Fuzzy Systems and Soft Computing
ISSN : 1819-4362

ENHANCING EMOTION RECOGNITION THROUGH COMPUTER VISION: A MODULAR APPROACH AND FUTURE RESEARCH DIRECTIONS

Dr. R. Velmurugan, Associate Professor, Presidency College, Chennai -5, India.

Abstract:

Human emotions and thoughts are reflected in facial expressions, providing a plethora of social cues such as emotion, motivation, goals, and focus of attention. Facial expressions are a powerful silent communication tool, and examining them offers a deeper understanding of human behaviour. AI-based facial expression recognition (FER) has become a significant area of study with applications in dynamic analysis, pattern identification, interpersonal interaction, mental health monitoring, and many other fields. The Covid-19 pandemic has increased the need for an effective FER analysis framework due to the growing volume of visual data produced by images and videos on online platforms. Additionally, variations in facial expressions indicating emotions in adults and senior citizens must be considered in FER research. While there has been considerable research in this area, a comprehensive overview of past work and future directions is lacking. This paper provides a thorough evaluation of AI-based FER methodologies, including datasets, feature extraction techniques, algorithms, and recent breakthroughs in facial expression identification. To the best of the authors' knowledge, this is the only review paper addressing all aspects of FER across various age brackets, aiming to significantly impact the research community in the coming years.

Keywords:

Social cues, Facial emotion recognition (FER), feature extraction, machine learning, facial expressions, Convolutional Neural Network, Deep Learning

1. Introduction

Affective computing is an interdisciplinary field that involves studying and developing systems capable of understanding and interpreting human emotions (Banafa, 2016). One of the primary motivations for research in this field is the simulation of empathy: endowing machines with the ability to detect and interpret users' emotional states and generate adaptive behavior based on the recognized information. Facial expressions are crucial in communication and convey complex mental states during interaction. In non-verbal communication, the face transmits emotions (Darwin and Prodger, 1996). Using machine learning techniques such as face recognition, information obtained from facial expressions can be processed to infer emotional states. Affective computing, which recognizes user emotional states, aims to enrich the form of user-machine interaction. A system with this capability could generate more appropriate responses considering users' emotional states.

The application of affective computing offers a wide range of possibilities. In marketing, analyzing emotions is instrumental in determining the impact of a given advertisement or product on the public. An increasing number of companies are investing in projects related to affective computing, such as the detection and prevention of stress in workers or the development of video games capable of adapting to players. This paper presents the development of software capable of detecting a user's emotions through computer vision techniques using AI algorithms, considering the theories of emotions and how to evaluate emotions with different algorithms to determine people's emotions.

The development of software capable of detecting user emotions through computer vision techniques using AI algorithms, specifically neural convolutional networks, involves face recognition performed using the framework Multitask Cascade Convolutional Networks (MTCNN). This approach considers the theories of emotions and evaluates emotions with different algorithms to determine people's emotions. The paper is structured as follows: Section 1 Introduction, Section 2 Materials and Methods, Section 3 Results, and Section 5 Discussions.

2. Materials and methods

This section delves into the foundational elements, encompassing the psychological dimensions of emotions and the various classification theories. It also extends to the technical aspects of facial recognition, exploring current techniques employed for object recognition and image classification. Furthermore, it addresses the intricacies of developing emotion recognition software using AI algorithms.

2.1. Background

Emotions are pivotal in mammals, providing essential information for survival and environmental adaptation. The perception of emotions is defined as the ability to take appropriate actions or direct thoughts and identify emotions in oneself or others. It is crucial to differentiate between emotions and feelings. Emotions arise unconsciously and rapidly, requiring no explicit mental processing. In contrast, feelings are consciously elaborated from the emotions experienced by the individual. Emotions primarily manifest as physical responses characterized by specific physiological activation patterns. However, it is noteworthy that different emotions can share similar physiological responses; for instance, both fear and anger increase heart rate. Additionally, the same emotion can elicit various responses, such as fleeing, fighting, or experiencing paralysis in the face of danger or intense fear.

2.1.1 Psychology of emotion

Emotions play a vital role in mammals, providing essential information for survival and environmental adaptation. Emotion perception can be defined as the ability to take actions or direct thoughts appropriately and identify emotions in oneself or others. Emotions come and go involuntarily, requiring no intentional thought processes. They are primarily physical reactions, symbolized by the physiological activation pattern of a trait. Occasionally, certain physiological reactions may be shared by two or more emotions.

