in a manufacturing facility and achieved an accuracy rate of over 90%.

On the other hand, recent research has explored the potential of implementing emotion recognition technology based on deep learning (DL) [27–29] in video surveillance systems for large groups of people. These studies have shown promising results in detecting emotions such as Anger, Fear, Joy, and Sadness from facial expressions. For example, a study published in [30] explored using DL-based emotion recognition algorithms to detect aggression and violence in crowded public spaces. The system was able to detect aggression with an accuracy of over 90% by analyzing facial expressions obtained by unmanned aerial vehicles. Similarly, another work [31] demonstrated the potential of DL-based emotion recognition in detecting depression and anxiety among college students through facial expressions. The system achieved an accuracy rate of over 80% in identifying depression and anxiety from facial expressions.

In the above-mentioned regard, the use of high-precision facial landmark extractors is increasingly becoming a central aspect of geometrical feature extraction in emotion recognition. For instance, Kazemi and Sullivan [32] proposed a method to achieve exact face alignment in just one millisecond. This approach uses an ensemble of regression trees to predict facial landmark positions, offering a significant advantage in speed and precision. Such rapid, precise facial alignment can be instrumental in real-time emotion recognition, as it facilitates rapid extraction of geometrical features that directly contribute to recognizing the emotional state. Kansizoglou et al. further built upon this approach in [33]. Their research underscores the utility of recurrent neural networks (RNNs) for emotion recognition. By effectively capturing temporal dynamics, RNNs can model long-term emotional behavior based on a sequence of facial landmarks. This approach enables more accurate and nuanced emotion recognition by accounting for temporal variations in facial expressions, enhancing the understanding of emotional states over time. Lastly, in [34], the authors presented a novel approach to mitigate the influence of identity-related attributes on emotion recognition. Their identity-invariant facial landmark frontalization technique ensures that the extracted facial landmarks are invariant to personal identities,

focusing instead on emotion-expressive facial movements. This is particularly useful in emotion recognition, as it helps to avoid potential biases and improves the system's generalizability. The analyzed studies underscore the importance of high-precision facial landmark extractors in emotion recognition. Adopting efficient face alignment techniques, long-term behavior modeling, and identity-invariant facial landmark extraction makes achieving more accurate and robust emotion recognition systems possible.

Nevertheless, despite the promising results mentioned above, such studies and other related research often lack interpretability and transparency in their findings [35,36]. It can be problematic to explain how the modern DL models arrived at their conclusions, which raises concerns around potential biases and ethical considerations raised by a novel concept of FACTS (Fairness, Accountability, Confidentiality, Transparency, and Safety) [37] in AI. Furthermore, DL models require a large amount of data to train, making it challenging to ensure the models accurately capture the diversity of emotions and expressions [38]. This can lead to potential biases and limitations [39] in the ability of these models to generalize to different situations in public spaces. Therefore, there is a need for continued research into the designing of interpretable and more transparent emotion recognition models that can effectively detect emotions in safety video surveillance systems.

Overall, the goal of this study is to improve the accuracy of identifying changes in the emotional state of a person through facial expressions by designing a novel technique to detect abnormal behavior of a group of people in a crowd via their facial expressions in systems that meet security requirements.

The conducted analysis identified and highlighted the following visual facial features that may indicate abnormal behavior of a group of people, which should be detected as soon as possible from the video stream of the video surveillance system:

- Rapid change of emotional state. In a normal conversation or interaction, people usually do not move quickly from one emotion to another. If the system detects that a person's facial expression fluctuates rapidly between Happiness, Anger, Fear, etc., this may indicate abnormal behavior.
- High-intensity facial expressions. People usually show a moderate level of emotion in public places. If one or a group of people frequently show excessive levels of emotion, such as very high levels of fear, this may indicate abnormal behavior caused by an emergency.
- Inconsistency in facial expressions. If different parts of the face express different or contradictory emotions (e.g., smiling with the mouth but not the eyes), this is often seen as a sign of insincere or abnormal emotional behavior.

We will consider the features mentioned above as potential indicators of abnormal behavior to achieve the goal of this study.

To achieve the goal of this study, we propose a novel, straightforward model, a method of geometric interpretation, and a hyperplane classification method to recognize changes in human emotional states at a fast pace from video surveillance cameras in areas with large crowds. One notable aspect of the authors' method is the utilization of visual analytics and the "human-in-the-loop" principle. These strategies facilitate generating transparent and easily understandable machine-learning solutions for the problem.

The key contributions of this research include:

- A new model for representing facial expressions of human emotional states, which, unlike analogs, stably groups and separates the main classes of emotions, which is provided using video surveillance cameras with low-resolution and can detect sudden changes in emotional state;
- A new method of geometric interpretation of facial areas, which, unlike analogs, makes it possible to transparently obtain the characteristic features of facial activity, which allows for analyzing low-resolution images and video frames with low computational complexity;
- An improved method of hyperplane classification for identifying facial expressions of emotional states, which, unlike analogs, allows for building a hyperplane of separation

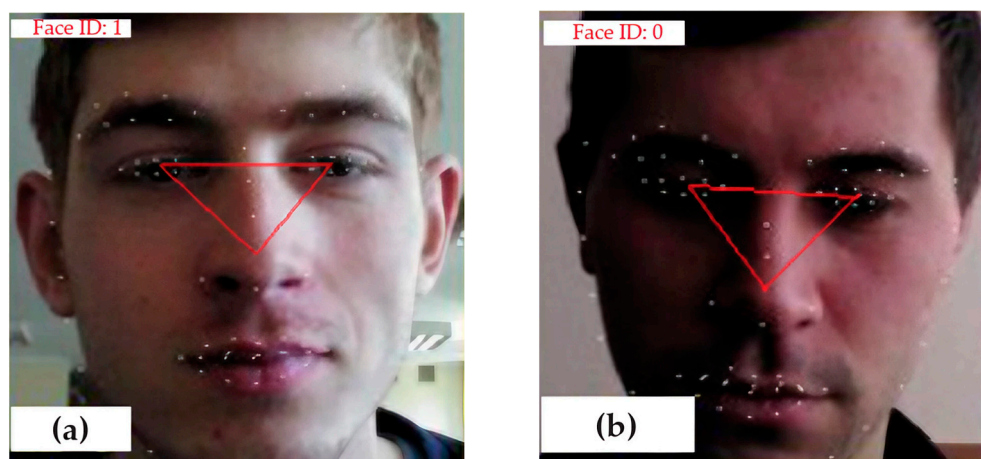
- in the vector space of features based on the “human-in-the-loop” principle, which makes it possible to obtain classifiers for detecting sudden changes in emotional states;
- A novel technique for identifying sudden changes in emotional states based on the designed model of representation of facial expressions of human emotional states, the method of the geometric interpretation of facial areas, and the method of hyperplane classification, which differs from analogs in its simple model and transparent and understandable feature selection and classification, which allows for localizing groups of people with sudden changes in emotional states based on external video recording materials, in particular through storyboarding of video frames, with relatively high accuracy.

## 2. Materials and Methods

### 2.1. Simple Model

#### 2.1.1. Analysis of Areas of the Face That Reproduce Emotional Facial Expressions

One of the ways to determine the parts of the face responsible for emotional facial expressions is the use of the Intel Real Sense video camera [40], the images of which contain automatically detected points of facial features. That is, it is possible to analyze images of people’s faces, with the primary emotional states reproduced on them and feature points selected with the help of the Intel Real Sense video camera (Figure 1).



**Figure 1.** Facial feature points are highlighted by the Intel Real Sense camera. The red triangles here in both images, (a) Face ID: 1 and (b) Face ID: 0, represent geometric shapes that reflect a visual feature from the center of the nose to the centers of both eyes in the image.

To analyze the images, we performed transformations; an example is shown in Figure 1. In step 1, the face image is normalized for comparison on one basis. To do this, the face area is centered and normalized by the distance between the eyes. In the 2nd step, informative points are selected based on the analysis of the movement of feature points in the set of input images; that is, those whose movement during emotional facial expressions is the most significant. In the final stage (3 steps), areas of the face are determined, the changes of which form the visual perception of emotion with different thresholds of movements. A set of possible states is formed for each selected area.

Figure 2 shows examples of point displacement graphs. The conclusions that can be drawn from these graphs are that the most extensive movements during the manifestation of different emotions were recorded in the points that belong to the following areas of the face (provided that only one-half of the face was considered): the top of the right eyelid, the bottom of the right eyelid, the left side of the right eyelid, the right side of the right eyebrow, the left side of the right eyebrow, the center of the right eyebrow, the right side of the nose, the right side of the lips, the center of the upper lip, and the right side of the lower lip.

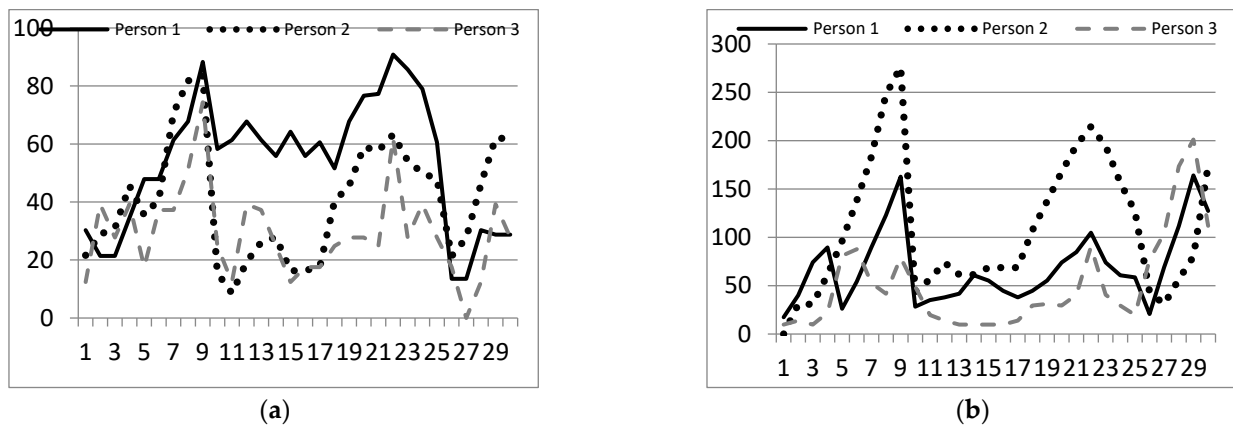


Figure 2. Shifting points for emotions (a) Anger and (b) Fear.

It can be seen from Figure 2 that some points shift more when the corresponding emotions are manifested, so it is necessary to track the shift of only the most informative points. As a result of the research and processing of the collected data, the main areas of the face were highlighted; the changes in facial expressions directly affect the display of emotional states.

Next, we determined the sets of qualitative characteristics of displacements of points or groups of points (Table 1).

Table 1. A set of values of qualitative characteristics for each part of the face. The green dots on each image represent the key landmarks of the corresponding muscle group.

Area of the Face	Qualitative Characteristics of Facial Areas					
	Unchanged	Slightly Lowered	Lowered	Slightly Raised	Raised	
Eyebrows		×				
Eyes	Upper eyelids					
	Outer corners of the eyes		×	×	×	
Lips	Corner of lips		×			
	Lower part of upper lip					
	Upper part of lower lip					

The behavior of the feature points of the face when certain emotions are manifested was investigated experimentally, and qualitative indicators for these emotions were determined. Based on the study’s results, the model’s main parameters are proposed according to qualitative criteria for evaluating the main areas of the face.

### 2.1.2. Synthesis of the Simple Model

Based on the need to detect facial expressions using ordinary cameras with a low resolution or at a long distance, and according to the results from Table 1, we introduced

the following gradation for areas of the face: (1) eyes: {open, squinted, normal}; (2) lips: {stretched, compressed, normal}; (3) eyebrows: {raised, lowered, normal}. According to the given gradation obtained in the research, the mimic expressions of emotions are presented as follows (Table 2).

**Table 2.** Representation of facial expressions of emotions with qualitative characteristics.

Areas of the Face	Anger	Fear	Joy	Sadness
eyes	squinted	flattened	normal	normal
lips	normal	normal	stretched	compressed
eyebrows	omitted	raised	raised	omitted

The presentation of facial expressions in terms of emotional states given in Table 2 will serve as a baseline for the following creation of the model according to which detection will be carried out. Empirically determined features are formally presented as follows: (1)  $x_1$ —a feature of facial expressions with eyes; (2)  $x_2$ —a feature of facial expressions with lips; (3)  $x_3$ —a feature of facial expressions with eyebrows.

$x_1, x_2, x_3 \in [0, 1]$ ; moreover,  $x_1 \in [0, 0.2]$ —for squinting eyes;  $x_1 \in [0.4, 0.6]$ —for normal eyes;  $x_1 \in [0.8, 1]$ —for open eyes;  $x_2 \in [0, 0.2]$ —for pursed lips;  $x_2 \in [0.4, 0.6]$ —for normal lips;  $x_2 \in [0.8, 1]$ —for stretched lips;  $x_3 \in [0, 0.2]$ —for lowered eyebrows;  $x_3 \in [0.4, 0.6]$ —for normal eyebrows; and  $x_3 \in [0.8, 1]$ —for raised eyebrows.

Existing unused gaps ( $[0.2, 0.4]$ ,  $[0.6, 0.8]$ ) in the proposed synthetic model serve to model a good separation between different emotional states during their classification [41].

The validity of the proposed model was checked on the synthesized data because the real input data will belong to the same intervals as the artificially created ones. The validation results are given in Section 3.3.

