IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 16 March 2023, accepted 10 April 2023, date of publication 13 April 2023, date of current version 20 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3267007

RESEARCH ARTICLE

Identification of Emotions From Facial Gestures in a Teaching Environment With the Use of Machine Learning Techniques

WILLIAM EDUARDO VILLEGAS-CH[1], JOSELIN GARCÍA-ORTIZ[1], AND SANTIAGO SÁNCHEZ-VITERI[2]

[1]Escuela de Ingeniería en Ciberseguridad, FICA, Universidad de Las Américas, Quito 170125, Ecuador
[2]Departamento de Sistemas, Universidad Internacional del Ecuador, Quito 170411, Ecuador

Corresponding author: William Eduardo Villegas-Ch (william.villegas@udla.edu.ec)

This work was supported by the University of Las Américas.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Educational models currently integrate a variety of technologies and computer applications that seek to improve learning environments. With this objective, information technologies have increasingly adapted to assume the role of educational assistants that support the teacher, the students, and the areas enrolled in educational quality. One of the technologies that are gaining strength in the academic field is computer vision, which is used to monitor and identify the state of mind of students during the teaching of a subject. To do this, machine learning algorithms monitor student gestures and classify them to identify the emotions they convey in a teaching environment. These systems allow the evaluation of emotional aspects, based on two main elements, the first is the generation of an image database with the emotions generated in a learning environment such as interest, commitment, boredom, concentration, relaxation, and enthusiasm. The second is an emotion recognition system, through the recognition of facial gestures using non-invasive techniques. This work applies techniques for the recognition and processing of facial gestures and the classification of emotions focused on learning. This system helps the tutor in a modality of face-to-face education and allows him to evaluate emotional aspects and not only cognitive ones. This arises from the need to create a base of images focused on the spontaneous learning of emotions since most of the works reviewed focus on these acted-out emotions.

INDEX TERMS Computer vision, emotion recognition, neural networks, teaching.

I. INTRODUCTION

Smart classrooms aim to improve the pedagogical activities that take place in them, through the analysis of the classroom context and its adaptation using different forms of content presentation or varying teaching methodologies [1]. The classroom context refers to all the metrics that make it possible to determine what happens inside the school, for example, the type of activity that is being carried out in the classroom, and the disposition or level of attention of the students concerning the class, among others [2]. Based on this information, a smart classroom determines what needs to be improved in the classroom and recommends it to the teacher. These recommendations can range from a change in teaching practice to modifying the way content is presented, for example, using slides or augmented reality.

Currently, smart classrooms propose the use of different information and communication technologies (ICT) that contribute to the identification of the emotional state of students regarding the contents, to offer more experiential forms of teaching that reinforce memory and understanding [3]. The recognition of facial gestures in people is a complex multidisciplinary problem that has not yet been fully resolved [4]. The incorporation of new technologies

The associate editor coordinating the review of this manuscript and approving it for publication was Kah Phooi (Jasmine) Seng.

such as depth sensors or high-resolution cameras, as well as the greater processing capacity of current devices, allow the development of new technologies capable of detecting different gestures and acting in real time [5]. Facial expressions focus on the identification of gestures, which express affective states such as joy, surprise, fear, sadness, anger, and disgust. It is difficult to find databases of facial images that represent learning-focused secondary emotions, such as bored, engaged, excited, focused, interested, and relaxed [6]. The use of emotion recognition systems to detect emotions focused on teaching can help tutors to assess emotional aspects and not only cognitive processes [7].

In several of the works reviewed, the existing advances in the field of information technologies (IT) are mentioned, which have revealed new forms of support for education strategies in the higher technical-professional field. Especially in the evaluation of emotional aspects of people with the application of different biometric techniques, with emphasis on the recognition of facial patterns, which allows for capturing relevant information for the analysis and development of strategies in the educational field. In the work carried out by [8], it is mentioned that the recognition of biometric patterns is a method for the identification of individuals through facial characteristics. This analysis can be developed practically by various tools and applied to various disciplines, where it is required to use the recognition of faces and their emotions.

Another group of works mentions that facial recognition can be a three-dimensional object that is subject to different degrees of luminosity, positioning, expressions, and other factors that need to be identified from patterns generated from the acquired images [9]. Another of the works reviewed mentions that most emotion recognition systems analyze the voice, as well as all the words that are pronounced or written. For example, a high, shaky, rushed tone can indicate fear, and more complex systems also analyze gestures and even consider the environment, along with facial expressions. Emotion recognition systems typically learn to link emotion to its outward manifestation from large, categorized data sets. Gartner estimates that in the immediate future, one in ten devices will have emotion recognition technology.

In works such as [10], a study is carried out on existing technological solutions on the market that use artificial intelligence techniques to identify students' emotions during a learning activity. These solutions use facial and voice recognition, and text analysis techniques to identify students' emotions in real-time. For example, Affectiva is a platform that uses facial recognition technology to measure students' emotions as they interact with digital learning content. Emotiva is a voice analysis system that can detect emotions in real-time in students' speech. Mursion is a simulation platform that uses artificial intelligence technology to simulate interactions between students and teachers, allowing teachers to measure students' emotions in real-time and adjust their teaching accordingly [11]. Smart Sparrow uses machine learning algorithms to measure students' attention and emotion as they interact with adaptive learning content. Knewton is an adaptive learning platform that uses data analytics technology to monitor student progress and adjust learning content based on their needs and emotions. BrainCo uses wearable sensors to measure students' brain activity and detect their levels of concentration and emotion during learning. Classcraft is a gamification platform that uses data analytics technology to measure students' motivation and engagement while playing online educational games. Importantly, the selection of a platform should be based on the specific needs of each learning scenario and the ethics and privacy of student data.

While these solutions offer significant benefits, some issues are important to be aware of the main one is precision, since, this can vary depending on the quality of the input data and the complexity of the algorithms used. It is important to note that these solutions are not foolproof and may make mistakes when identifying students' emotions. The collection and use of student emotional data may raise ethical and privacy issues. It is important to ensure that students are informed about the use of their data and that appropriate steps are taken to protect their privacy. The algorithms used in these solutions may have unintended biases, which can lead to inaccurate or discriminatory results. It is important to consider the diversity of the student population and ensure that solutions are fair and equitable. Some of these solutions can be expensive and may not be available to all educational institutions. It is important to carefully weigh the costs and benefits before investing in a solution.

This work proposes the design of an emotion identification system, based on the recognition of gestures in the faces of students during the teaching process in a specific subject [12]. The gestures on the faces are generally aligned to the emotion of a person, it is this factor that is taken advantage of by an artificial intelligence algorithm that uses a neural network that has been previously trained with the use of a data set that contains a large volume of frames generated, through streaming videos, which allow classifying gestures and determining the emotion generated by a student [13]. The proposed algorithm is developed in Python and uses several libraries that are responsible for managing neural networks, as well as the use of functions for data analysis.

The novelty of having a system that is responsible for identifying emotions lies in its ability to provide real-time feedback on the emotional state of students. This can help educators tailor their teaching and improve student learning. In addition, these types of systems can help identify long-term patterns of behavior and emotions, which can help educators develop more effective interventions to help students learn [14].

For example, if an emotion recognition system detects that a student is frustrated, the educator can step in and offer additional help to address the challenges the student is facing. Similarly, if the system detects that a student is bored or disinterested, the educator can modify her teaching to keep the student more engaged. In addition, the integration of

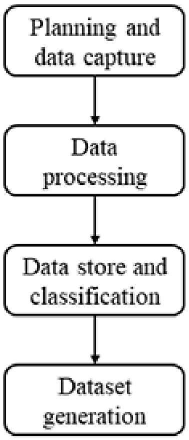

FIGURE 1. Methodology for data processing and generation of a dataset.

emotion recognition systems can provide greater objectivity to the learning assessment process since it can help educators to assess the impact of different teaching methods on the emotional state of students. In general, emotion recognition systems have the potential to improve the learning experience of students and help educators improve the quality of their teaching.