In 1994, Paul Ekman, an anthropologist and psychologist, identified six emotions that are independent of societal influences. This range includes surprise, fear, rage, excitement, and contempt. These emotions are referred to as fundamental or universal emotions due to their intimate ties to the evolutionary patterns and survival strategies of the human species. Disdain was later added to this original collection of universal feelings.

The collection of universal emotions has spawned numerous theories that attempt to combine two or more fundamental emotions to explain the wide range of emotions people experience. Emotion classification attempts have resulted in factorial models and fuzzy category-based circumplex models. According to Russell, these models suggest responding to people's emotional states by employing the opposing extremes of the emotional categories.

The eight fundamental emotions in the circumplex model are anticipation, anger, grief, aversion, joy, trust, and surprise. This model shows how the various emotion categories relate to one another. The cone's vertical dimensions indicate the strength of the emotion, while the cone parts show how similar the feelings are to one another in terms of intensity. The secondary emotions that arise from mixing two fundamental emotions are represented by the blanks in the model that is being displayed.

2.1.2 Facial recognition

The field of face recognition began in the 1960s when W. Bledsoe's research team experimented to see if a computer could identify faces. To enable facial recognition, Bledsoe's team attempted to create correlations between the minute details of the human face.

Although Bledsoe's studies were not entirely successful, they played a crucial role in establishing the foundation for using biometric data in facial recognition (Bledsoe, 1966). It took a long time for face recognition techniques to advance significantly, but in 2001, Paul Viola and Michael Jones released an object identification system with hit rates never seen before.

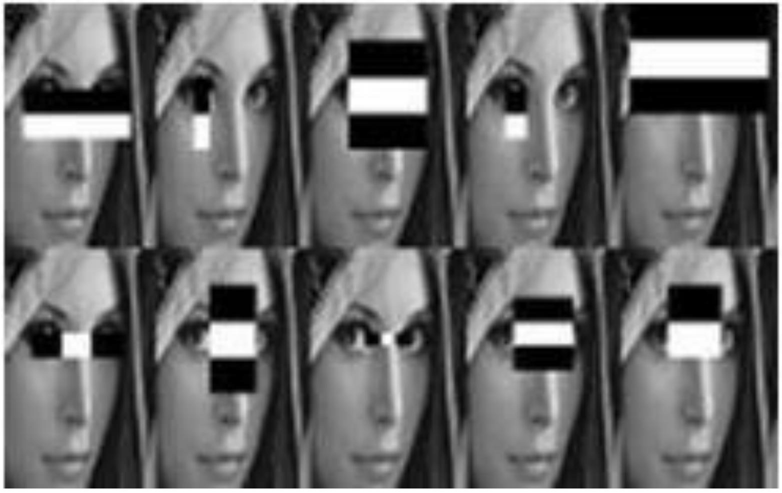


Figure 1: Facial Recognition

Following the quick "Integral Image" method's acquisition of features, they are classified. The AdaBoost classification algorithm (Schapire, 2013) is adapted for the Viola-Jones technique, which manages the process of selecting a smaller subset of features. Developed by Schapire and Freund, the AdaBoost method (Schapire, 2013) is a classifier that learns by combining several smaller classifiers. AdaBoost was initially designed to enhance the performance of existing classification methods.

AdaBoost trains each classifier independently. These classifiers then process the data that its predecessor failed to classify accurately. For each iteration, the algorithm selects the feature set with the lowest error rate. A region in an image is passed to the next classifier if it is recognized by one of the classifiers.