According to the results of the qualitative analysis of facial areas (Table 2), a set of qualitative characteristics of displacements of feature points or groups of feature points was formed (Table 3).

**Table 3.** Qualitative characteristics of areas of the human face.

	Anger	Fear	Joy	Sadness
Mouth	closed	open	open	closed
Corners of the lips	omitted	raised	raised	omitted
Eyes	open or squinted	wide open	open or squinted	squinted
Eyebrows (bridge of the nose)	erected	erected	normal	erected
Eyebrows	normal	lifted up	lifted up	normal
The corners of the eyebrows are external	normal	raised	raised	normal
The corners of the eyebrows are internal	normal	raised		normal

Given the need to identify changes in the emotional state of facial expressions using ordinary cameras with a low resolution and according to the results of Table 3, the following gradation was introduced for the features located on the parts of the face: (1) mouth: {open/closed/closed or open}; (2) corners of the lips: {lowered/raised}; (3) eyes: {wide open/open (normal)/squinted}; (4) eyebrows (bridge of the nose): {reduced to the bridge of the nose/normal}; (5) eyebrows: {raised/normal}; (6) outer corners of the eyebrows: {raised/normal}; (7) inner corners of the eyebrows: {raised/normal}.

The above presentation of facial expressions in terms of emotional states is the basis for the synthesis of the model, according to which identification will be carried out:

$$f : P \rightarrow \langle \mathbf{X}, \mathbf{W} \rangle, \tag{1}$$

where  $P$  is the matrix of pixels of the input video recording of a crowd of people,  $\mathbf{X}$  is a feature vector of facial expressions of emotions on a person’s face,  $\mathbf{X} = (x_i)_{i=1}^7$ , and  $\mathbf{W}$  is a weight vector of the model of identification of an emotional state on a person’s face.

Empirically determined features that form vector  $\mathbf{X}$  are formally presented as follows:  $x_1$  is a feature of mimicry of the facial area with the mouth;  $x_2$  is a feature of facial expressions with the corner of the mouth;  $x_3$  is a feature of facial expressions with eyes;  $x_4$  is a feature of mimicry of the part of the face with the bridge of the nose between the eyebrows;  $x_5$  is a feature of facial expressions with eyebrows;  $x_6$  is a feature of facial expressions with the outer corners of the eyebrows; and  $x_7$  is a feature of facial expressions with the inner corners of the eyebrows.

The feature vector of facial expressions in terms of emotional states  $x_i, i = \overline{1,7}$ , is described in Table 4.

**Table 4.** Qualitative characteristics of the features of facial expressions by emotional states.

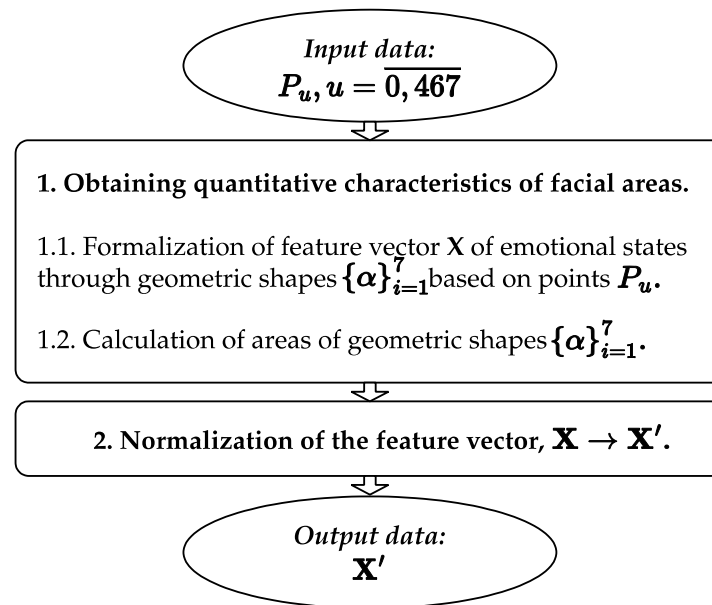
#	Quality Feature	Anger	Change Criterion	Fear	Change Criterion
$x_1$	Mouth	closed	[0;0.3]	open	[0.3;0.7]
$x_2$	Corners of the lips	raised	[0.3;0.7]	omitted	[0;0.3]
$x_3$	Eyes	squinted	[0;0.2]	wide open	[0.5;1]
$x_4$	Eyebrows (bridge of the nose)	reduced to the bridge of the nose	[0;0.3]	divorced	[0.7;1]
$x_5$	Eyebrows	omitted	[0;0.3]	lifted up	[0.7;1]
$x_6$	The corners of the eyebrows are external	omitted	[0;0.3]	lifted up	[0.6;1]
$x_7$	The corners of the eyebrows are internal	omitted	[0;0.3]	lifted up	[0.6;1]
#	Quality feature	Joy	Change criterion	Sadness	Change criterion
$x_1$	Mouth	open or closed	[0.6;1]	closed	[0;0.3]
$x_2$	Corners of the lips	raised	[0.7;1]	omitted	[0;0.3]
$x_3$	Eyes	squinted or wide open	[0.2;0.5]	squinted	[0;0.3]
$x_4$	Eyebrows (bridge of the nose)	normal	[0.3;0.7]	reduced to the bridge of the nose or normal	[0.1;0.5]
$x_5$	Eyebrows	elevated or normal	[0.3;0.7]	omitted or normal	[0.2;0.6]
$x_6$	The corners of the eyebrows are external	elevated or normal	[0.3;0.7]	omitted or normal	[0.2;0.6]
$x_7$	The corners of the eyebrows are internal	elevated or normal	[0.3;0.7]	omitted or normal	[0.2;0.6]

Mimic manifestations naturally have limited states and, in particular, are characterized by a typical set of features of external manifestations of these states. Under these principles, in [18], the limits of symptom manifestations were empirically determined. Note that there is a natural distribution, and the limits indicated in Table 3 correspond to the most typical manifestations.

### 2.2. The Method of the Geometric Interpretation of Facial Areas

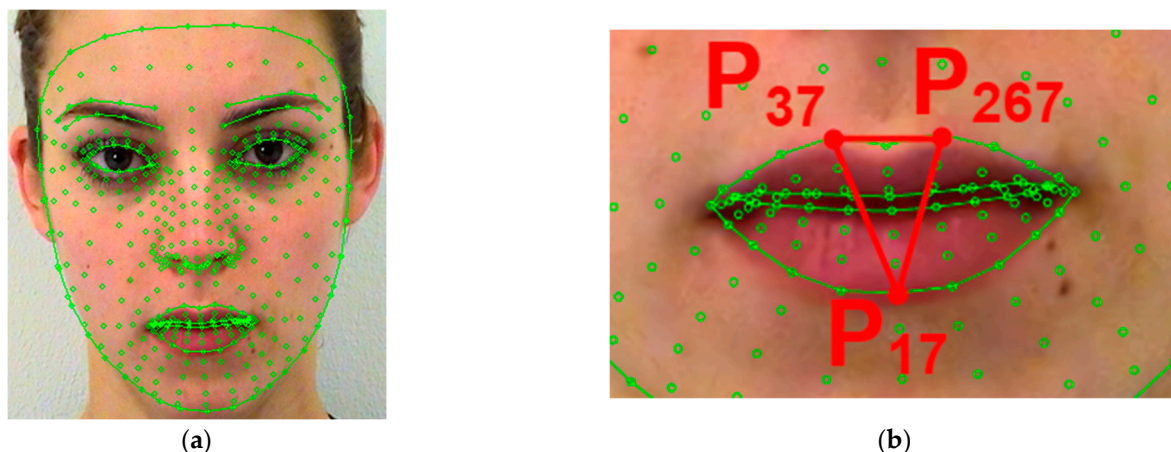
A method of the geometric interpretation of facial areas is proposed to obtain quantitative characteristics of qualitative features according to the given model. The method is used for the automated determination of facial expressions of human emotions in the form

of quantitative characteristics of geometric shapes on a person's face. The scheme of the method is shown in Figure 3.



**Figure 3.** Scheme of the method of the geometric interpretation of facial areas. Arrows here represent consistent transition from one method block to another.

The input information for the method is a vector  $P_u$  containing automatically marked landmarks on the face image. Such landmarks can be obtained in various known ways. For example, in this work, this task was completed using the open-source toolkit called MediaPipe Face Mesh [42]. Figure 4a shows an example of face geometry description using MediaPipe Face Mesh. In Figure 4a, 486 markers  $P_u, u = \overline{0,467}$ , are represented by green circles.



**Figure 4.** Human face with automatically marked landmarks: (a) all landmarks; (b) “Mouth” face area with geometric shape.

In step 1 of the method, the quantitative characteristics of the areas of the face are calculated. In step 1.1, the vector of features  $X$  from model (1) is formalized by geometric shapes  $\{\alpha\}_{i=1}^7$ , whose ends lie at points  $P_u$ . The representation of facial expressions of emotions by qualitative characteristics is presented in Table 5.



**Table 5.** Presentation of mimic expressions of emotions with geometric shapes.

#	Area of the Face	State	Feature	Type of Shape
$\alpha_1$	Mouth	Open/closed/closed or partially open	$\alpha_1 \in \{0..1\}$ , 0—closed, 1—open	A triangle describing the mouth
$\alpha_2$	Corners of the lips	Lowered/raised	$\alpha_2 \in \{0..1\}$ , 0—omitted, 1—raised	The ratio of the segments to the corners of the lips
$\alpha_3$	Eyes	Wide open/open (normal)/squinted	$\alpha_3 \in \{0..1\}$ , 0—squinted (almost closed), 1—wide open	A rectangle describing the left eye
$\alpha_4$	Eyebrows (bridge of the nose)	Reduced to the bridge of the nose/normal	$\alpha_4 \in \{0..1\}$ , 0—normal, 1—erected	A quadrangle describing the bridge of the nose
$\alpha_5$	Eyebrows	Raised up/normal	$\alpha_5 \in \{0..1\}$ , 0—normal, 1—raised	A triangle describing the upper part of the face to the eyebrows
$\alpha_6$	The corners of the eyebrows are external	Raised/normal	$\alpha_6 \in \{0..1\}$ , 0—normal, 1—raised	Section to the outer corner of the eyebrows
$\alpha_7$	The corners of the eyebrows are internal	Raised/normal	$\alpha_7 \in \{0..1\}$ , 0—normal, 1—raised	Section to the inner corner of the eyebrows

In step 1.2, the areas of geometric shapes  $\{\alpha\}_{i=1}^7$  are calculated. The type of shapes indicated in Table 5 is determined empirically for each part of the face. The distance of the segments forming the shapes  $\{\alpha\}_{i=1}^7$ , is calculated according to the Euclidian distance formula. Next, we provide a detailed description of the shapes  $\{\alpha\}_{i=1}^7$ , for each part of the face.

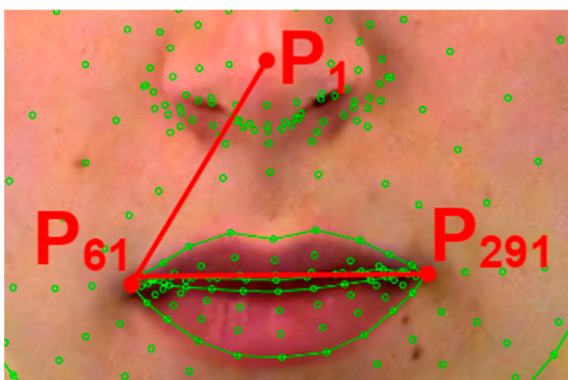
The area of the face “Mouth” is described by a triangle, the ends of which lie in points  $P_{17}$ ,  $P_{37}$ , and  $P_{267}$  (Figure 4b). The quantitative characteristic  $\alpha_1$  is area  $\Delta P_{17}P_{37}P_{267}$ :

$$\alpha_1 = \sqrt{p_0(p_0 - \overline{P_{17}P_{37}})(p_0 - \overline{P_{37}P_{267}})(p_0 - \overline{P_{267}P_{17}})}, \tag{2}$$

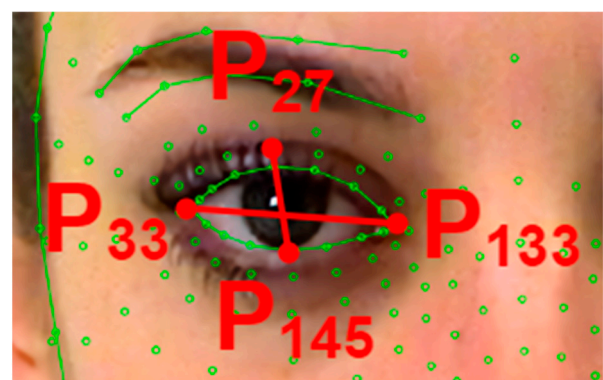
where  $p_0 = \frac{\overline{P_{17}P_{37}} + \overline{P_{37}P_{267}} + \overline{P_{267}P_{17}}}{2}$ .