II. MATERIALS AND METHODS

For the development of the method, three fundamental bases are considered, image databases, affective computing, and emotion recognition systems with artificial intelligence. These bases guarantee the functioning of the identification of the emotions of the students, through the gestures that their faces generate in a didactic environment.

A. IMAGE DATABASE

Databases (DB) have had great growth in recent years in the analysis and construction of multidimensional data warehouses. These are created for a specific purpose and represent a collection of large volumes of data that can be text, images, audio, etc. In general, data warehouses represent a fundamental part of the infrastructure for the development of data recognition and processing applications [15], [16]. Figure 1 shows the methodology of a data warehouse, which generally integrates a data planning and capture process, processing, data storage and classification, and dataset generation. The extraction stage consists of planning and data capture; In this stage, the sources and types of data that are collected, the necessary infrastructure to acquire them, and the process that those involved will follow are specified. The transformation stage consists of data processing that allows the identification and extraction of characteristics from the captured data, then maximum and time difference techniques are applied. The last stage consists of storing and classifying the information using a title or label that allows the information to be recognized, through a loading process [17].

Planning and data capture is the process by which an image of the real world is obtained through a sensor (camera, scanner, etc.) that will then be processed and manipulated. An image is a two-dimensional function symbolized by f(x,y), where x and y are spatial coordinates. A continuous image can be represented by a matrix of N rows and M columns. This matrix is the digital representation of the image. For the correct treatment and study of the image in an artificial vision system, it is important to consider certain factors such as light, interference, image background, and resolution. The nature of the light, its position, and how the light reflects off the object can affect the quality of the image that you want the vision system to process. Inadequate lighting can overload the image with unnecessary information such as shadows, reflections, high contrasts, etc. which decreases the performance of the vision system. The two relevant lighting factors are light intensity and the position of the light source.

Preprocessing is the transformation of one image into another, that is, through one image a modified one is obtained, the purpose of which is to make the subsequent analysis simpler and more reliable. There are countless image preprocessing techniques, but very few satisfy the low computational cost. Within these, there are methods in the space domain and the frequency domain. In the transformation of data, the use of techniques and filters that allow for to improvement of the image is common, among these is the handling of histograms. That is to say that given the digital representation of an image by means of the arrangement of N rows by M columns, an M $\times$ N matrix is determined, in which the digital representation of bitmap will be given by the distribution function f(m,n), for $n \in [0, N-1]$, and $m \in [0, M-1]$, typically N and M are powers of 2. Another parameter to consider is the resolution of an image, this is the number of pixels that describe it, and a typical measurement is in terms of "pixels per inch" (PPI). Therefore, the quality of the representation as well as the size of the image depends on the resolution, which in turn determines the memory requirements for the graphic file to be generated.

Another important parameter in handling images is the size of the image, these are its actual dimensions in terms of width and height once printed, while the file size refers to the amount of physical memory needed to store the image information, the digitized image on any computer storage medium. Certainly, the resolution of the image strongly conditions these two concepts, since the number of pixels in the digitized image is fixed and therefore increasing the size of the image reduces the resolution and vice versa. Contrast is another factor widely used in the transformation of the data, this consists of increasing or decreasing the straight line with a slope of 45 degrees that represents the gray (with the precaution of not exceeding the limits 0-255) between input and output. The transformation corresponding to the contrast change is:

$$v_O(m,n) = (v_I(m,n) - 2^{y-1}) tan\phi + 2^{y-1} \tag{1}$$

where Y is the scale in bits, VI and VO are the input and output values, respectively valued in the pixel (m,n); and the angle ϕ corresponds to the properties of the linear transformation of contrasts, specifically the slope The data warehouse contains information relevant to the object of study, therefore, it must be tested to identify its strengths and weaknesses. Based on the experience and analysis of the tests, data is added or removed from the warehouse. This mechanism is repeated continuously to have a balanced database, which implies that the classes or labels have approximately the same number of images [18]. In addition, the data must be representative and correspond to spontaneous images, avoiding the storage of actuated images. The data must also be reliable, without errors or repetitions.

B. AFFECTIVE COMPUTING

Affective computing is an area of Artificial Intelligence (AI) that arises from the need to provide computer equipment with a certain capacity to interact with people. This task is carried out using computer vision techniques and machine learning algorithms, the objective of machine-human interaction is for the system to be capable of producing an effective response in people [19]. Affective computing is interdisciplinary and is applicable in different areas such as computer science, psychology, and cognitive aspects. In addition, it represents an important role in the development of intelligent interfaces applied to education or educational software. According to [20] and [21], affective computing is subdivided into four research areas, such as:

- The analysis and characterization of affective states that identify through natural interactions the relationships between effect and cognitive processes in learning.
- Automatic recognition of affective states by analyzing facial expressions and extracting features from linguistic expressions, posture, gaze tracking, and heart rate, among others.
- The adaptation of the systems to respond to a particular affective state of the users.
- The design of avatars that show appropriate affective states for a better interaction with the user.

Affective computing, from an interpretive point of view, requires that the concept of emotion be precisely defined since it can confuse with concepts such as affect, feeling, or motivation. With this consideration, it is established that affection is a process of social interaction between two or more people [22]. Giving affection is something that is transferred, for example, giving a gift, visiting a sick person, etc. Feelings are the mental expression of emotions; that is when the emotion is encoded in the brain and the person can identify the specific emotion they are experiencing, joy, grief, anger, loneliness, sadness, shame, etc. Motivation is a set of processes involved in the activation, direction, and persistence of behavior which allows us to cause changes in life in general. For its part, emotion is a state of mind produced by an event or memory and occurs every day in our daily lives, which plays an important role in non-verbal communication.

Concerning emotions, these are classified into two groups, primary or basic and secondary or alternative. In [23], he identified six basic emotions, anger, disgust, fear, happiness, sadness, and surprise, and shows some characteristics that appear on the person's face as shown in Figure 2.

Secondary or alternative emotions are complex emotions that appear after the primary emotions and depend on the situation and the context of the person. For example, a person who is afraid (primary emotion) may turn that into anger or rage (secondary emotion) and provoke an aggressive reaction. A model of valence and intensity dimensions is also used in this case to describe an emotion more precisely [25], [26], as shown in Figure 3.

Emotions generally generate expressions, and these are classified into internal and external expressions. Internal expressions can be signals generated by the body, such as blood pressure, sweating, and electroencephalography signals, and external expressions can be facial, the sound of the voice, body posture, body movement body language [29].

All the activities carried out by people generate emotions, according to the environment where this work is carried out, emotions are also focused on learning. These emotions are produced in students when they perform different activities, manifesting a variety of affective states in learning contexts [22]. Among the emotions aligned to learning are commitment, boredom, frustration, stress, focus, interest, relaxation, etc. According to several of the related works, the emotions of confusion, frustration, or bored are presented in students when they perform exercises that require certain techniques or information with which they are not familiar and can be considered negative for student learning.

On the [30] a model related to the emotions of learning is proposed, in which a model divided into four quadrants is shown, where quadrant I show an evaluation of admiration, satisfaction or curiosity resulting in positive constructive learning, Quadrant II depicts an appraisal of disappointment, perplexity, or confusion resulting in negative constructive learning, Quadrant III depicts an appraisal of frustration, rejection, or misconceptions that result in negative learning and the last Quadrant IV, shows an assessment of optimism, new research that translates into positive learning [21], [31], [32].