The area of the face “Corners of the lips” is described by the ratio of segments  $P_1P_{61}$  and  $P_{61}P_{291}$  (Figure 5a). The quantitative characteristic  $\alpha_2$  is presented as follows:

$$\alpha_2 = \frac{\overline{P_1P_{61}}}{\overline{P_{61}P_{291}}}. \tag{3}$$



(a)



(b)

**Figure 5.** A geometric shape describing the areas of the face: (a) “Corners of the lips”; (b) “Eyes.”.

The area of the face “Eyes” is described by the ratio of segments  $P_{27}P_{145}$  and  $P_{33}P_{133}$  for the left eye (Figure 5b). The quantitative characteristic  $\alpha_3$  is the following product:

$$\alpha_3 = \overline{P_{27}P_{145}} \cdot \overline{P_{33}P_{133}}. \tag{4}$$

The area of the face “Eyebrows (bridge of the nose)” is described by the ratio of segments  $P_9P_{168}$  and  $P_{107}P_{336}$  (Figure 6a). The quantitative characteristic  $\alpha_4$  is the following product:

$$\alpha_4 = \overline{P_9P_{168}} \cdot \overline{P_{107}P_{336}}. \tag{5}$$

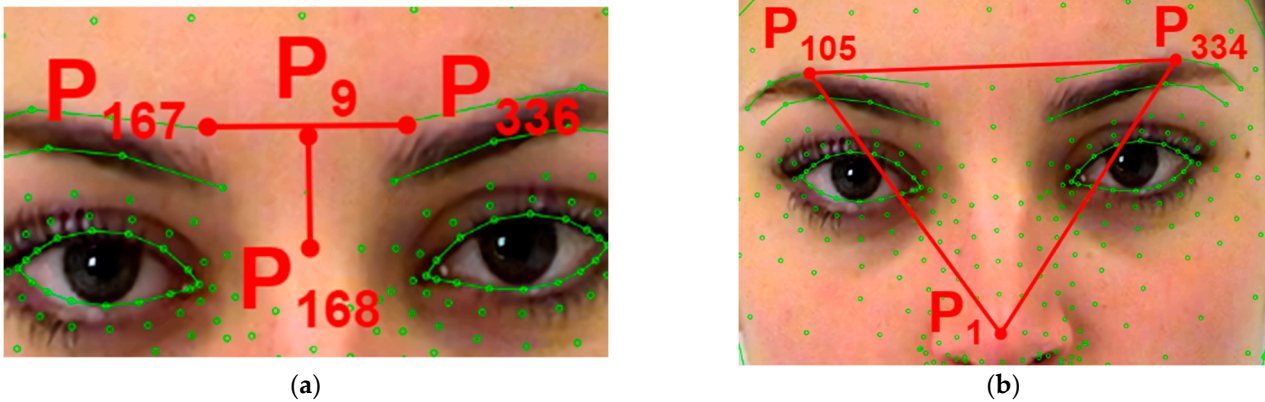


Figure 6. A geometric shape describing the parts of the face: (a) “Eyebrows (bridge of the nose)”; (b) “Eyebrows”.

The area of the face “Eyebrows” is described by a triangle, the ends of which lie in points  $P_1$ ,  $P_{105}$ , and  $P_{334}$ . (Figure 6b). The quantitative characteristic  $\alpha_5$  is the area  $\Delta P_1P_{105}P_{334}$ , calculated as follows.

$$\alpha_5 = \sqrt{p_0(p_0 - \overline{P_1P_{105}})(p_0 - \overline{P_{105}P_{334}})(p_0 - \overline{P_{334}P_1})}, \tag{6}$$

where  $p_0 = \frac{\overline{P_1P_{105}} + \overline{P_{105}P_{334}} + \overline{P_{334}P_1}}{2}$ .

The area of the face “Outer corners of the eyebrows” is described by the segment  $P_{63}P_{145}$  (Figure 7a). The quantitative characteristic  $\alpha_6$  is the length of the segment  $\overline{P_{63}P_{145}}$ .

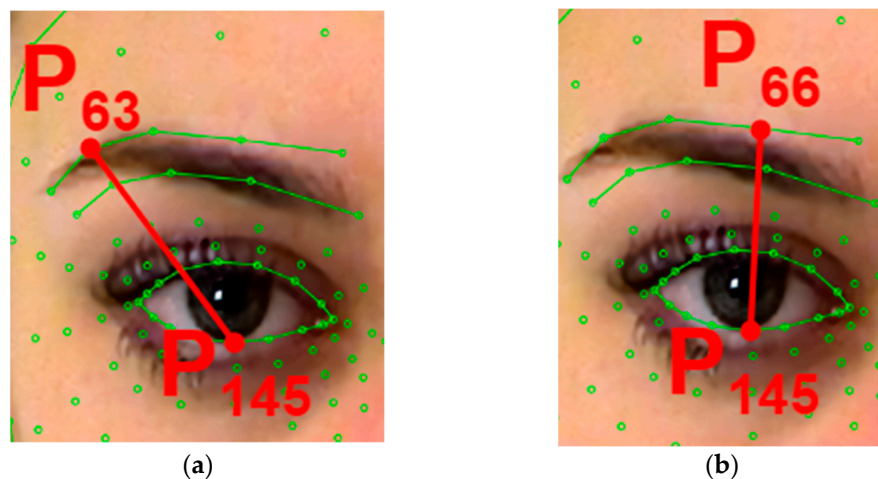


Figure 7. A geometric shape describing the area of the face: (a) “Outer corners of the eyebrows”; (b) “Corners of the inner eyebrows”.

The area of the face “Corners of the inner eyebrows” is described by the segment  $P_{66}P_{145}$  (Figure 7b). The quantitative characteristic  $\alpha_7$  is the length of the segment  $\overline{P_{66}P_{145}}$ . In step 2 of the method, the feature vector  $X$  is normalized by:

$$x_i' = \frac{\alpha_i - \alpha_{i\min}}{\alpha_{i\max} - \alpha_{i\min}}, \tag{7}$$

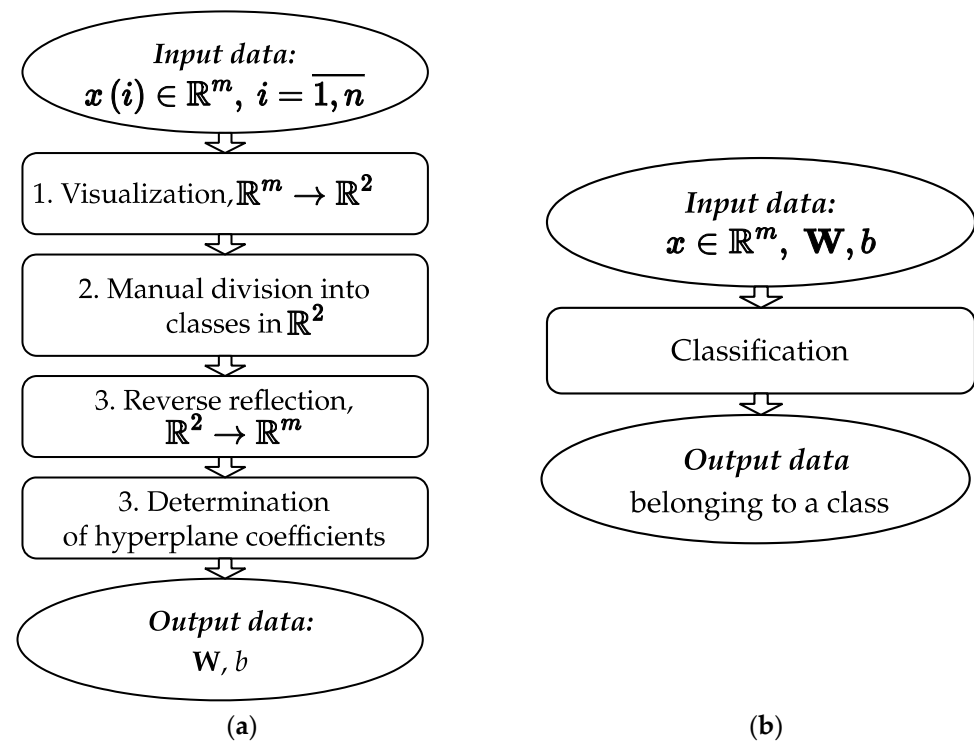
where  $\alpha_i$  is the quantitative characteristic of the  $i$ -th characteristic,  $i = \overline{1,7}$ ,  $\alpha_{i\min}$  is the minimum value of the  $i$ -th characteristic, determined empirically,  $\alpha_{i\max}$  is the maximum value of the  $i$ -th characteristic, determined empirically, and  $x_i'$  is the normalized value of the  $i$ -th characteristic,  $x_i' \in \mathbf{X}'$ ,  $x_i' \in [0; 1]$ .

The output data of the proposed method are represented through the normalized feature vector  $\mathbf{X}'$  used for further identification of emotional states.

So, the method of the geometric interpretation of facial areas makes it possible to represent a person's face, detected on a video recording, into a normalized feature vector  $\mathbf{X}'$ .

### 2.3. Hyperplane Classification Method

The method of hyperplane classification according to the above model of mimic manifestations of emotional state is intended to identify changes in a person's emotional state. The method is built according to the direction of visual analytics [43] and the "human-in-the-loop" principle [44], which distinguishes it from similar approaches by the transparency and interpretability of the decisions made according to it. That is, the method implements the principles of trust in decisions made with its help. The scheme of the method is shown in Figure 8.



**Figure 8.** Scheme of the method of hyperplane classification of the emotional state by facial expressions: (a) training; (b) classification. Arrows here represent consistent transition from one method block to another.

The input data of the method are a set of  $n$  points in the feature space ( $m$  number of features):  $x(i) \in \mathbb{R}^m$ ,  $i = \overline{1, n}$ . In step 1 of the method, input data are visualized in two-dimensional space:  $\mathbb{R}^m \rightarrow \mathbb{R}^2$ .

The mapping  $\mathbb{R}^m \rightarrow \mathbb{R}^2$  is obtained by solving an optimization problem using evolutionary methods, similar to the problem of multidimensional scaling (MDS) [45]—determining the mutual location of points (vectors) in the space of smaller dimensions. The points are located so that the pairwise distances between them in the new space differ as little as possible from the empirically measured distances in the feature space of the studied objects. The Euclidean distance is used to calculate the distance. The similarity function uses the

hypothesis that the smaller the distance between the objects, the more they are similar and vice versa. As a result, we obtain a set of points:  $x'(i) \in \mathbb{R}^2, i = \overline{1, n}$ .

In step 2, visual analytics is directly implemented in the “human-in-the-loop” direction, namely: (1) the separation ability of the training sample is visually evaluated in space  $\mathbb{R}^2$  according to the proposed vector space model; (2) human-built lines (when possible) separating grouped clouds of points (classes),  $x'(i) \in \mathbb{R}^2, i = \overline{1, n}$ ; (3) the coordinates of the beginnings and ends of these lines are memorized; (4) if it is impossible to separate the obtained classes with a single line, piecewise continuous lines are formed (the end of one line is the beginning of another); and (5) for each line, in addition to the two start and end points, there are also  $m - 2$  points located evenly between the start and end points. As a result, a set of new points is obtained:  $x'^L(i, j) \in \mathbb{R}^2, i = \overline{1, l}, j = \overline{1, m}$ , where  $l$  is the number of lines, which is the result of visual analytics.

In step 3, we carry out the inverse mapping  $\mathbb{R}^2 \rightarrow \mathbb{R}^m$  (analogous to step 1) for the sets of points:  $x'(i), x'^L(k, j) \in \mathbb{R}^2, i = \overline{1, n}, j = \overline{1, m}, k = \overline{1, l}$ , as a result of which we obtain (if possible) the set  $x^L(k, j) \in \mathbb{R}^m, j = \overline{1, m}, k = \overline{1, l}$ .

In step 4, we form systems of linear algebraic equations for the set of points  $x^L(k, j) \in \mathbb{R}^m, j = \overline{1, m}, k = \overline{1, l}$  located on the corresponding hyperplanes to determine the coefficients of these hyperplanes. There will be as many such systems as there are lines in  $\mathbb{R}^2$ . For the  $i$ -th hyperplane, the system of equations will have the form:

$$\begin{cases} w_1x_1^L(i, 1) + w_2x_2^L(i, 1) + \dots + w_mx_m^L(i, 1) + b = 0; \\ w_1x_1^L(i, 2) + w_2x_2^L(i, 2) + \dots + w_mx_m^L(i, 2) + b = 0; \\ \dots \\ w_1x_1^L(i, m) + w_2x_2^L(i, m) + \dots + w_mx_m^L(i, m) + b = 0. \end{cases} \tag{8}$$