C. EMOTION RECOGNITION SYSTEM

For the design of emotion recognition systems, various methods of extraction and classification of biometric parameters are used. Parameters are extracted from user-generated gestures, for example, by using a person's facial expressions. Facial expression analysis is applied in different areas of interest such as education, video games, and telecommunications, to name a few [33]. In addition, it is one of the most used in human-computer interactions. Facial expression recognition is an intelligent system that identifies a person's face and from it obtains certain characteristics that it analyzes and processes to know the person's affective state [34]. The objective of facial recognition is, from the incoming image, to find a series of data of the same face in a set of training images in a database. The great difficulty lies in ensuring that

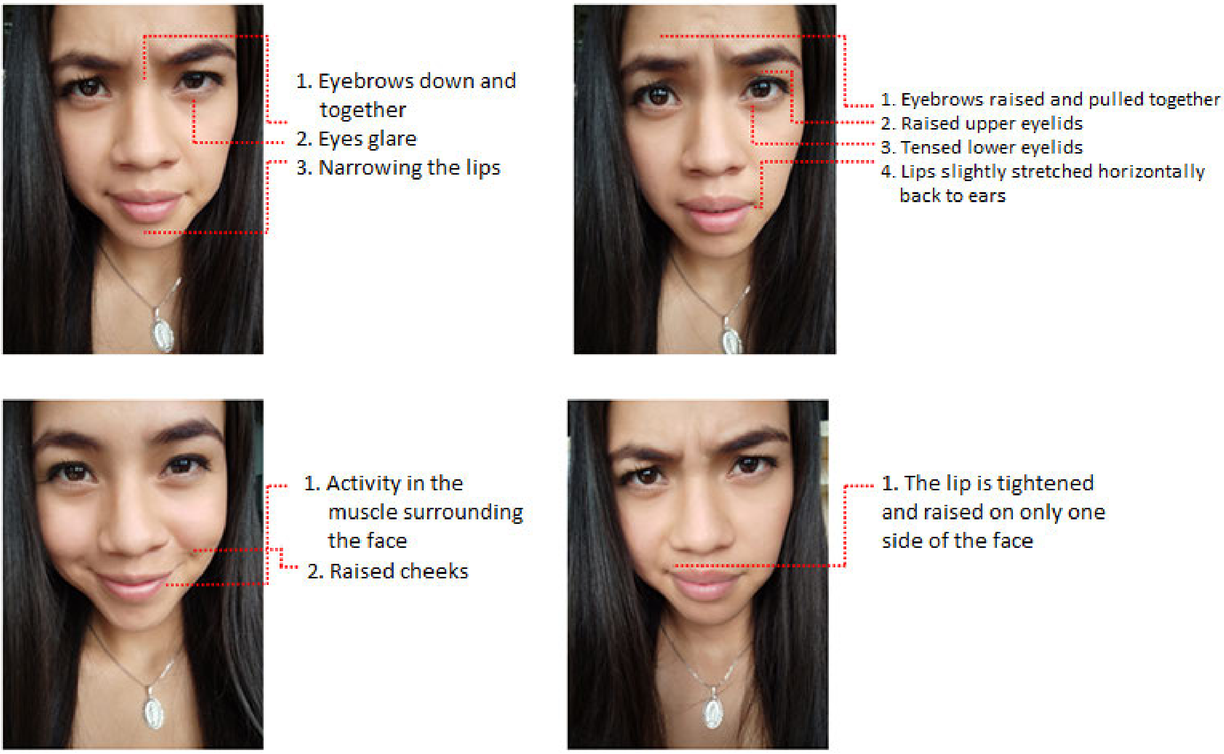

FIGURE 2. **Instances of discernable facial expressions that convey various emotions through gestures [24].**

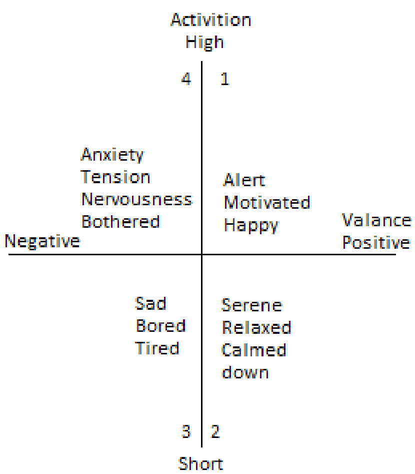

FIGURE 3. **Model of valence and intensity dimensions that describe an emotion [27], [28].**

this process is carried out in real-time, something that is not within the reach of similar works that have been previously reviewed.

The emotion recognition system through gesture classification requires meeting various technical requirements. The selection of the algorithms to use will depend on the specific use of the system, but it is considered essential to have experience in the required programming language, such as Python since it is widely used in the development of image recognition systems. In addition, there are several Python image processing libraries, such as OpenCV, Pillow, and Scikit-image, among others, which are used to read, process, and manipulate images, and it is important to analyze their use according to system requirements. Likewise, for the implementation of image recognition, experience in machine learning libraries such as TensorFlow, Keras, PyTorch, and Scikit-learn, among others, is required.

Concerning hardware, it is considered that most image recognition algorithms require a large number of hardware resources for processing, so it is important to have a computer with a good amount of RAM and a high-end processor speed, the use of a graphics card can speed up the processing of information. In addition, it is necessary to have a data set that is representative of the images to be recognized. This data is used to train the machine learning model. For the proper choice of algorithms and the selection of hyperparameters, it is necessary to know statistics and model theory.

Regarding the design of the algorithm in Python for the recognition of emotions, several technical data are taken into account, such as the programming language, where Python is used with the OpenCV, NumPy libraries. The choice of the emotion detection method through face detection and analysis of facial expressions. The algorithm uses a pre-trained Machine Learning classification model to recognize facial expressions. The model is trained on labeled data sets to recognize six universal emotions, happiness, sadness, fear, surprise, disgust, and anger. Regarding the input data, these are images or video sequences, the output data are the emotions recognized for each image or video frame. In choosing the data set, it is important to choose a quality data set that contains a variety of emotions and facial expressions. This data is used to train and test the emotion recognition model.

In addition, in the development of the algorithm, it is necessary to select the characteristics that will be used to recognize emotions. Characteristics may include the intensity of facial expressions, the position of the eyes and mouth, and the duration of the expressions. There are different machine-learning algorithms and signal-processing techniques that can be used to recognize emotions. It is important to choose the appropriate classification model for the selected data set and features. Before feeding the data to the model, it is important to perform proper preprocessing such as data normalization, denoising, and dimensionality reduction. Another important parameter is the evaluation of the emotion recognition model using a test data set to measure its accuracy and generalizability. Likewise, once the emotion recognition model has been developed, it is important to integrate it with an application so that it can be used in real life.

Regarding the evaluation, several techniques and metrics are considered to evaluate the emotion recognition algorithm. For example, we consider the precision that measures the proportion of correct classifications made by the model. It is one of the most common metrics and is used to assess the overall performance of the model. Sensitivity measures the model's ability to correctly identify a specific emotion. It is especially important if you want to recognize specific emotions, such as sadness or happiness. Specificity measures the model's ability to correctly identify the lack of a specific emotion. For example, if you want to identify the absence of negative emotions such as sadness or anger. The F value is a measure that combines precision and sensitivity in a single measure. It is useful when you want to find a balance between accuracy and the ability to correctly identify a specific emotion. The confusion matrix is a table showing the actual rankings and the model's predictions. It is a useful tool for evaluating false positive and false negative rates and for identifying patterns of error in the model.

D. METHODOLOGY

For the development of this work, an institution of higher technical education is considered. The students participating in the study are the third level, and as per the regulations of the institution, the objective of the system was explained to the population and their consent was requested to be able to capture facial images for research purposes only. The data capture is carried out in two stages, in which the first 20 students participate, of which 12 are men and 8 women, with an age range of 18 to 20 years. In the second stage, 18 students participate, of which 10 are men and 8 women, who are in the same age range. The two groups of students take the subject of databases and are part of two parallels, the first group belongs to parallel "A" and the second to "B," the teacher is the same for both parallels, therefore, the methodology is the same for the total population.

To capture the data, a controlled environment is established that corresponds to a laboratory with a maximum capacity of 24 students. An Intel RealSense depth camera has been installed in this laboratory to capture images of the participants. The laboratory has 24 computers that the students use to solve different exercises on the subject. The emotion recognition application is housed in the institution's data center, where two virtual servers have been created, one of which processes and stores the image signals, and the other captures and stores images with a specific name. The name consists of an identifier, a number, and the exact date, and time.

The class sessions are separate for the two groups of students, however, the methodology for each session is similar, as are the various activities that consist of theoretical components and practical development with exercises related to a specific problem. During the design stage of the activities, they evaluate different learning outcomes, these are based on the eight types of learning exposed in the learning theory of [34]. Within the classes, the teacher has three sessions, each of 60 minutes. As it is a three-hour class, the attention of the students can be diminished, therefore, it is the teacher's task to use a methodology that allows for improving the interest of the students. The class schedule is added to the continuous sessions since this is in the evening and 75% of the students work. These factors hinder the teacher's efforts to maintain the interest and concentration of students in the different topics of the subject. Among the techniques used by the teacher is giving higher priority to practical sessions and activities that require student participation.

The mechanics of the class that uses the recognition system has as components that each student has at their disposal a computer and a webcam, in addition to the depth camera that each laboratory has. Having the cameras allows the identification of gestures when the students are carrying out practical activities on their computers [35]. Instead, when student review is needed for conceptual review of a topic or if the teacher is the focus of the class, depth cameras have the full perspective for monitoring.

To carry out facial recognition, a database of several faces that express a facial gesture is needed. These faces should denote a variety of expressions such as happiness, sadness, boredom, surprise, etc. Another aspect that these images must have had was the variation in light conditions, whether people wear glasses or not, even if they close or wink. It is recommended that these images be collected in the environment or environment where facial recognition will be applied. The variety of images obtained from the faces contributes to the performance of the algorithms applied in the system. There is no exact number of images and no guide to precisely how many images are needed for an AI algorithm to classify emotions. The only existing reference is that the larger the volume of images and the more varied, the better to increase the precision in the algorithm.

Initially, many images are needed to train a classifier so that it can discern between the presence and non-presence of an object. For example, in the first stage, the algorithm detects if there is a face, therefore happy images are required, that is, photos with faces, and negative images, which are images that do not contain faces. Then, the features of all the images are extracted, and for this, an automatic learning approach is used, and the training proceeds as shown in figure 4.

OpenCV has several pretrained classifiers, not only for human faces but also for eyes, smiles, and animal faces, among others. To use face detection with hair cascade in OpenCV it is necessary to use the detectMultiScale module that helps to detect objects according to the classifier used. This allows obtaining a bounding rectangle where the object to be found within an image is located, for this, the following arguments must be specified:

- Image: It is the image where the face detector will act.
- ScaleFactor: This parameter specifies how much the image will be scaled down. For example, if 1.1 is entered, it means that the image will be reduced by 10%, with 1.3 it will be reduced by 30%, thus creating a pyramid of images. It is important to note that if we give a very high number, some detections are lost. While for very small values such as 1.01, it will take longer to process, since there will be more images to analyze, in addition to the fact that false positives can increase, which are detections presented as objects or faces, but which are not they are.
- MinNeighbors: This parameter specifies how many neighbors each candidate rectangle must have to retain it, that is, a window is obtained that will go through an image looking for faces. So, this parameter relates to all those delimiting rectangles of the same face. Therefore, minNeighbors specifies the minimum number of bounding boxes, or neighbors, that a face must have to be detected as such.

The detection of the students' gestures, while they are receiving a synchronous class, requires the generation of a data set that the recognition algorithm takes for its training, for this *haarcascade_frontalface* is used. This algorithm allows the storage of a specific number of faces for system training, this process can be done from a video provided by the students or from streaming. In this case, the number of images is 350 per student, and the students have been asked to simulate many expressions in the videos, where gesticulation helps to improve the accuracy of the training. For this task, a convolutional network designed in Kaggle Notebooks is used with the Keras, Tensorflow, and OpenCV libraries for image processing [36]. In addition, the Matplotlib library is used in the design to generate a graph where the model is evaluated. To do this, the loss and precision values of both the training and validation phases are identified. The confusion matrix is generated with the Sklearn library, to evaluate the accuracy of the [37] classification.

The data set generated for the first training session and tests corresponds to the first group of students that have been named as parallel "A" and six classes are evaluated, commitment, boredom, frustration, focused, interested and neutral, with a total of 7,000 images corresponding to 20 students,

TABLE 1. Dataset generated for the first training session with the detection of six emotions applied in the tests of a convolutional neural network with 7,000 images.

Commitment	Boring	Frustration	Focused	Interested	Neutral
1850 Total	1680	592	1140	922	816 7000

Table 1 indicates the number of images that each category has.

The 7,000 images correspond to the gestures identified within each of the six categories and these are processed as part of the training of the convolutional network. Figure 5 shows the architecture of the convolutional network, which has three layers, each of which has a convolution with an activation function and an additional max polling layer. At the end of these the flatten, dropout layers, and an output layer with an activation function were added.

Facial expression recognition depends on four steps shown in Figure 6, the first step is to detect faces in an image by applying the oriented gradient histogram algorithm. Next, the facial landmark estimation algorithm is used, which allows the identification of 68 landmarks on each face. In the third step, 128 measurements are created for each face through deep learning, which corresponds to the unique characteristics of the faces, finally, with the unique characteristics of each face, the name of the person is determined.

Figure 7 illustrates the steps of the system to achieve accurate identification of emotions. The first stage validates if the image received by the recognizer contains a face; if no face is found, the image is discarded. Next, a grayscale filter is applied to remove color channels and detect important parts of the face such as the nose, eyes, eyebrows, and mouth. In the next stage, facial points are marked on the detected parts, using an initial reference point in the center of the nose, and various facial points are identified in different areas of the face. Geometric calculations are then performed to determine the distance between the initial reference point and each facial point detected on the face. The result of these calculations is a matrix of facial features that are processed using a support vector algorithm together with the corresponding emotion labels so that the system can learn to classify facial expressions. Finally, the trained neural network is fed more facial feature vectors to test whether the algorithm has learned to classify gestures and recognize emotions.

For the construction of the model, it is possible to use a video previously stored in a repository, for which the face must simulate several gestures that will later be separated into frames and used for training the artificial neuron. In addition, in the process of data acquisition and training of the neuron, it is possible to acquire a greater volume of data with the use of streaming video. Obtaining the images from a streaming video has the advantage that the frames represent the most natural state possible of the people and even help to identify the variations of the environment such as lighting, objects, attenuation, etc.

For the generation of the streaming video, it is necessary to specify the faceClassif function where the face detector to

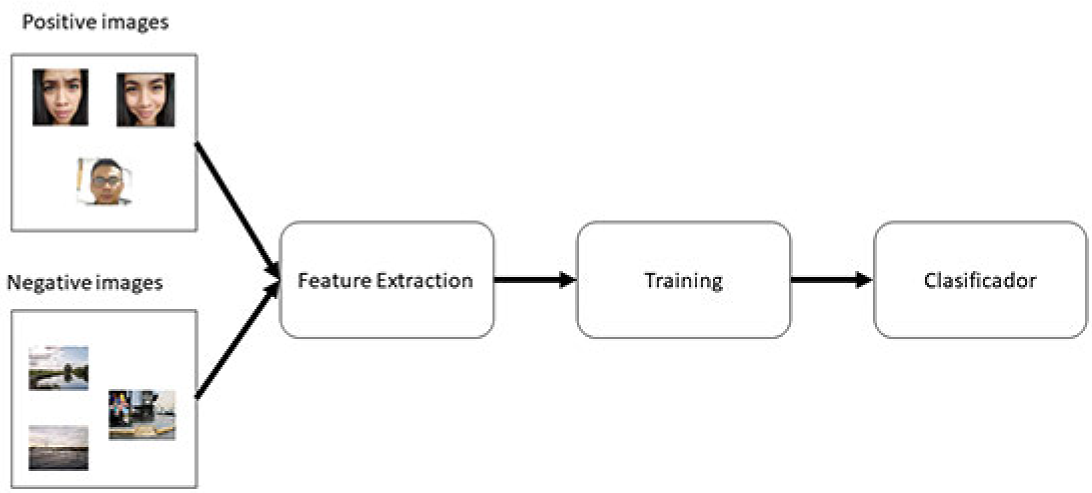

FIGURE 4. Development of a database for storing and organizing images.