Next, we define coefficients  $w_i$  and  $b$ . Let us present (8) in the form of the equation of a hyperplane that passes through  $m$  points:

$$\det \begin{pmatrix} x_1 & x_2 & \dots & x_m & 1 \\ x_1^L(i, 1) & x_2^L(i, 1) & \dots & x_m^L(i, 1) & 1 \\ x_1^L(i, 2) & x_2^L(i, 2) & \dots & x_m^L(i, 2) & 1 \\ \vdots & \vdots & \dots & \vdots & 1 \\ x_1^L(i, m) & x_2^L(i, m) & \dots & x_m^L(i, m) & 1 \end{pmatrix} = 0, \tag{9}$$

and expand (9) by the first line:

$$\begin{aligned} & \overbrace{(-1^{1+1}) \det \begin{pmatrix} x_2^L(i, 1) & x_3^L(i, 1) & \dots & x_m^L(i, 1) & 1 \\ x_2^L(i, 2) & x_3^L(i, 2) & \dots & x_m^L(i, 2) & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ x_2^L(i, m) & x_3^L(i, m) & \dots & x_m^L(i, m) & 1 \end{pmatrix}}^{w_1} x_1 + \\ & + \overbrace{(-1^{1+2}) \det \begin{pmatrix} x_1^L(i, 1) & x_3^L(i, 1) & \dots & x_m^L(i, 1) & 1 \\ x_1^L(i, 2) & x_3^L(i, 2) & \dots & x_m^L(i, 2) & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ x_1^L(i, m) & x_3^L(i, m) & \dots & x_m^L(i, m) & 1 \end{pmatrix}}^{w_2} x_2 + \dots + \\ & + \overbrace{(-1^{1+k}) \det \begin{pmatrix} x_1^L(i, 1) & \dots & x_{k-1}^L(i, 1) & x_{k+1}^L(i, 1) & \dots & x_m^L(i, 1) & 1 \\ x_1^L(i, 2) & \dots & x_{k-1}^L(i, 2) & x_{k+1}^L(i, 2) & \dots & x_m^L(i, 2) & 1 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \\ x_1^L(i, m) & \dots & x_{k-1}^L(i, m) & x_{k+1}^L(i, m) & \dots & x_m^L(i, m) & 1 \end{pmatrix}}^{w_k} x_k + \dots + \end{aligned} \tag{10}$$

$$\begin{aligned}
 & \overbrace{\left( -1^{1+m} \det \begin{pmatrix} x_1^L(i, 1) & x_2^L(i, 1) & \cdots & x_{m-1}^L(i, 1) & 1 \\ x_1^L(i, 2) & x_2^L(i, 2) & \cdots & x_{m-1}^L(i, 2) & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_1^L(i, m) & x_2^L(i, m) & \cdots & x_{m-1}^L(i, m) & 1 \end{pmatrix} \right)}^{w_m} x_m + \\
 & + \overbrace{\left( -1^{2+m} \det \begin{pmatrix} x_1^L(i, 1) & x_2^L(i, 1) & \cdots & x_m^L(i, 1) \\ x_1^L(i, 2) & x_2^L(i, 2) & \cdots & x_m^L(i, 2) \\ \vdots & \vdots & \cdots & \vdots \\ x_1^L(i, m) & x_2^L(i, m) & \cdots & x_m^L(i, m) \end{pmatrix} \right)}^b.
 \end{aligned}$$

The classification takes place according to the obtained vector of weighting coefficients of the hyperplane  $W$ . The linear classifier  $d(\mathbf{X}')$  is defined as follows:

$$d(\mathbf{X}^*) = \mathbf{W}^T \mathbf{X}^* + b, \tag{11}$$

where  $\mathbf{X}^* = (x_1^*, x_2^*, \dots, x_m^*, 1)^T$  is a normalized feature vector that defines the image of the classification object;  $\mathbf{W} = (w_1, w_2, \dots, w_m)^T$  is the vector of weight coefficients of the hyperplane classifier; and  $b$  is a free coefficient (scalar).

The output data of the hyperplane classification method are the results of the classifier (5). Belonging to a class is determined by the rule of relation to the classifier; that is, the object's location relative to the class line is determined.

To classify new data, their location in multidimensional space is determined by determining their position relative to the hyperplane. Substituting the data coordinates into the hyperplane equation, their location is determined from the set  $\{-1, 1\}$ . If the result is  $<0$ , then the investigated element is located "to the right" of the hyperplane, and accordingly, the person's face corresponds to the emotional state of one class; if the result  $> 0$ , the element is located "to the left" of the hyperplane and the detected face corresponds to an emotional state of another class; if the result is  $=0$ , then the element is located on the hyperplane, and the detected face corresponds to an undefined state.

#### 2.4. Technique for Detecting a Sudden Change in the Emotional State of a Crowd

As one of the options, the above model and methods are proposed to be used for the timely (in real time) detection of the atypical behavior of a group of people in a crowd. By atypical behavior, we understand the occurrence of mimic manifestations of negative emotions (for example, Fear, Anger, etc.). To achieve this, we propose analyzing streaming video from surveillance cameras installed in places of mass gatherings. Timely detection of such situations will allow the relevant security services to react, identify the source, and minimize the consequences of an atypical situation.

Next, we consider the preliminary settings, limitations, and main steps of the proposed technique.

##### 2.4.1. Presets and Limitations

Applying the geometric interpretation of facial areas means using input information in the form of a vector  $P_u$  containing automatically marked landmarks on the image of the face. Various tools can be used to obtain vector values from a face image. For the geometric interpretation method to function, the following landmarks must be present among these values:

- Two points above the upper lip and a point in the middle of the lower lip (Figure 4b).
- The points of the tip of the nose and corners of the lips (Figure 5a).
- The points of the corners of the eyes and the middle of the lower and upper eyelids (Figure 5b).

- The points of the inner tips of the eyebrows and the middle of the bridge of the nose (Figure 6a).
- The points in the middle of the eyebrows and the tip of the nose (Figure 6b).
- The points of the outer tips of the eyebrows and the middle of the lower eyelid (Figure 7a).
- The points in the middle of the eyebrows and lower eyelids (Figure 7b).

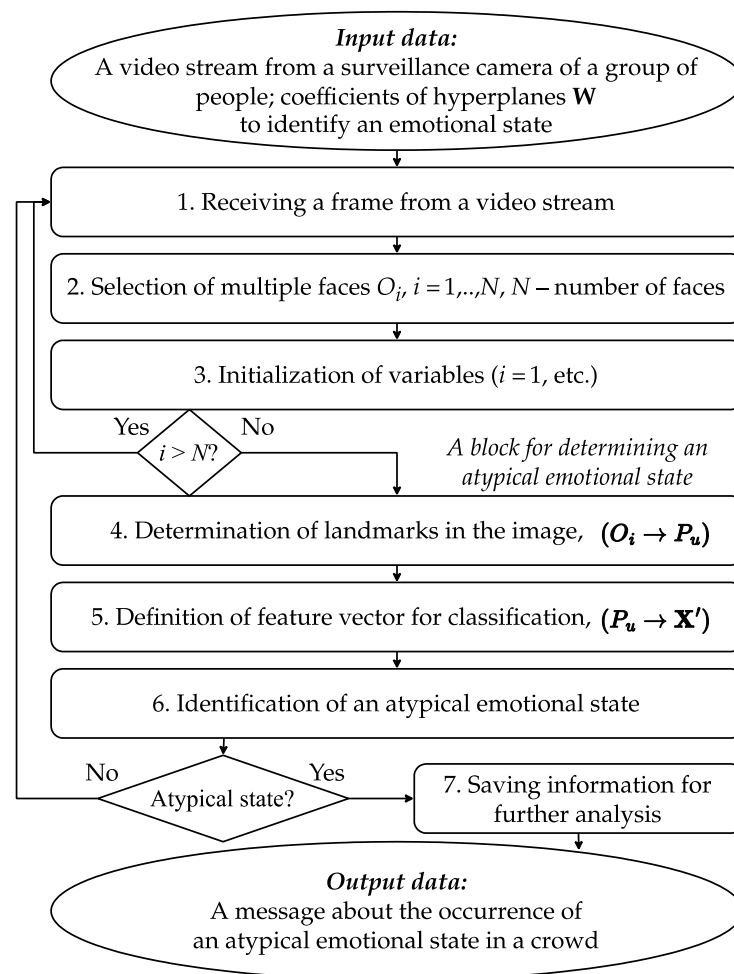
Also, for the geometric interpretation method to function, it is necessary to set the minimum and maximum values of the landmark points based on the values of the landmarks used.

For the correct hyperplane classification, a necessary condition is the use of the training set of datasets containing images of faces characteristic of the location in which the proposed technique will be used, namely:

- Corresponding categories of faces both by age and gender.
- Appropriate categories of faces according to racial and cultural characteristics.
- Considering the peculiarities of clothing.
- Considering climatic features (seasons, etc.).

#### 2.4.2. Primary Steps of the Technique

The generalized scheme of the proposed technique with its primary steps is shown in Figure 9.



**Figure 9.** The scheme of the proposed technique for detecting a sudden change in the emotional states in a crowd. Arrows here represent consistent transition from one technique block to another.

The technique's input is a video stream from a surveillance camera at a location containing a crowd of people. Also, the input information is pre-trained classifiers for identifying negative emotional states (Fear, Anger, etc.).

The frames of the video stream are sequentially processed. It is possible to consider not the entire sequence of frames but specific time intervals. In step 1, we select a frame with an image. In step 2, we select frames with faces on the frame (well-known approaches can be used for this). We obtain a set of faces for analysis. Next, in step 3, we initialize the changes required for work (counter faces, etc.).

In the logical block, we determine whether the frames with faces in the frame have not ended. If finished—go to step 1—obtain the next frame of the image; otherwise—go to the steps to determine the emotional state of the  $i$ -th face (steps 4–6).

In step 4, we define landmarks in the image. For this, a mechanism for determining specific points of the face using MediaPipe Face Mesh [42] was used. According to this method, the face is described by points in the amount of 468, which is the basis for the so-called geometric shapes of the method of geometric interpretation.

In step 5, using the proposed method of geometric interpretation, we obtain a vector of features for classification ( $P_u \rightarrow \mathbf{X}'$ ). The elements of the feature vector are the quantitative characteristics of the seven parts of the face. After that, the resulting vector is normalized to bring the feature values to the range [0; 1]. That is, the data formed by the normalized vector of facial features  $\mathbf{X}' = (x_1', x_2', \dots, x_7')$  are presented in the form of the above-considered model for recognizing facial expressions of emotions. The result of this mapping determines the normalized feature vector  $\mathbf{X}'$ , which, together with the vector of weights  $\mathbf{W}$ , forms the input data for the next step.

Next, in step 6, we identify the emotional state using the proposed hyperplane classification method. That is, the result of this step is the belonging of the input data to one of two classes,  $d(\mathbf{X}') \in \{-1, 1\}$ , where 1 corresponds, for example, to the emotional state "Target Class" and 0 to the emotional state "Not Target Class" (all other emotions).

When changes in the emotional state to the opposite are detected, data are generated for further analysis, and a corresponding message is issued to the services responsible for law and order. Therefore, the use of the above-proposed models and methods for identifying mimic expressions of emotions in the proposed technique makes it possible, based on the input data in the form of a video recording of a crowd of people, to obtain the resulting data in the form of emotional state assessments for the identification of abnormal emotional manifestations in a crowd of people.

It is worth noting that the results obtained with the help of the proposed technique are transparent and interpretable, which meets the criterion of confidence in the decisions obtained with their help.

### 3. Results and Discussion

We present the results of validation experiments and discuss the proposed models and methods.

#### 3.1. Dataset for Conducting Experiments

##### 3.1.1. Facial Expression Recognition Dataset

The Facial Expression Recognition (FER+) [46] dataset extends the original FER2013 dataset. This dataset is designed for building and testing facial emotion recognition models. The full FER+ dataset contains more than 35,000 grayscale images, each  $48 \times 48$  pixels in size, and reflects the results of a large-scale study of human emotions. FER+ differs from FER2013 because it has increased annotation reliability with multiple independent labels for each image, effectively reducing annotation noise.

In this study, a small set of FER+small images with five basic human emotional states was compiled to test the proposed IT: Anger, Fear, Joy, Neutral, and Sadness. The total number of images in FER+small is 6236. Each category of emotions in the selected dataset has a different number of samples, which reflect real-life facial expressions.

### 3.1.2. Amsterdam Dynamic Facial Expression Set

The Amsterdam Dynamic Facial Expression Set (ADFES) [47] is a comprehensive reference dataset containing images and videos of faces showing different emotional states. The ADFES is available for academic purposes and can be accessed through a request to the Amsterdam Interdisciplinary Center for Emotion.

The ADFES reference set stands out above other facial recognition datasets due to its focus on dynamic facial expressions. It contains sequences of images derived from video recordings of models that show a change in facial expression from a neutral state to a peak emotional state. This set includes 2764 videos of 22 models (12 male and 10 female) expressing emotions, with each model captured from five different angles to provide a comprehensive and detailed representation of facial expressions.

In this study, to test the proposed technique, a  $720 \times 576$  high-resolution image was extracted from the original ADFES and compiled into a small set, ADFES\_small, with five basic human emotional states: Anger, Fear, Joy, Neutral, and Sadness. The total number of images in ADFES-small is 110; all images included in ADFES-small are different from those used to determine the weights for the hyperplane classifier. Each emotion category in the selected dataset has the same number of samples—110.

## 3.2. Qualitative Analysis of the Proposed Technique

### 3.2.1. Qualitative Comparison of the Proposed Simple Model with the FACS System

Today, the facial action coding system (FACS), developed by Professor Paul Ekman [10], is the generally accepted methodology for describing changes in emotional states on the human face. FACS serves as a human-observer recognition system for detecting subtle changes in facial features. It presents the human face in the form of 98 fully controlled models, the so-called action units (AUs), each responsible for specific facial expressions. Currently, the FACS coding system's AUs are considered the benchmark for identifying different emotional facial expressions. Work [10] highlights and describes in detail four universal human emotional states, i.e., Anger, Fear, Joy, and Sadness. These emotions are manifested by different facial expressions and described by a unique set of AUs.

The FACS system characterizes the degree of expressiveness of changes in the condition of the muscles with five levels: A—minimal, B—insignificant, C—obvious, D—extreme, and E—maximum. In the proposed simple model, the change in an emotional state is determined according to a predetermined standard against the approach with active units in Ekman's work. That is, the same features are defined for different emotions but with different meanings of facial expressions.