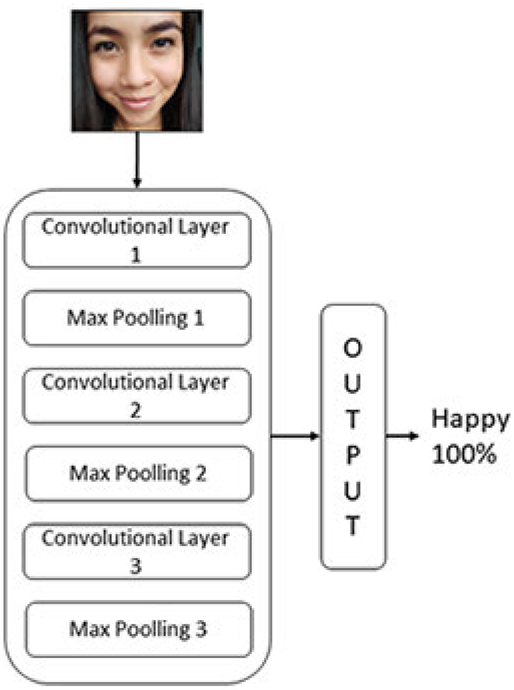

FIGURE 5. The design and structure of a convolutional neural network utilized in an emotion recognition model.

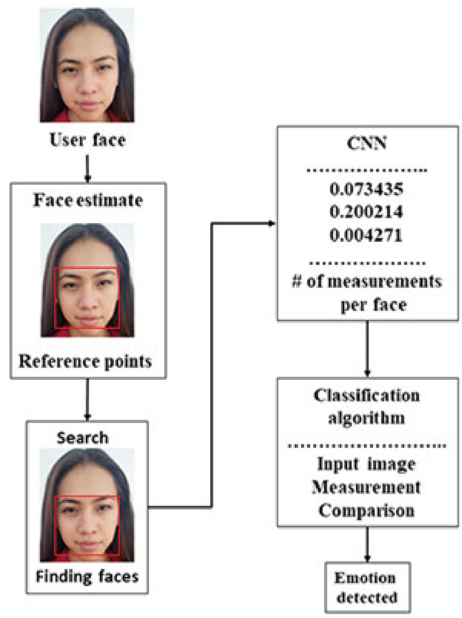

FIGURE 6. Stages of detection implemented within a convolutional neural network.

be used is assigned, in addition, a counter initialized to "0" is established, which is responsible for counting all the faces that are stored for training. The library reads each frame and resizes it to 640 pixels wide with the grayscale application. In the next step, a function is created where all the detected faces are stored. To analyze each one of them, the cycle is generated, in which a rectangle is drawn that surrounds each face and then it is cut and resized to 150 × 150 pixels to that all stored faces have the same size, and we store each face in the repository and finally increment the counter by 1.

For training, the obtModel function is created, for this function the method to be used and facesData are necessary, the array where the faces with their different emotions will be stored with the labels, which are the labels of each of the faces corresponding to each emotion. There are several options for training the neural network if the method to use is EigenFaces, then *cv2.face.EigenFaceRecognizer_create*() will be assigned to *emotion_recognizer*. If EigenFaces is used, *cv2.face.FisherFaceRecognizer_create*() will be assigned to *expression_recognizer*. Or if the method used is LBPH, *cv2.face.LBPHFaceRecognizer_create*() will be assigned to *expression_recognizer*.

III. RESULTS

In the first evaluation of the system, a dataset consisting of 54 images was used, which correspond to 18 students belonging to groups "A and B" out of a total of 38 students.

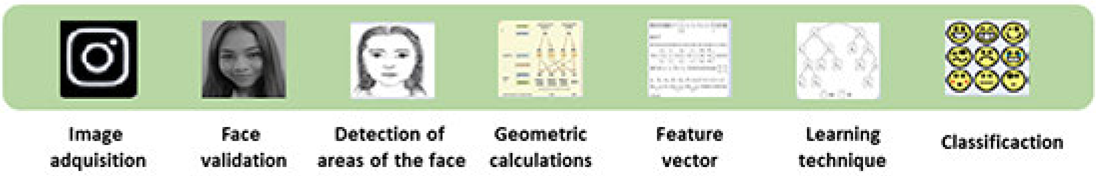

FIGURE 7. Proposed framework for recognizing gestures and identifying emotions using architectural design [24].

TABLE 2. Description of the test set for algorithm verification.

Description	Number of images	Models used
Angry	6	2,5,8,13,21,25
Contented	9	5,6,11,13,23,26,28,35,37
Disgusted	7	2,10,13,18,28,30,35
Fearful	8	3,10,16,21,25,26,37,18
Happy	4	2,6,11,37
Neutral	6	3,5,10,16,23,26
Sad	8	6,8,16,21,25,28,30,35
Surprised	6	3,8,11,18,23,30

For the dataset, the students were asked to model three images that are included in the description of Table 2. The students were randomly selected, and what is sought is to carry out a test to evaluate the operation of the emotion recognition algorithm, model subjects have been identified as, 2, 3, 5, 6, 8, 10, 11, 13, 16, 18, 21, 23, 25, 26, 28, 30, 35 and 37.

In the table, the first column indicates the description or name of the evaluated emotion, the second column contains the number of images present in the set of tests related to each emotion, and the last column "Models used" indicates the model number detected. In this training stage, it was necessary to generate the feature matrix for the test data set. It had 54 rows (number of images) and 136 columns (number of features). The matrix with the results was stored in a file with a "txt" extension. This task is performed by a routine written in the Python language.

Once the "X" matrix is obtained, the next step evaluates the 136 characteristics of each one of the images in the hypothesized functions corresponding to the 8 emotional classifiers. The result of this operation is a matrix with 8 rows and 54 columns. The number of rows corresponds to the number of classifiers (of the different emotions). In this way, row 1 will contain the outputs of the classifier associated with the emotion with code 1 (Anger), 2 (Happy), etc. The number of columns for your part is equal to the number of images in the test set. Thus, for example, in column 1 the outputs of the 8 classifiers for the first image were obtained. The final output of the algorithm is for each image the row number for which the maximum probability value was calculated. The results obtained made it possible to build the confusion matrix in Table 3.

From the matrix shown in the table, it is possible to deduce that the operation in the first evaluation of the system, presented several errors with which the operation must be adjusted, for example, in code 1 (Angry) six images were expected, but The result obtained was five hits and one image was recognized as neutral. This can be observed in several

TABLE 3. Confusion matrix built with the results obtained for the test with 54 images.

Prediction	1	2	3	4	5	6	7	8	Accuracy
1	5	1			1				0.71
2		7			1				0.87
3			7					1	0.87
4		1		7		1			0.78
5					3				1
6	1					3			0.75
7							8		1
8				1		1		5	0.4
Recall	0.83	0.77	1	0.87	0.75	0.5	1	0.83	80 %

FIGURE 8. Application of filters and identification of facial points.

of the emotions evaluated in the confusion matrix. Another result is that in the disgusted and sad emotions, the algorithm has recognized 100 % of the images.

After the training stage and the first evaluation, the operation of the system is evaluated in a production environment. In this, the system is implemented obtaining several results whose fundamental base is the identification of the gestures on the faces of the students. Figure 8 shows the result of the process carried out on the images, which consists of applying a gray filter to the figure (left-right). Subsequently, the parts of the face are detected and the initial center point is found, which will be used as a reference for the different facial points.

Once the facial points have been obtained, the pertinent geometric calculations are made from the distance between the initial reference point (center of the nose) to the other facial points and their respective position coordinates (x, y), obtaining the matrix of features. Table 4 shows the precision obtained by the emotion recognition system, resulting from the algorithm applied with a cross-validation of 10 instances for the training of the neural network. The right column shows the average accuracy of each emotion class. According to the literature review, the emotions that are present in learning are established in the first column, in addition, what is sought with this limitation in emotions is to adjust the system so that there is not a high percentage of false positives.

TABLE 4. Accuracy obtained in cross-validation in emotion recognition.

Classification	Average Accuracy
Bored	68.57 %
Hooked	71.83 %
Excited	70.99 %
Focused	75.82 %
Interested	92.07 %
Relaxed	100 %
Average accuracy	52.78 %
Standard deviation	13.70 %

TABLE 5. Result of existing correlations between emotions.

Classification	Engaged/bored	Relaxed	Excited
Interested	-0.501	-0.316	-0.225
Engaged/bored	1	0.223	0.441
Focused	0.134	-0.435	0.987

The statistical analysis process was applied to the resulting data to find anomalous results referring to a possible correlation between the gender of the participants and their emotions. For this, the test of the person carried out on the samples of the entire population is considered. In the process, each gesture was ordinarily categorized, according to the results obtained, there is no evidence of an existing relationship between the gender of the participants and their emotions. In a subsequent analysis, the bivariate correlation calculation was used to obtain the Pearson correlation coefficient, between the average of the emotions presented by each participant. Table 5 presents the results of the tests with negative correlation values. Among the most significant results, positive correlations were found in hooked/bored and excited (.441), and in focused and excited (.987), this means that when the focused value increases, the exciting value also intensifies.

For the evaluation of the system, a greater number of sessions is added, for which a second data set is considered that is generated from the second group named parallel "B". The generation of the second data set maintains the same guidelines, such as the generation of 350 images from a video or streaming, the objective is to increase the number of facial expressions for the recognition of emotions. In this session the images correspond to the students while they carry out an activity, considering the experience of the group "A". The activity proposed for the detection of mood is the reading of concepts on the application of relationships in databases, the objective is to improve the detection of boring emotion, this being the one that had the greatest impact in group "A". A total of 6300 captured images belonging to the 18 students in the group were obtained, these images are attached to the process to create an additional evaluation session of the system. Table 6 shows the total number of images in the corpus, annexing the sessions where the boredom emotion was captured.

Table 7 shows the mean precision and the standard deviation of the two sessions that were considered for the evaluation of the proposed model. According to the results obtained, it can be identified that the performance of the machine learning algorithm in the personality trait recognition task

TABLE 6. Samples taken for the generation of the image dataset.

Emotion	Session 1	Session 2
Bored	1850	890
Hooked	1680	1540
Excited	592	1175
Focused	1140	974
Interested	922	933
Relaxed	816	788

TABLE 7. Results of the precision and standard deviation taken in two sessions for the evaluation of the recognition model.

Classification	Session 1	Session 2
Average accuracy	65.7%	69.9%
Standard deviation	2.3%	1.3%

has been effective. For this, an evaluation was carried out using quintuple-crossed controls. The precision calculation is performed using the standard deviation, for which it is possible to use two formulas. One of them is used if the measured data represents an entire population, while a second formula is used in case the measured data comes from only a sample of the population. For this work, a set of samples is used, where formula (3) for the standard deviation to be used for a set of samples is presented below:

$$\sigma\sqrt{\frac{\sum(x-\mu)^2}{n}} \tag{2}$$

$$\sigma\sqrt{\frac{\sum(x-\mu)^2}{n-1}} \tag{3}$$

As with calculating the mean deviation, the first step is to find the mean of the data values using the set of measurements for each factor. In the second step, the square of each variance is calculated by subtracting the data value from the mean for each data point and squaring the result. The square of the difference will always be positive in each case of the five sample data values. Finally, the square root of the result is calculated, which represents the variance of the data set. The standard deviation is the square root of the variance, using this calculation the precision of the balance is represented by giving the mean, plus or minus the standard deviation. For example, the accuracy of AGR is 65.7%.

The results of the final validation of the method are presented in Figure 8, where it was carried out with a comparison of the emotion recognition system between the two groups participating in the study. Five people were selected, three men and two women, aged between 19 and 22 years. The participants carried out two activities; reading a scientific article and carrying out three design exercises for a relational database using Microsoft SQL. The results were obtained by recording and counting the times the recognizer matched the emotion classification with the student's feelings.

IV. DISCUSSION

Based on the literature review on various facial expression recognition systems, it was determined that for implementation the Python programming language is aligned with

TABLE 8. Comparison of the emotions captured by the system among 5 randomly selected users in the development of two learning activities.

Classification	A 1	A 2	A 3	A 4	A 5	Total
Coincidence	33	47	53	48	55	236
Total data	44	75	61	81	69	330
Average	0.75	0.626	0.868	0.592	0.797	77.66%

A= Participant

the needs exposed in this work. Furthermore, the OpenCV library for parsing and processing images or videos is suitable for training support vector machines [38]. In several works, the use of headbands with sensors for the identification of emotions is proposed, however, the process with devices that people must use can cause rejection because they are considered invasive. Therefore, the inclusion of AI and computer vision was planned in the design, for which it was determined to validate that the image received by the recognizer contains a face and that if it is not found, discard the image. Subsequently, in the process, a gray filter is applied to eliminate the different color channels to later detect some important parts of the face such as the nose, eyes, eyebrows, and mouth [39], [40]. By applying a facial point technique, it is possible to detect an initial reference point in the center of the nose and to identify various facial points on parts of the face [41].

According to the results obtained, it is established that the proposed method can be applied in an educational environment, where the ideal environment is when students have a personal computer. Having personal cameras guarantee the results obtained from the emotion recognition system, in addition, it is important to establish that the performance tests were carried out in computer laboratories, where activities depend on the exclusive use of computer equipment. Among the tests carried out, the system was also evaluated in traditional classrooms, in which only one camera is available, even though the model is capable of recognizing people, its effectiveness in recognizing gestures and subsequent emotion is limited to the line of sight of the camera with the student. This means that if a student does not have a direct camera angle, the model assumes that a person does not exist and generates a person not detected notification. In addition, the number of students marks another limitation for the identification of gestures [42], [43]. Other works that use sensors to detect emotions may have greater success in an environment such as the one described, however, costs become a limitation that must be considered.

This work guarantees the application of the model in face-to-face educational environments, focusing on providing comprehensive help to the teacher [42]. Considering that in certain chairs it is a requirement to have computer devices, it is possible to implement the system and act as an assistant to the tutor. Having accurate information about the emotions reflected by the students, they can adjust their teaching methodologies. The potential of the system allows it to be integrated into the student service system, for this it will be necessary to add methods and libraries that allow not only to identify of gestures and emotions in students, but it will also be necessary for the system to recognize people [33]. The inclusion of these characteristics in the system will be presented in the second stage of this work, in addition, it has been proposed to improve the recognition of emotions so that the system can be applied to a hybrid education model and verify its functionality.

There are several comparisons between an emotion recognition system like the one proposed and other works that use different techniques. Some comparisons focus on the accuracy of emotion recognition models, while others consider factors such as processing speed and ease of implementation. For example, in a study published in 2020 [12], various emotion recognition algorithms were compared using standard databases. The results indicated that the deep neural network models had the best overall accuracy results. However, models based on spectral features were also found to have comparable results and were faster to train and evaluate.

Another 2019 [21] study compared different signal processing methods for emotion recognition in speech. In this case, several machine learning algorithms were compared, including convolutional neural networks and support vector machines. The results indicated that methods based on spectral features and prosodic features obtained the best overall accuracy results. The choice of the best method for emotion recognition will depend on the context of the application, the quality of the available data, and the skills of the development team. Therefore, it is important to do a thorough evaluation of the available options before making a decision.

In another study carried out in 2020 [11], different signal processing methods for emotion recognition in EEG (electroencephalogram) were compared. The methods evaluated included convolutional neural networks, recurrent neural networks, and SVMs (support vector machines). The results indicated that recurrent neural networks had the best accuracy results overall. The 2020 [42], [44] study compared different deep-learning models for emotion recognition in facial images. The models evaluated included convolutional neural networks and recurrent neural networks. The results indicated that the models based on convolutional neural networks obtained the best precision results. In a 2019 study [22], different feature selection methods for emotion recognition in the speech were compared. The methods evaluated included LDA (linear discriminant analysis), SVM, and decision trees.