We compared the proposed simple model of facial expressions of emotional states (1) based on the qualitative characteristics of facial features (Table 4) with the FACS human facial expression classification system. The following emotions were taken for comparison: Anger, Fear, Joy, Neutral, and Sadness. A qualitative comparison of the proposed model with the FACS system is presented in Table 6.

It is worth noting from Table 6 that the proposed model does not track the lowering of the jaw. FACS examines the sequence of different facial expressions. Sometimes these manifestations are opposite, which indicates a redundancy of features and leads to manifestations of ambiguity, as in the case of the Fear condition (raising the eyebrows, then lowering them). On the other hand, the proposed model tracks one pattern of facial expressions (seven qualitative characteristics of facial areas), which most characteristically allows for the separation of emotional states.

A comparative analysis of the FACS system and the proposed model based on identifying and selecting influential features allows for identifying changes in the emotional state of facial expressions without reducing quality performance indicators. Using the model allows the determination of the set of necessary emotional states and forms the necessary set of characteristic features that create an emotional manifestation and are characterized by the appropriate parameters for separating emotions by facial expressions.



**Table 6.** Comparison of qualitative features of the model with the system of classification of human facial expressions.

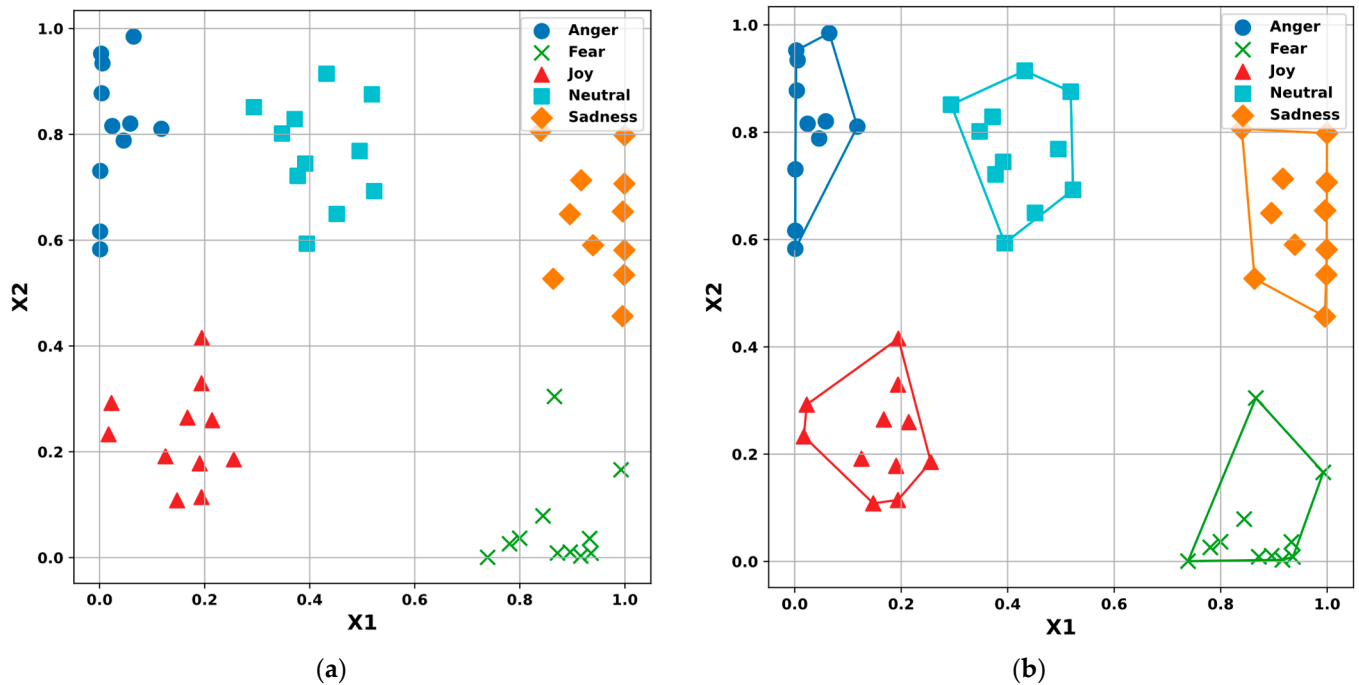
Emotional State	FACS	Synthetic Model, $X = (x_i)_{i=1}^3$	Simple Model, $X = (x_i)_{i=1}^7$
Anger	(1) lowered eyebrows (AU4); (2) raised upper eyelids (AU5); (3) raised lower eyelids (AU7); (4) compressed lips (AU23).	(1) squinting eyes (lowered upper eyelid and raised lower eyelid); (2) lowered eyebrows; (3) tight lips.	(1) the mouth is closed; (2) the corners of the lips are raised; (3) eyes are squinted; (4) eyebrows (bridge of the nose) are drawn together; (5) eyebrows are lowered; (6) outer corners of the eyebrows are lowered; (7) the corners of the eyebrows are lowered.
Fear	(1) the inner parts of the eyebrows are raised (AU1); (2) the outer parts of the eyebrows are raised (AU2); (3) lowered eyebrows (AU4); (4) raised upper eyelids (AU5); (5) raised lower eyelids (AU7); (6) lips are stretched (AU20); (7) drooping jaw (AU26).	(1) lowered upper eyelids and raised lower eyelids; (2) raised eyebrows; (3) normal lips.	(1) the mouth is ajar; (2) the corners of the lips are down; (3) eyes are wide open; (4) eyebrows (nose bridge) are brought together; (5) eyebrows are lowered; (6) outer corners of the eyebrows are lowered; (7) the corners of the eyebrows are lowered.
Joy	(1) cheeks are raised (AU6); (2) stretched corners of the lips (AU12).	(1) lowered upper eyelids and raised lower eyelids; (2) eyebrows in normal condition; (3) stretched lips.	(1) the mouth is ajar or open; (2) the corners of the lips are raised; (3) eyes are squinted or open; (4) eyebrows (nose bridge) are normal; (5) eyebrows are raised up or normal; (6) outer corners of eyebrows are raised up or normal; (7) inner eyebrow corners are raised up or normal.
Neutral	(1) the cheeks are raised (R12A); (2) the corners of the lips are turned out (R14A).	-	(1) the mouth is closed; (2) the corners of the lips are down; (3) eyes are squinted or open; (4) eyebrows (nose bridge) are normal; (5) eyebrows are normal; (6) outer corners of eyebrows are normal; (7) internal corners of eyebrows are normal.
Sadness	(1) the inner parts of the eyebrows are raised (AU1); (2) lowered eyebrows (AU4); (3) lowered corners of the lips (AU15).	(1) upper and lower eyelids are lowered; (2) eyebrows are lowered; (3) lips are compressed.	(1) the mouth is closed; (2) the corners of the lips are down; (3) eyes are squinted; (4) eyebrows (nose bridge) are reduced to the bridge of the nose or normal; (5) eyebrows are lowered or normal; (6) outer corners of the eyebrows are lowered or normal; (7) inner eyebrow corners are lowered or normal.

At the same time, the unification of the number of qualitative structural features of facial expressions has shown its effectiveness because significant features are determined, which collectively create conditions for a good separation of these groups and, therefore, separation of emotional states.

### 3.2.2. Synthesis of the Simple Model

The validity of the proposed simple model was checked on the test dataset described in Section 3.1.

Based on the images of the corresponding emotions from the given dataset, according to Table 2, features  $x_1$ ,  $x_2$ , and  $x_3$  are formed at the appropriate intervals. The generated input data were visualized for five emotional states (Anger, Fear, Joy, Neutral, and Sadness) in two-dimensional space using MDS and visualized in Figure 10.



**Figure 10.** Visualization of emotional states (a) in two-dimensional space and (b) with piecewise linear class separators.

As can be seen from Figure 10a, the synthesized data are grouped by emotions, which confirms the ability of the proposed model to be used for the classification of emotional states. Next, following the steps of the proposed hyperplane classification method (using visual analytics and the “human-in-the-loop” principle), piecewise linear separators for classes corresponding to emotional states are specified (Figure 10b), and hyperplane parameters are obtained. With the help of the obtained parameters of hyperplanes, a decision tree was built for hyperplane classification of mimic manifestations of emotional states.

Next, we provide a concise comparative analysis of two emotions based on our synthetic model. For example, as can be seen in Figure 10b, the classes of emotions Anger and Sadness can be visually divided into two groups each. In Table 7, for the Sadness emotion, this is explained by the fact that some of the respondents in the photo have squinted eyelids ( $x_1 \in [0, 0.2]$ ), while the others have normal eyelids ( $x_1 \in [0.4, 0.6]$ ).

**Table 7.** The value of the features for emotions “Anger” and “Sadness.”.

























Sadness				Anger			
Face	Eyes	Eyebrows	Lips	Face	Eyes	Eyebrows	Lips
	0.2	0.1	0.5		0.8	0.1	0.1
	0.5	0.2	0.5		0.9	0	0.2

Table 7. Cont.

Face	Sadness			Face	Anger		
	Eyes	Eyebrows	Lips		Eyes	Eyebrows	Lips
	0.2	0	0.5		0.9	0	0
	0.2	0.1	0.5		0.9	0	0.1
	0.5	0.2	0.5		0.2	0	0.1
	0.5	0.1	0.5		0.2	0	0.2
	0.5	0.2	0.5		0.1	0	0.1
	0.1	0	0.5		0.2	0	0
	0.2	0.1	0.5		0.1	0	0.2
	0.5	0.2	0.5		0.1	0	0.1
	0.5	0.2	0.5		0	0	0.1
	0.2	0.1	0.5		0.8	0.1	0.1

In Table 7, for the emotion Anger, some of the respondents in the photo have their eyelids narrowed ( $x_1 \in [0, 0.2]$ ), and others have their eyelids widened ( $x_1 \in [0.8, 1]$ ).

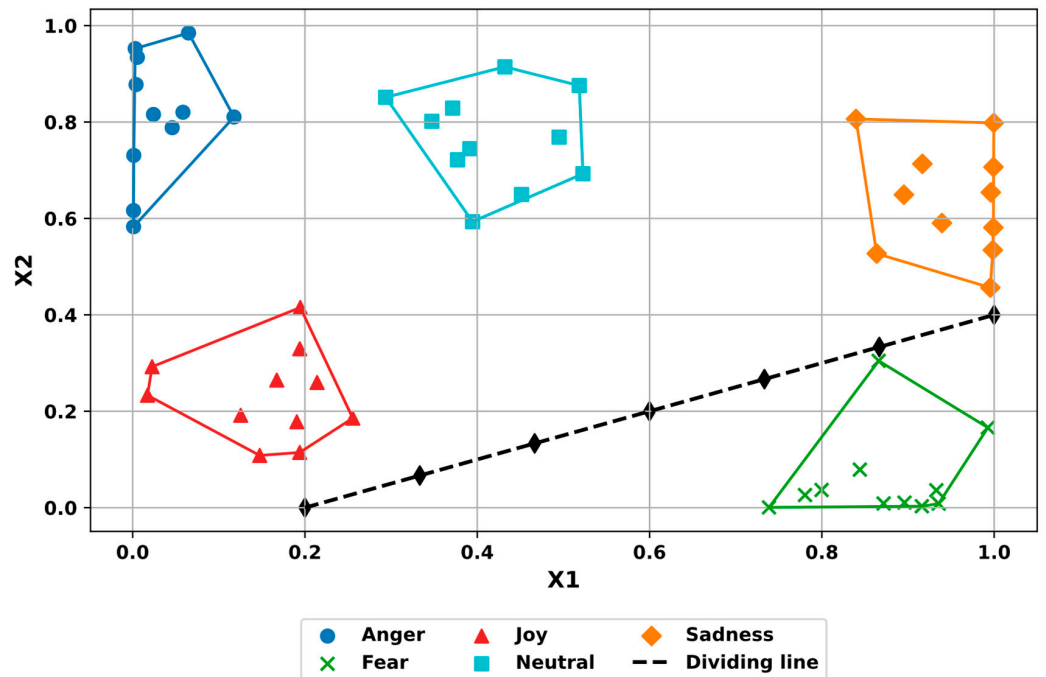
### 3.3. Validation of the Model and Methods for Determining Emotional States

First, the method of the geometric interpretation of facial expressions was applied to the training dataset. The input data of the method were 110 unique images of human faces obtained from the ADFES dataset. As a result of applying the method, the matrix of normalized values  $\mathbf{X}' = (x'_{ijk})$ ;  $i = \overline{1, 7}$ —facial features of a person;  $j = \overline{1, 110}$ —objects of a training dataset; and  $k = \overline{1, 5}$ —researched emotions was received. Preparation of weights was conducted according to the hyperplane classification of emotional state by facial expressions. The input data of the method were a matrix of normalized values  $\mathbf{X}' = (x'_{ijk})$ ,  $i = \overline{1, 7}$ ,  $j = \overline{1, 110}$ , and  $k = \overline{1, 5}$ .

The first step in applying the method of hyperplane classification was the mapping of the matrix in the visualization of normalized values  $\mathbf{X}' = (x'_{ijk}) \in \mathbb{R}^7$ ,  $i = \overline{1, 7}$ ,  $j = \overline{1, 110}$ ,  $k = \overline{1, 5}$  in two-dimensional space, i.e.,  $\mathbb{R}^7 \rightarrow \mathbb{R}^2$ . To do this, we solved the optimization

problem according to the evolutionary algorithm. To effectively find the values of  $\mathbf{W}$  weights, the ADFES training set of 110 human faces was reduced to a subset of fifty-five faces, i.e., eleven models were left, each of which depicts five emotions: Anger, Fear, Joy, Neutral, and Sadness.