The results indicated that LDA and SVM had the best overall accuracy results. A 2018 study [16] compared different signal processing methods for emotion recognition in video signals. The methods evaluated included SVM, convolutional neural networks, and recurrent neural networks. The results indicated that convolutional neural networks had the best accuracy results.

V. CONCLUSION

The gesture recognition system for the identification of emotions allows the teacher to obtain an additional variable, with which it is possible to improve the teaching method. According to the results obtained, it has been identified that students generate a variety of gestures during the teaching process, and these are generally linked to a certain emotion. By identifying the emotions of the student during a teaching process, the teacher provides feedback on her method and can make decisions that allow him to improve the teaching environment. For the teacher, having real-time information on the state of mind of the students is undoubtedly an advantage over any model of supervision of the impact of teaching on the student. About other works, it should be noted that there are several recognition algorithms and libraries for image processing, however, the applicability of these is focused on students with other types of abilities, our proposal is focused on monitoring an environment common, and we prepare a future work that adapts to the needs established in the hybrid and online educational models, these models being the future of education where the role of the teacher will need ICT with greater emphasis.

In the training of the neural network, a validation accuracy percentage of over 70% has been achieved, this percentage is adequate considering the results found in similar works that have used other platforms. These results guarantee the effectiveness of the model and the classification of student gestures, with which it is possible to identify with certainty the emotions of a certain student population.

Among the positive characteristics of the developed model, it can be established that generating the training with the students' images, allows for improved recognition of gestures. For this, the data set worked on was asked of all the students, through a video, however, in certain cases it is better to use streaming to obtain the images and generate the data set. The objective is to prevent participants from acting unconsciously and for the images obtained to be as natural as possible.

In the validation of results, it is recommended to make a comparison of the emotion recognition system with a greater number of sessions. In addition, it is important to include a greater variety of gender in the group, as well as to include in the study people who are in a higher age range than the one included in the study. By opening the age range, the aim is to build a training data set that can be used for the future inclusion of the system in an online education model. In these models, it is common for students to be older, which implies that their traits physical is different.

When analyzing the data, it is identified that the classes of the data set were unbalanced, so a greater number of sessions was carried out, to increase the number of facial expressions of the boring emotion. In the same way, teaching activities should be varied and include those that usually present problems in students, such as reading literary articles, this helps to increase the production of emotions and adjust the different parameters of the model.

According to the results obtained from the systems, there is a percentage of precision that guarantees its use in an education system and provides important information to tutors. however, several limitations may affect the effectiveness of the AI algorithm designed to detect emotions in university students. Among these limitations, it stands out that AI algorithms can be trained with limited or biased data sets, which can affect the accuracy of emotional detection's. This can lead to erroneous results and a lack of reliability in the results. Difficulty detecting subtle emotions, sometimes emotions can be subtle and hard to see, even for humans. Therefore, AI algorithms may also have a hard time detecting emotions that are not so obvious.

AI algorithms can have a hard time understanding the context in which an emotion occurs. Without context, algorithms can misinterpret emotions and provide incorrect results. Often people experience multiple emotions simultaneously, which can make it challenging to detect accurate emotions. AI algorithms may also have difficulty detecting mixed emotions and may provide inaccurate or conflicting results. Furthermore, emotions and their ways of expression can vary between cultures and geographic regions. Therefore, AI algorithms designed to detect emotion in college students may not be accurate across cultures.

REFERENCES

[1] N. Agustini, N. Nursalam, T. Sukartini, G. K. Pranata, N. W. Suniyadewi, and I. D. A. Rismayanti, "Teaching methodologies regarding palliative care competencies on undergraduate nursing students: A systematic review," *J. Int. Dental Med. Res.*, vol. 14, no. 3, pp. 1302–1308, 2021.

[2] A. K. H. AlSaedi and A. H. H. Alasadi, "A new hand gestures recognition system," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, p. 49, Apr. 2020.

[3] H.-Y. Lai, H.-Y. Ke, and Y.-C. Hsu, "Real-time hand gesture recognition system and application," *Sensors Mater.*, vol. 30, no. 4, pp. 869–884, 2018.

[4] J. Zhou, D. Yungbluth, C. N. Vong, A. Scaboo, and J. Zhou, "Estimation of the maturity date of soybean breeding lines using UAV-based multispectral imagery," *Remote Sens.*, vol. 11, no. 18, p. 2075, Sep. 2019.

[5] R. Kobai and H. Murakami, "Effects of interactions between facial expressions and self-focused attention on emotion," *PLoS ONE*, vol. 16, no. 12, Dec. 2021, Art. no. e0261666.

[6] E. G. Krumhuber, D. Küster, S. Namba, D. Shah, and M. G. Calvo, "Emotion recognition from posed and spontaneous dynamic expressions: Human observers versus machine analysis," *Emotion*, vol. 21, no. 2, pp. 447–451, 2021.

[7] F. Makhmudkhujaev, M. Abdullah-Al-Wadud, M. T. B. Iqbal, B. Ryu, and O. Chae, "Facial expression recognition with local prominent directional pattern," *Signal Process., Image Commun.*, vol. 74, pp. 1–12, May 2019.

[8] K. R. Scherer, H. Ellgring, A. Dieckmann, M. Unfried, and M. Mortillaro, "Dynamic facial expression of emotion and observer inference," *Frontiers Psychol.*, vol. 10, p. 508, Mar. 2019.

[9] M. Aziz and M. Aman, "Decision support system for selection of expertise using analytical hierarchy process method," *IAIC Trans. Sustain. Digit. Innov.*, vol. 1, no. 1, pp. 49–65, Apr. 2021.

[10] B. Gaudelus, J. Virgile, S. Geliot, and N. Franck, "Improving facial emotion recognition in schizophrenia: A controlled study comparing specific and attentional focused cognitive remediation," *Frontiers Psychiatry*, vol. 7, p. 105, Jun. 2016.

[11] S. Volynets, D. Smirnov, H. Saarimäki, and L. Nummenmaa, "Statistical pattern recognition reveals shared neural signatures for displaying and recognizing specific facial expressions," *Social Cogn. Affect. Neurosci.*, vol. 15, no. 8, pp. 803–813, Oct. 2020.

[12] G. Lou and H. Shi, "Face image recognition based on convolutional neural network," *China Commun.*, vol. 17, no. 2, pp. 117–124, Feb. 2020.

[13] Y. Takahashi, S. Murata, H. Idei, H. Tomita, and Y. Yamashita, "Neural network modeling of altered facial expression recognition in autism spectrum disorders based on predictive processing framework," *Sci. Rep.*, vol. 11, no. 1, Jul. 2021, Art. no. 14684.

[14] G. Yang, X. Xi, and Y. Yin, "Finger vein recognition based on a personalized best bit map," *Sensors*, vol. 12, no. 2, pp. 1738–1757, Feb. 2012.

[15] R. Shyam and Y. N. Singh, "Identifying individuals using multimodal face recognition techniques," *Proc. Comput. Sci.*, vol. 48, pp. 666–672, Jan. 2015.

[16] A. Villar, M. T. Zarrabeitia, P. Fdez-Arroyabe, and A. Santurtún, "Integrating and analyzing medical and environmental data using ETL and business intelligence tools," *Int. J. Biometeorol.*, vol. 62, no. 6, pp. 1085–1095, Jun. 2018.

[17] W. Villegas-Ch, X. Palacios-Pacheco, and S. Luján-Mora, "Artificial intelligence as a support technique for university learning," in *Proc. IEEE World Conf. Eng. Educ. (EDUNINE)*, Mar. 2019, pp. 1–6.

[18] X. Palacios-Pacheco, W. Villegas-Ch, and S. Luján-Mora, "Application of data mining for the detection of variables that cause university desertion," in *Proc. Int. Conf. Technol. Trends*, in Communications in Computer and Information Science, vol. 895, 2019, pp. 510–520.