As a result of solving the optimization problem, matrix  $\mathbf{X}' = (x'_{ijk}) \in \mathbb{R}^7, i = \overline{1,7}, j = \overline{1,110}$ , was mapped to matrix  $\mathbf{X}^* = (x^*_{ijk}) \in \mathbb{R}^2, i = \overline{1,2}, j = \overline{1,55}, k = \overline{1,5}$ . Reflection  $\mathbb{R}^7 \rightarrow \mathbb{R}^2$  for five emotional states, Anger, Fear, Joy, Neutral, and Sadness, is visualized in Figure 11.



**Figure 11.** An example of visualization of input data for five emotional states (Anger, Fear, Joy, Neutral, and Sadness) in two-dimensional space. The dividing line here is manually constructed as a result of reproducing the “human-in-the-loop” principle and reflects a hyperplane in two-dimensional space; the dividing line aims to distinguish the target class “Fear” from all other classes that represent four other emotions.

From Figure 11, we see that the values of matrix  $\mathbf{X}^* = (x^*_{ijk}) \in \mathbb{R}^2$ , obtained by the method of the geometric interpretation of facial expressions, were stably grouped by different emotional states. Figure 11 also shows a dividing line that was manually drawn on the coordinate plane, which clearly distinguishes the target emotional state class (in this case, the emotional state “Fear”) from all other emotion groups. This result confirms the ability of the proposed model (1) to be used to classify emotional states. Later, this dividing line was used in the hyperplane classification method to build a hyperplane in  $\mathbb{R}^7$ .

The next step of the method of hyperplane classification was inverse reflection  $\mathbb{R}^2 \rightarrow \mathbb{R}^7$ . To do this, we solved the optimization problem according to the evolutionary algorithm. Here we found matrix  $\mathbf{X}^{*L} = (x^{*L}_{1i}) \in \mathbb{R}^7, i = \overline{1,7}$ , for one dividing line built on seven points. As a result, the dividing line  $L^2 \in \mathbb{R}^2$  was mapped to hyperplane  $L^7 \in \mathbb{R}^7$  with coordinates  $\mathbf{X}^{*L}$ .

The constructed system (8) was laid out in the first line, according to block 5 of the method of hyperplane classification. So, due to applying a method of hyperplane

classification of an emotional state on facial expressions, eight weight coefficients of a separating hyperplane were received:

$$W = \begin{pmatrix} 0.005292 & 0.002342 & 0.026911 & 0.004783 \\ -0.00685 & -0.00971 & -0.03693 & 0.032351 \end{pmatrix}. \tag{12}$$

Following the obtained weighting factors (12), a linear classifier was constructed:

$$d(X') = 0.005292x'_1 + 0.002342x'_2 + 0.026911x'_3 + 0.004783x'_4 - 0.00685x'_5 - 0.00971x'_6 - 0.03693x'_7 + 0.032351. \tag{13}$$

Linear classifier (13) was later used to classify emotional manifestations and, consequently, identify emotional facial states by mimicking manifestations for information systems that meet security requirements.

### 3.4. Quantitative Comparison of the Proposed Technique with the Analogs

Experimental testing of the proposed technique was performed using two reference datasets. To ensure the purity of the experiment, small balanced datasets were selected from the reference datasets. The performance of the technique was evaluated by the statistical indicators of Accuracy, Precision, Recall, and F1-Score. Four more ML and DL models were used for the comparison with the proposed technique.

1. Model\_1: EmoFan [48] is a model proposed in a research paper. It classifies 20 different emotional categories using a combination of convolutional neural networks (CNN)s and closed recurrent units.
2. Model\_2: OpenFace [10,49] is an open-source package of tools and models for ML/DL developed at Carnegie Mellon University that enables users to recognize facial expressions and categorize a person’s emotional state.
3. Model\_3: DeepFace [50] is a software framework that contains implementations of various ML/DL models and is designed to process photo and video materials with human faces, in particular, to identify emotional states by facial expressions.
4. Model\_4: LHC-Net [51] is a new model based on CNNs and an integrated module of self-awareness designed to identify emotional states by facial expressions.
5. Model\_5: The proposed technique (see Figure 9) with the constructed classifier (13) for identifying changes in a person’s emotional state based on facial expressions for photo and video surveillance.

The results of the experimental tests on five models (Model\_1, Model\_2, Model\_3, Model\_4, and Model\_5), which were evaluated by the ability to classify five emotions (Anger, Fear, Joy, Neutral, and Sadness) according to the selected statistical indicators using the FER+small reference set, are presented in the form of discrepancy matrices in Figure 12.

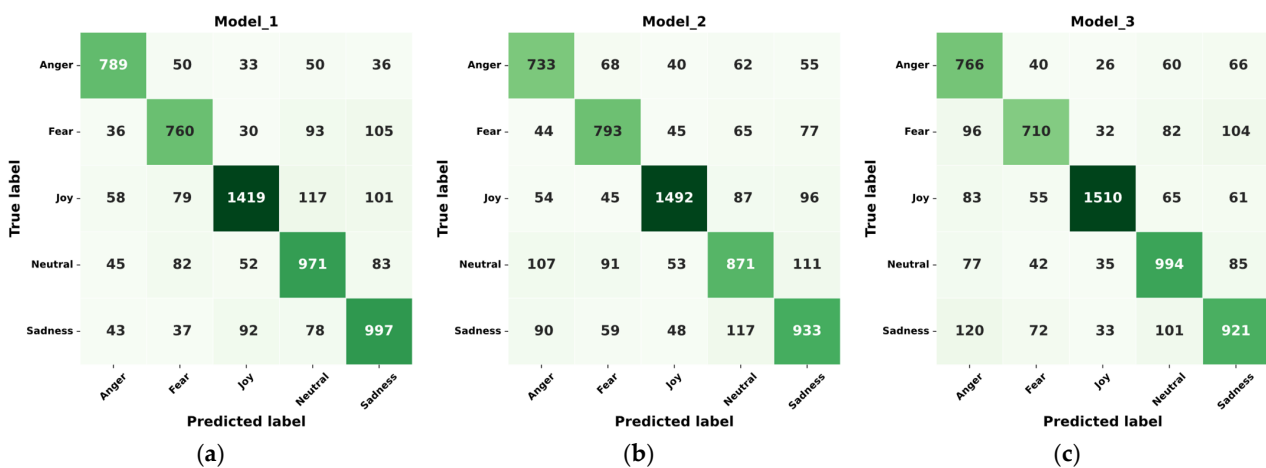


Figure 12. Cont.

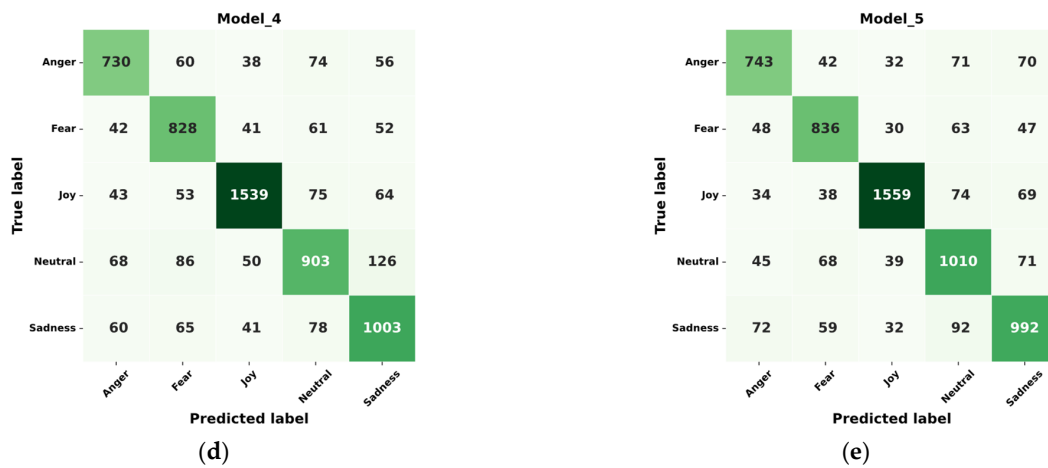


Figure 12. Confusion matrices obtained by (a) Model\_1, (b) Model\_2, (c) Model\_3, (d) Model\_4, and (e) Model\_5 based on the FER+small dataset.

A visual comparison of the models by statistical indicators for each emotional state on the FER+small dataset is shown in Figure 13.

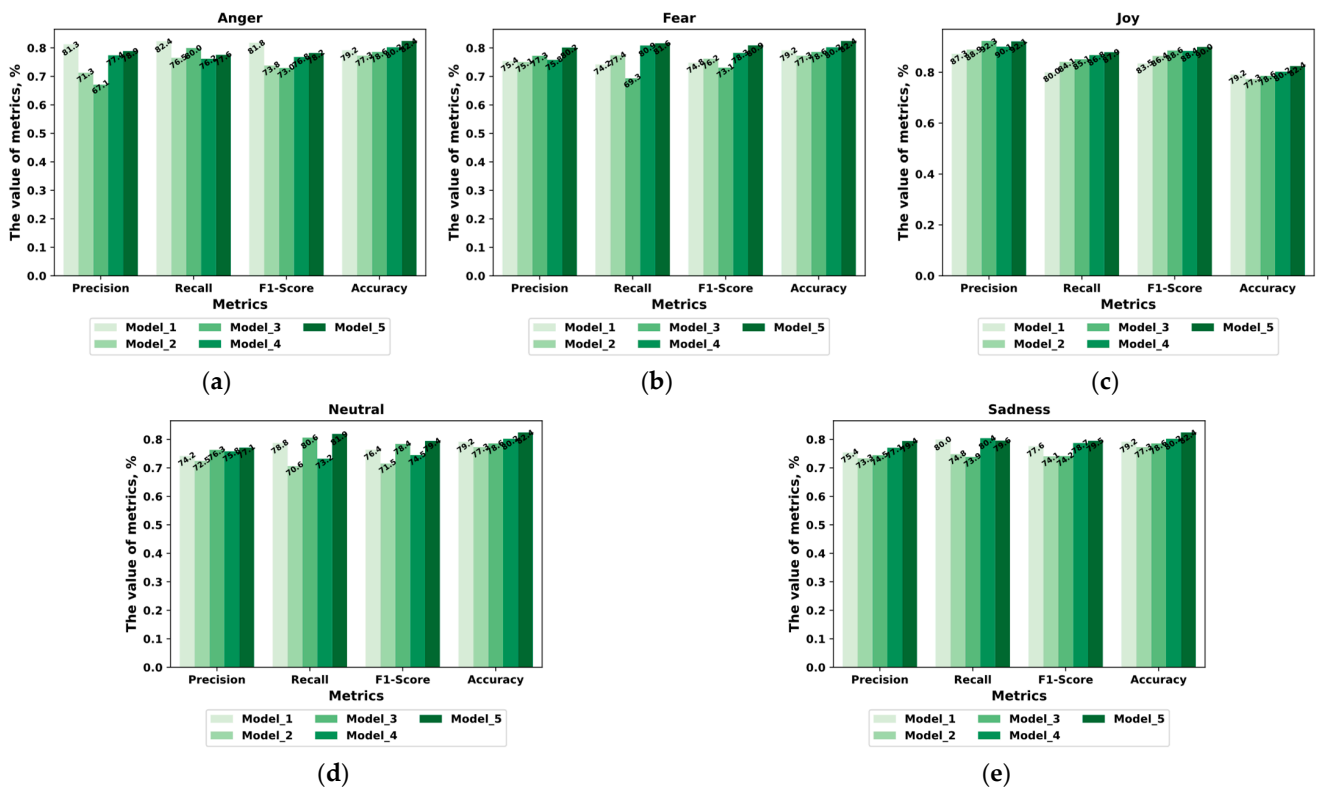


Figure 13. Comparison of models by Precision, Recall, F1-score, and Accuracy for emotions: (a) Anger, (b) Fear, (c) Joy, (d) Neutral, and (e) Sadness on the FER+small dataset.

The results of experimental testing of the five models (Model\_1, Model\_2, Model\_3, Model\_4, and Model\_5), which were evaluated by their ability to classify five emotions (Anger, Fear, Joy, Neutral, and Sadness) according to the selected statistical indicators from the ADFES-small reference set, are presented in the form of discrepancy matrices in Figure 14.

A visual comparison of the models by statistical indicators for each emotional state on the ADFES-small dataset is shown in Figure 15.

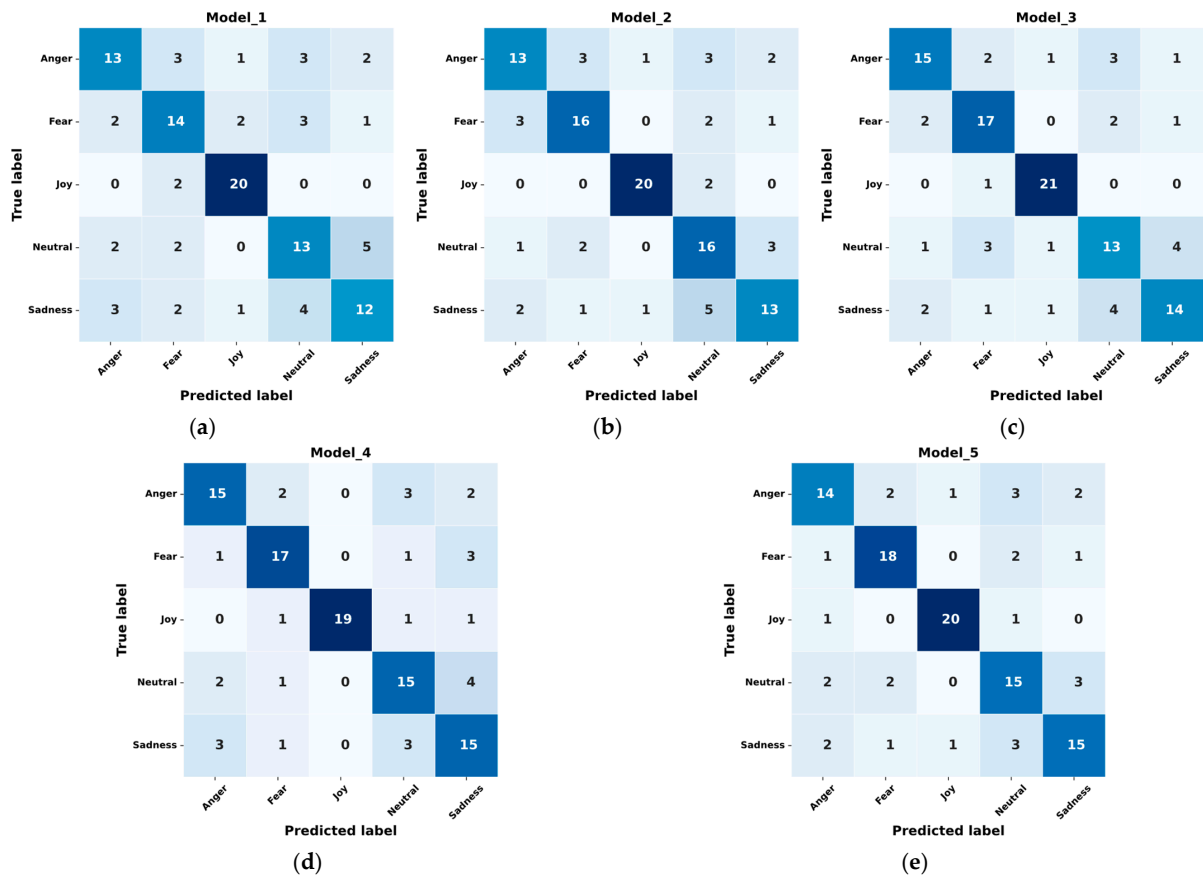


Figure 14. Confusion matrices obtained by (a) Model\_1, (b) Model\_2, (c) Model\_3, (d) Model\_4, and (e) Model\_5 on the ADFES-small dataset.

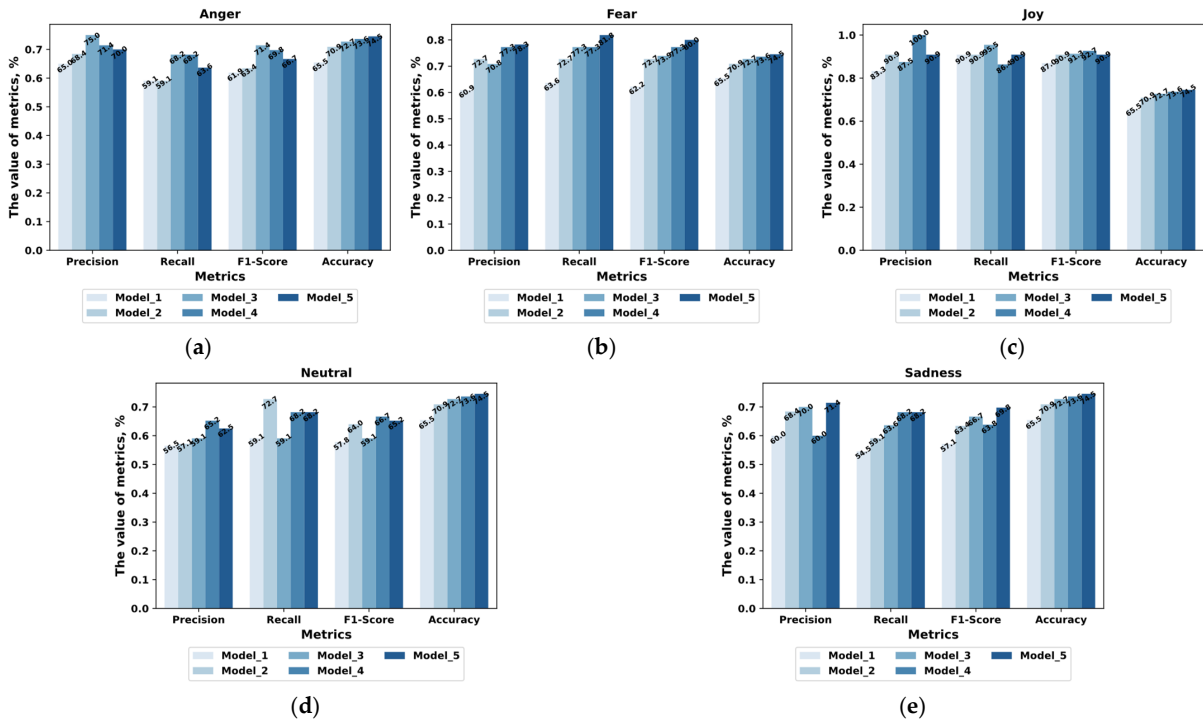


Figure 15. Comparison of models by Precision, Recall, F1-score, and Accuracy for emotions: (a) Anger, (b) Fear, (c) Joy, (d) Neutral, and (e) Sadness on the ADFES-small dataset.

It can be observed from Figures 12–15 that Model\_5 demonstrates the best results, while maintaining a high level of performance for all statistical indicators and demonstrating overall accuracy for all emotional states. The high Precision and Recall values indicate that Model\_5 can obtain reliable results in the task of classifying true positive cases and limit the number of false positives and false negatives. In particular, Model\_5 perfectly recognizes the emotion Fear with Precision—80.15% and 78.26%, Recall—81.64% and 81.82%, and F1-Score—80.89% and 80.00%, which demonstrates the ability of Model\_5 to correctly identify this emotional state by facial expressions with a high level of reliability.

The differences in the values of the indicators between Model\_5 and its analogs are also revealed. Comparatively, Model\_5 outperforms all other models in terms of Accuracy (82.42% and 74.55%), indicating its better overall ability to identify emotional states. In addition, Model\_5 demonstrated high Precision, Recall, and F1-Score values for all emotions, which emphasizes its balanced performance. High Precision indicates that the model has fewer false positives, and high Recall indicates fewer false negatives. In general, a high F1-Score also indicates that Model\_5 maintains a balance between Precision and Recall, suggesting that it is successfully applicable to emotion recognition in security-critical systems.

The key aspect that positively distinguishes Model\_5 for the task at hand is its ability to successfully process the emotion Fear. Stable operation of Model\_5 with this emotion can be especially valuable in video surveillance systems that meet security requirements, where accurate identification of the emotional state Fear can allow for the swift and efficient localization of a group of targeted individuals in a crowd.

Evaluation of the effectiveness of the proposed technique (Model\_5) in comparison with analogs was performed by pairwise comparison of the results of the technique with modern identification models that achieved high performance on both datasets. According to Figures 13–15, we can see that the proposed technique (Model\_5) outperforms the state-of-the-art identification models in the following indicators:

- Accuracy—by 2.20% and 0.91% (hereinafter, the first value corresponds to the FER+small dataset, the second to the ADFES-small dataset).
- Precision—by 2.30% for the FER+small dataset.
- Recall—by 2.21% and 0.91%.
- F1-Score—by 2.26% and 0.47%.
- False positive rate—by 0.55% and 0.23%.
- False negative rate—by 2.21% and 0.91%.

The practical implications of these results lie in the technique's (Model\_5) automation of input conversion. The technique turns low-resolution video frames of human faces ( $704 \times 480$ ) into output data identifying groups exhibiting sudden negative emotional shifts. The unique combination of a new facial expression representation model, geometric interpretation method, and enhanced hyperplane classification strategy resulted in a high classification accuracy of human emotional states (up to 82.42%). Using the technique gives security personnel a reliable tool to understand crowd dynamics and predict potential security risks during large gatherings.

In conclusion, the proposed technique showed decent and balanced characteristics in all statistical indicators. The technique shows excellent accuracy and recall, providing a low number of false positives and negative results. Its overall accuracy across all emotional states is superior to its peers, and its stable performance across all emotions, especially the Fear emotion, makes it suitable for deployment in video surveillance systems that meet security requirements. The strengths of the proposed techniques include ease of use, high accuracy of emotional state identification, and a high level of interpretability of classification results.

### 3.5. Limitations of the Proposed Technique

It is worth citing certain limitations for the use of the proposed technique. The authors see the following main limitations of the proposed technique.



1. To be able to apply the method of the geometric interpretation of facial areas, the following landmarks must be present when using third-party tools to automatically detect feature points (markers) on a face image:

- Two points above the upper lip;
- A point in the middle of the lower lip;
- Points of the tip of the nose and corners of the lips;
- Points of the corners of the eyes and the middle of the lower and upper eyelids;
- Points on the inner tips of the eyebrows and the middle of the bridge of the nose;
- Points of the middle of the eyebrows and the tip of the nose;
- Points of the outer tips of the eyebrows and the middle of the lower eyelid;
- Points of the middle of the eyebrows and lower eyelids.

The requirement for the mandatory presence of the above points is not critical since the study found that such applications generally provide such a possibility.

In addition, for the geometric interpretation method to function, it is necessary to set the minimum and maximum values of the landmarks used, taking into account the values of the landmarks. Obtaining the minimum and maximum values of the landmarks for the datasets to be used is also not a critical limitation—calculating the required values is a simple technical task.

2. The limitation of the hyperplane classification method is the need to use datasets with faces of people that are typical for the location where the proposed technique will be used, namely:

- Relevant categories of faces by age and by gender;
- Racial and cultural characteristics;
- Consideration of clothing and climatic features (time of year, etc.).

This restriction is more significant than the previous one, yet not critical. One can use existing datasets that satisfy the above limitations or create a unique dataset and test it using the visual analytics technique presented in this study for further application.

3. Emotional states are dynamic and often evolve, which is an aspect our current model does not fully account for. In our technique, we analyze each frame of the video individually, assigning an emotional state based on the facial expression in that frame. This process can overlook the temporal continuity of emotional states in the video sequences, potentially leading to inaccuracies when rapid emotional shifts occur.

However, there are additional approaches that emphasize the importance of temporal context in emotion recognition. For example, an active learning paradigm [52] can consider the temporal sequences in video streams, thereby improving emotion recognition accuracy. As such, a potential direction for future work could be integrating an active learning paradigm into our technique. This could involve designing a model that accounts for the temporal context of emotional states through techniques like RNNs or other sequence modeling approaches. Such a modification could enhance the model's ability to understand the fluid nature of emotions, improving its overall accuracy in emotion recognition from video streams.

As discussed, the above limitations of the proposed technique are not critical and have straightforward steps to eliminate them. In this regard, the research findings confirm that the proposed technique allows for localizing groups of people with a sharp change in emotional state based on the external photo and video recording materials with high accuracy.

#### 4. Conclusions

This study proposes a novel method for detecting sudden shifts in emotional states. This method utilizes a unique model to interpret facial expressions related to emotional states, a geometric interpretation technique for facial areas, and a hyperplane classification method. Distinguished by its simplicity and transparency in feature extraction and classification, the proposed technique enables the accurate localization of groups exhibiting

sudden emotional changes. These groups are identified from the external photo and video surveillance system, enhancing the technique's reliability. Our contribution is designed to accurately detect changes in emotional states through facial expressions, facilitating identifying abnormally behaving groups within a crowd—crucial for outdoor surveillance systems adhering to security protocols.

Experimental testing confirmed the efficacy of this new technique. The developed method surpassed its counterparts by improving classification accuracy by 2.20% and 0.91% on the FER+small and ADFES-small datasets, respectively, precision by 2.30% on the FER+small dataset, completeness by 2.21% and 0.91% on the FER+small and ADFES-small datasets, and F1-score by 2.26% and 0.47% for the FER+small and ADFES-small datasets. It also decreased false positive and false negative rates by 0.55% and 0.23%, and 2.21% and 0.91% for the FER+small and ADFES-small datasets, respectively.

In conclusion, the experimental testing using the developed software prototype validates the scientific claims of the proposed technique. Its implementation improves the reliability of abnormal behavior detection via facial expressions by 0.91–2.20%, depending on different emotions and environmental conditions. Moreover, it decreases the error probability in identifying sudden emotional shifts by 0.23–2.21% compared to existing counterparts.

Future research will aim to improve the approach quantitatively and address the limitations discussed in the paper. Moreover, our further investigations will include a leave-one-speaker-group-out cross-validation strategy to evaluate the technique's ability to generalize across diverse conditions during video surveillance.