[19] W. Villegas-Ch, S. Luján-Mora, and D. Buenaño-Fernandez, "Data mining toolkit for extraction of knowledge from LMS," in *Proc. 9th Int. Conf. Educ. Technol. Comput.*, Dec. 2017, pp. 31–35.

[20] J. Liu, J. Tong, J. Han, F. Yang, and S. Chen, "Affective computing applications in distance education," in *Proc. the Int. Conf. Educ. Technol. Inf. Syst. (ICETIS)*, 2013, pp. 932–935.

[21] H.-J. So, J.-H. Lee, and H.-J. Park, "Affective computing in education: Platform analysis and academic emotion classification," *Int. J. Adv. Smart Converg.*, vol. 8, no. 2, pp. 8–17, 2019.

[22] E. Yadegaridehkordi, N. F. B. M. Noor, M. N. B. Ayub, H. B. Affal, and N. B. Hussin, "Affective computing in education: A systematic review and future research," *Comput. Educ.*, vol. 142, Dec. 2019, Art. no. 103649.

[23] A. Alblushi, "Face recognition based on artificial neural network: A review," *Artif. Intell. Robot. Develop. J.*, vol. 1, no. 2, pp. 116–131, Jun. 2021.

[24] W. Villegas-Ch., J. García-Ortiz, I. Urbina-Camacho, and A. Mera-Navarrete, "Proposal for a system for the identification of the concentration of students who attend online educational models," *Computers*, vol. 12, no. 4, p. 74, Apr. 2023, doi: 10.3390/computers12040074.

[25] E. Roidl, F. W. Siebert, M. Oehl, and R. Höger, "Introducing a multivariate model for predicting driving performance: The role of driving anger and personal characteristics," *J. Saf. Res.*, vol. 47, pp. 47–56, Dec. 2013.

[26] K. R. Scherer and E. Coutinho, "How music creates emotion: A multifactorial process approach," in *The Emotional Power of Music*, T. Cochrane, B. Fantini, and K. R. Scherer, Eds. Oxford, U.K.: Oxford Univ., 2013.

[27] J. Chai, "Notice of retraction: Study on harmonious human-computer interaction model based on affective computing for web-based education," in *Proc. 1st Int. Conf. Inf. Sci. Eng. (ICISE)*, Dec. 2009, pp. 3455–3458.

[28] B. García, E. L. Serrano, S. P. Ceballos, E. J. Cisneros-Cohernour, G. C. Arroyo, and Y. E. Díaz, "Las competencias docentes en entornos virtuales: Un modelo para su evaluación," *Revista Iberoamericana de Educación a Distancia*, vol. 21, no. 1, p. 343, Sep. 2017.

[29] M. Á. R. Catalán, R. Pérez, R. B. Sánchez, O. B. García, and L. V. Caro, "Las emociones en el aprendizaje online," *Revista Electrónica de Investigación y Evaluación Educativa*, vol. 14, no. 1, pp. 1–23, Oct. 2014.

[30] E. Ivanova and G. Borzunov, "Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions," *Proc. Comput. Sci.*, vol. 169, pp. 244–248, Jan. 2020.

[31] R. W. Picard, "Affective computing," *Pattern Recognit.*, vol. 2, no. 321, pp. 85–112, 1995.

[32] P. Näykki, J. Isohätälä, and S. Järvelä, "'You really brought all your feelings out'—Scaffolding students to identify the socio-emotional and socio-cognitive challenges in collaborative learning," *Learn., Culture Social Interact.*, vol. 30, Sep. 2021, Art. no. 100536.

[33] N. A. A. Aziz, N. H. A. Aziz, S. N. M. B. S. Ismail, C. T. Khan, M. A. Hasnul, M. A. Rahman, T. A. Rahman, and K. A. Aziz, "Awareness and readiness of Malaysian university students for emotion recognition system," *Int. J. Integr. Eng.*, vol. 13, no. 6, pp. 299–309, 2021.

[34] P. Partila, M. Voznak, and J. Tovarek, "Pattern recognition methods and features selection for speech emotion recognition system," *Sci. World J.*, vol. 2015, pp. 1–7, Jan. 2015, doi: 10.1155/2015/573068.

[35] T. M. Wani, T. S. Gunawan, S. A. A. Qadri, M. Kartiwi, and E. Ambikairajah, "A comprehensive review of speech emotion recognition systems," *IEEE Access*, vol. 9, pp. 47795–47814, 2021, doi: 10.1109/ACCESS.2021.3068045.

[36] D. Singh, "Human emotion recognition system," *Int. J. Image, Graph. Signal Process.*, vol. 4, no. 8, pp. 50–56, Aug. 2012.

[37] S. Emami and V. P. Suciu, "Facial Recognition using OpenCV," *J. Mobile, Embedded Distrib. Syst.*, vol. 4, no. 1, pp. 38–43, 2012.

[38] J. Sigut, M. Castro, R. Arnay, and M. Sigut, "OpenCV basics: A mobile application to support the teaching of computer vision concepts," *IEEE Trans. Educ.*, vol. 63, no. 4, pp. 328–335, Nov. 2020, doi: 10.1109/TE.2020.2993013.

[39] M. Naveenkumar and V. Ayyasamy, "OpenCV for computer vision applications," in *Proc. Nat. Conf. Big Data Cloud Comput. (NCBDC)*, 2016, pp. 1–5.

[40] H. Adusumalli, D. Kalyani, R. K. Sri, M. Pratapteja, and P. V. R. D. P. Rao, "Face mask detection using OpenCV," in *Proc. 3rd Int. Conf. Intell. Commun. Technol. Virtual Mobile Netw. (ICICV)*, Feb. 2021, pp. 1304–1309, doi: 10.1109/ICICV50876.2021.9388375.

[41] Y. Kumar and M. Mahajan, "Machine learning based speech emotions recognition system," *Int. J. Sci. Technol. Res.*, vol. 8, no. 7, pp. 722–729, 2019.

[42] R. Alhalaseh and S. Alasasfeh, "Machine-learning-based emotion recognition system using EEG signals," *Computers*, vol. 9, no. 4, p. 95, Nov. 2020, doi: 10.3390/computers9040095.

[43] Y. Uranishi, "OpenCV: Open source computer vision library," *J. Inst. Image Inf. Telev. Eng.*, vol. 72, no. 9, pp. 736–739, 2018, doi: 10.3169/ITEJ.72.736.

[44] H. J. Yun and J. Cho, "Affective domain studies of K-12 computing education: A systematic review from a perspective on affective objectives," *J. Comput. Educ.*, vol. 9, no. 3, pp. 477–514, Sep. 2022.

WILLIAM EDUARDO VILLEGAS-CH received the master's degree in communications networks and the Ph.D. degree in computer science from the University of Alicante. He is currently a Professor of information technology with Universidad de Las Américas, Quito, Ecuador. He is also a Systems Engineer specializing in robotics and artificial intelligence. His research interests include web applications, data mining, e-learning, the design of ICT systems, and models and prototypes applied to different academic environments, especially with the use of big data and artificial intelligence as a basis for creating intelligent educational environments. He has participated in various conferences as a speaker on topics, such as ICT in education and how they improve educational quality and student learning.

JOSELIN GARCÍA-ORTIZ is currently pursuing the degree in cybersecurity engineering with Universidad de las Américas. She is a Software Engineer with Capmation Inc., Quito, Ecuador, where she applies her knowledge in software development and cybersecurity. Her research interests include the intersection of AI and its applications to education, in which she has research experience.

SANTIAGO SÁNCHEZ-VITERI was born in Quito, Ecuador, in 1984. He received the System Engineering degree from Salesiana Polytechnic University (UPS), Ecuador, in 2017, where he is currently pursuing the master's degree in telematics management. Since 2010, he has been with Universidad Internacional del Ecuador, where he is working in the area of telecommunications and networking of IT. He has been a Professor of computer science with Universidad Internacional del Ecuador. He is an administrator of computer servers and internet connection equipment. He is also an Administrator of the Moodle and CANVAS learning management systems with Universidad Internacional del Ecuador. He has participated in more than ten indexed articles on telecommunications, big data, and innovation.