**Author Contributions:** Conceptualization, O.B.; methodology, O.B. and P.R.; software, O.K.; validation, O.K. and P.R.; formal analysis, O.K.; investigation, O.K.; resources, P.R.; data curation, O.K.; writing—original draft preparation, O.K.; writing—review and editing, P.R.; visualization, P.R.; supervision, I.K. and O.B.; project administration, I.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Education and Science of Ukraine, state grant registration number 0121U112025, project title “Development of information technology for making human-controlled critical and safety decisions based on mental-formal models of machine learning.” This publication reflects the views of the authors only, and the Ministry of Education and Science of Ukraine cannot be held responsible for any use of the information contained therein.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The experimental data used to support the findings of this study are available from the corresponding author upon request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ADFES	Amsterdam Dynamic Facial Expression Set
AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
FACTS	Fairness, Accountability, Confidentiality, Transparency, and Safety
FACS	Facial Action Coding System
FER	Facial Expression Recognition
MDS	Multidimensional Scaling
ML	Machine Learning
RNN	Recurrent Neural Network

## References

1. Dilshad, N.; Hwang, J.; Song, J.; Sung, N. Applications and challenges in video surveillance via drone: A brief survey. In Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 21–23 October 2020; IEEE Inc.: New York, NY, USA, 2020; pp. 728–732. [\[CrossRef\]](#)
2. Elharrouss, O.; Almaadeed, N.; Al-Maadeed, S. A review of video surveillance systems. *J. Vis. Commun. Image Represent.* **2021**, *77*, 103116. [\[CrossRef\]](#)
3. Zhang, T.; Aftab, W.; Mihaylova, L.; Langran-Wheeler, C.; Rigby, S.; Fletcher, D.; Maddock, S.; Bosworth, G. Recent advances in video analytics for rail network surveillance for security, Trespass and Suicide Prevention—A Survey. *Sensors* **2022**, *22*, 4324. [\[CrossRef\]](#)
4. Gunduz, M.Z.; Das, R. Cyber-security on smart grid: Threats and potential solutions. *Comput. Netw.* **2020**, *169*, 107094. [\[CrossRef\]](#)
5. Sreenu, G.; Durai, M.A.S. Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *J. Big Data* **2019**, *6*, 48. [\[CrossRef\]](#)
6. Omarov, B.; Narynov, S.; Zhumanov, Z.; Kumar, A.; Khassanova, M. State-of-the-art violence detection techniques in video surveillance security systems: A Systematic Review. *PeerJ Comput. Sci.* **2022**, *8*, e920. [\[CrossRef\]](#)
7. De Gaspari, F.; Hitaj, D.; Pagnotta, G.; De Carli, L.; Mancini, L.V. Evading behavioral classifiers: A comprehensive analysis on evading ransomware detection techniques. *Neural Comput. Appl.* **2022**, *34*, 12077–12096. [\[CrossRef\]](#)
8. Drews, F.A.; Rogers, W.P.; Talebi, E.; Lee, S. The experience and management of fatigue: A study of mine haulage operators. *Min. Metall. Explor.* **2020**, *37*, 1837–1846. [\[CrossRef\]](#)
9. Park, J.; Park, J.; Shin, D.; Choi, Y. A BCI based alerting system for attention recovery of UAV Operators. *Sensors* **2021**, *21*, 2447. [\[CrossRef\]](#)
10. Ekman, P.; Friesen, W.V.; Hager, J.C. *The Facial Action Coding System: The Manual*; UT Research Nexus eBook: Salt Lake City, UT, USA, 2002.
11. Mehta, D.; Siddiqui, M.F.H.; Javaid, A.Y. Recognition of emotion intensities using machine learning algorithms: A comparative study. *Sensors* **2019**, *19*, 1897. [\[CrossRef\]](#)
12. Murugappan, M.; Mutawa, A. Facial geometric feature extraction based emotional expression classification using machine learning algorithms. *PLoS ONE* **2021**, *16*, e0247131. [\[CrossRef\]](#)
13. Saxena, A.; Khanna, A.; Gupta, D. Emotion recognition and detection methods: A comprehensive survey. *J. Artif. Intell. Syst.* **2020**, *2*, 53–79. [\[CrossRef\]](#)
14. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188. [\[CrossRef\]](#)
15. Kortli, Y.; Jridi, M.; Al Falou, A.; Atri, M. Face recognition systems: A survey. *Sensors* **2020**, *20*, 342. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Patrikar, D.R.; Parate, M.R. Anomaly detection using edge computing in video surveillance system: Review. *Int. J. Multimed. Inf. Retr.* **2022**, *11*, 85–110. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Danquah, P. Security operations center: A framework for automated triage, containment and escalation. *J. Inf. Secur.* **2020**, *11*, 225–240. [\[CrossRef\]](#)
18. Kalyta, O.; Krak, I.; Barmak, O.; Wojcik, W.; Radiuk, P. Method of facial geometric feature representation for information security systems. In Proceedings of the 3rd International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS-2022), Khmelnytskyi, Ukraine, 23–25 March 2022; Hovorushchenko, T., Savenko, O., Popov, P., Lysenko, S., Eds.; CEUR-WS: Aachen, Germany, 2022; Volume 3156, pp. 319–328.
19. Walambe, R.; Nayak, P.; Bhardwaj, A.; Kotecha, K. Employing multimodal machine learning for stress detection. *J. Healthc. Eng.* **2021**, *2021*, e9356452. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Giannakakis, G.; Grigoriadis, D.; Giannakaki, K.; Simantiraki, O.; Roniotis, A.; Tsiknakis, M. Review on psychological stress detection using biosignals. *IEEE Trans. Affect. Comput.* **2022**, *13*, 440–460. [\[CrossRef\]](#)
21. Juhong, A.; Pintavirooj, C. Face recognition based on facial landmark detection. In Proceedings of the 2017 10th Biomedical Engineering International Conference (BMEiCON-2017), Hokkaido, Japan, 31 August–2 September 2017; IEEE Inc.: New York, NY, USA, 2017; Volume 10, pp. 1–4. [\[CrossRef\]](#)
22. Zhang, J.; Yin, Z.; Chen, P.; Nichele, S. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Inf. Fusion.* **2020**, *59*, 103–126. [\[CrossRef\]](#)
23. Khan, A.R. Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges. *Information* **2022**, *13*, 268. [\[CrossRef\]](#)
24. Philpot, R.; Liebst, L.S.; Møller, K.K.; Lindegaard, M.R.; Levine, M. Capturing violence in the night-time economy: A review of established and emerging methodologies. *Aggress. Violent Behav.* **2019**, *46*, 56–65. [\[CrossRef\]](#)
25. Bera, A.; Randhavane, T.; Manocha, D. The emotionally intelligent robot: Improving socially-aware human prediction in crowded environments. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CM3K-2019), Long Beach, CA, USA, 16–20 June 2019; IEEE Inc.: New York, NY, USA, 2019; pp. 1–6.
26. Ejaz, S.; Islam, R.; Sifatullah, M.; Sarker, A. Implementation of principal component analysis on masked and non-masked face recognition. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 3–5 May 2019; 2019; pp. 1–5. [\[CrossRef\]](#)

27. Kant, K.; Shah, D.B. Emotion recognition in human face through video surveillance—A survey of state-of-the-art approaches. In Proceedings of the Information and Communication Technology for Competitive Strategies (ICTCS 2021), Rajasthan, India, 17–18 December 2021; Joshi, A., Mahmud, M., Ragel, R.G., Eds.; Springer Nature: Singapore, 2023; pp. 49–59. [\[CrossRef\]](#)
28. Tomar, A.; Kumar, S.; Pant, B. Crowd analysis in video surveillance: A review. In Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications (DASA-2022), Chiangrai, Thailand, 23–25 March 2022; IEEE Inc.: New York, NY, USA, 2022; pp. 162–168. [\[CrossRef\]](#)
29. Roemmich, K.; Schaub, F.; Andalibi, N. Emotion AI at work: Implications for workplace surveillance, emotional labor, and emotional privacy. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Hamburg, Germany, 23–28 April 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1–20. [\[CrossRef\]](#)
30. Srivastava, A.; Badal, T.; Saxena, P.; Vidyarthi, A.; Singh, R. UAV surveillance for violence detection and individual identification. *Autom. Softw. Eng.* **2022**, *29*, 28. [\[CrossRef\]](#)
31. Fan, L.; He, J.; Zheng, Y.; Nie, Y.; Chen, T.; Zhang, H. Facial micro-expression recognition impairment and its relationship with social anxiety in internet gaming disorder. *Curr. Psychol.* **2022**. [\[CrossRef\]](#)
32. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; IEEE Inc.: New York, NY, USA, 2014; pp. 1867–1874.
33. Kansizoglou, I.; Misirlis, E.; Tsintotas, K.; Gasteratos, A. Continuous emotion recognition for long-term behavior modeling through recurrent neural networks. *Technologies* **2022**, *10*, 59. [\[CrossRef\]](#)
34. Vonikakis, V.; Winkler, S. Identity-invariant facial landmark frontalization for facial expression analysis. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; IEEE Inc.: New York, NY, USA, 2020; pp. 2281–2285. [\[CrossRef\]](#)
35. Mayor-Torres, J.M.; Medina-DeVilliers, S.; Clarkson, T.; Lerner, M.D.; Riccardi, G. Evaluation of interpretability for deep learning algorithms in EEG emotion recognition: A case study in autism. *arXiv* **2023**, arXiv:2111.13208. [\[CrossRef\]](#)
36. Bethge, D.; Patsch, C.; Hallgarten, P.; Kosch, T. Interpretable Time-dependent convolutional emotion recognition with contextual data streams. In Proceedings of the Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems; CHI EA '23, Hamburg, Germany, 23–28 April 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1–9. [\[CrossRef\]](#)
37. Olteanu, A.; Garcia-Gathright, J.; de Rijke, M.; Ekstrand, M.D.; Roegiest, A.; Lipani, A.; Beutel, A.; Lucic, A.; Stoica, A.-A.; Das, A.; et al. FACTS-IR: Fairness, accountability, confidentiality, transparency, and safety in information retrieval. *SIGIR Forum* **2021**, *53*, 20–43. [\[CrossRef\]](#)
38. Umer, S.; Rout, R.K.; Pero, C.; Nappi, M. Facial expression recognition with trade-offs between data augmentation and deep learning features. *J. Ambient. Intell. Hum. Comput.* **2022**, *13*, 721–735. [\[CrossRef\]](#)
39. Wehrli, S.; Hertweck, C.; Amirian, M.; Glüge, S.; Stadelmann, T. Bias, Awareness, and ignorance in deep-learning-based face recognition. *AI Ethics* **2022**, *2*, 509–522. [\[CrossRef\]](#)
40. Zabatani, A.; Surazhsky, V.; Sperling, E.; Moshe, S.B.; Menashe, O.; Silver, D.H.; Karni, Z.; Bronstein, A.M.; Bronstein, M.M.; Kimmel, R. *Intel® RealSense™ SR300* coded light depth camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2333–2345. [\[CrossRef\]](#)
41. Barmak, O.; Krak, I.; Manziuk, E.; Lytvynenko, V.; Kalyta, O. Classification technology based on hyperplanes for visual analytics with implementations for different subject areas. In Proceedings of the 1st International Workshop on Intelligent Information Technologies & Systems of Information Security (IntellITSIS-2020), Khmelnytskyi, Ukraine, 10–12 June 2020; CEUR-WS: Aachen, Germany, 2020; Volume 2623, pp. 96–106.
42. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.-L.; Yong, M.G.; Lee, J.; et al. MediaPipe: A framework for building perception pipelines. *arXiv* **2019**, arXiv:1906.08172.
43. Krak, I.; Barmak, O.; Manziuk, E. Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology. *Comput. Intell.* **2022**, *38*, 921–946. [\[CrossRef\]](#)
44. Radiuk, P.; Kovalchuk, O.; Slobodzian, V.; Manziuk, E.; Krak, I. Human-in-the-loop approach based on MRI and ECG for healthcare diagnosis. In Proceedings of the 5th International Conference on Informatics & Data-Driven Medicine (IDDm-2022), Lyon, France, 18–20 November 2022; CEUR-WS: Aachen, Germany, 2022; Volume 3302, pp. 9–20.
45. Hout, M.C.; Papesh, M.H.; Goldinger, S.D. Multidimensional scaling. *WIREs Cogn. Sci.* **2013**, *4*, 93–103. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Barsoum, E.; Zhang, C.; Ferrer, C.C.; Zhang, Z. Training deep networks for facial expression recognition with crowd-sourced label distribution. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 279–283.
47. van der Schalk, J.; Hawk, S.T.; Fischer, A.H.; Doosje, B. Moving faces, looking places: Validation of the Amsterdam dynamic facial expression set (ADFES). *Emotion* **2011**, *11*, 907–920. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Toisoul, A.; Kossaiji, J.; Bulat, A.; Tzimiropoulos, G.; Pantic, M. Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nat. Mach. Intell.* **2021**, *3*, 42–50. [\[CrossRef\]](#)
49. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.-P. OpenFace 2.0: Facial behavior analysis toolkit. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; IEEE Inc.: New York, NY, USA, 2018; pp. 59–66. [\[CrossRef\]](#)

50. Serengil, S.I.; Ozpinar, A. HyperExtended LightFace: A facial attribute analysis framework. In Proceedings of the 2021 International Conference on Engineering and Emerging Technologies (ICEET), Istanbul, Turkey, 27–28 October 2021; IEEE Inc.: New York, NY, USA, 2018; pp. 1–4. [[CrossRef](#)]
51. Pecoraro, R.; Basile, V.; Bono, V. Local multi-head channel self-attention for facial expression recognition. *Information* **2022**, *13*, 419. [[CrossRef](#)]
52. Kansizoglou, I.; Bampis, L.; Gasteratos, A. An active learning paradigm for online audio-visual emotion recognition. *IEEE Trans. Affect. Comput.* **2022**, *13*, 756–768. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.