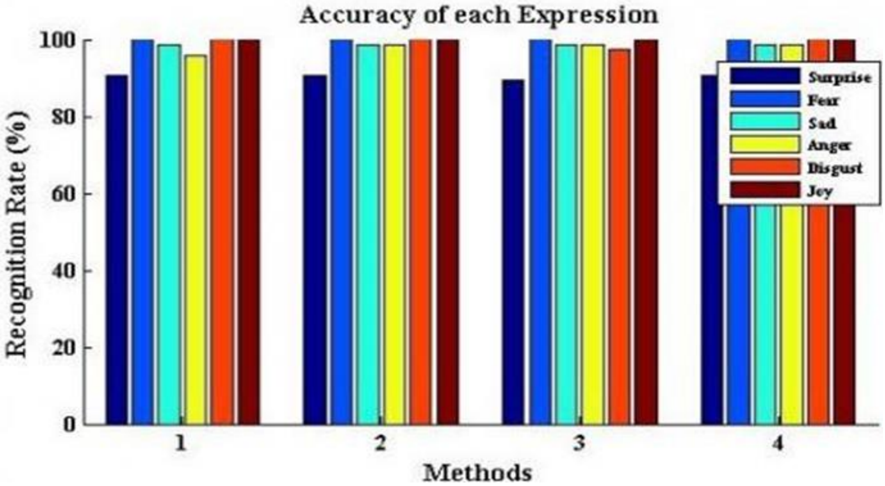


Figure 2: Accuracy Chart

2.1.3 CNN—convolutional neural networks

These days, the large availability of labeled data and ongoing advancements in GPU technology have been instrumental in creating novel Deep Convolutional Neural Network (DCN)-based facial recognition algorithms. The 2012 ImageNet competition served as the impetus for the widespread adoption of these neural networks. The convolutional layer is the most significant among the layers of neurons that make up CNNs. This layer takes a vector of pixel values as input and operates by applying a sequence of filters that pass over the image to generate the layer outputs.

Two essential parameters—stride and padding—along with the filter size employed in the convolutional layers, alter the behavior of the layer. The stride adjusts the filter's path through the image. Increasing the stride size causes the convolutional layer to focus on more distant regions of the image, suggesting a reduction in dimensionality. The Zero-Padding technique, which fills the edges with zeros, is commonly used to manage dimensionality reduction and ensure that the output is correctly sized.

Intermediate layers, or convolutional layers, are combined with other layers to remove non-linearity while preserving dimensionality, enhancing the neural network's robustness and preventing overfitting. According to Thomas et al. (2005), these layers use ReLU activation units, which are computationally more efficient than the conventional sigmoid activation function. As one moves deeper into the neural network, activation maps representing more intricate details of larger and more important areas of the image are obtained. The initial layers identify simpler visual components in smaller areas, while features identified in the initial layers contribute to higher-level representations in later layers.

The 2012 ImageNet competition was won by the AlexNet convolutional neural network architecture, marking a significant shift in computer vision. The development of AlexNet led to a surge in new architectures based on increasingly complex CNNs, consistently surpassing previous records. Additionally, AlexNet was the first to incorporate ReLU activation units, which have become the most widely used in the deep learning community.

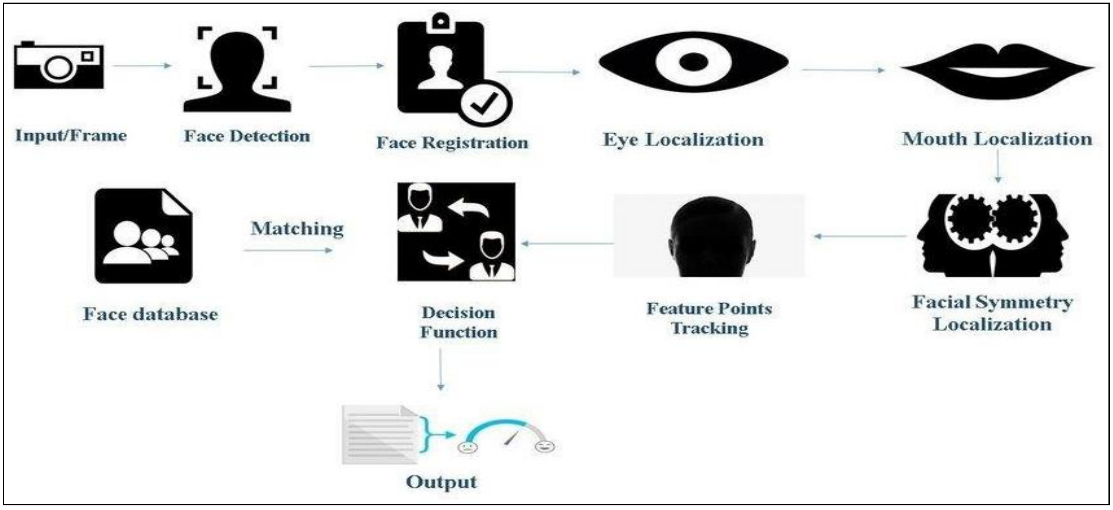


Figure 3: The Work Flow

2.1.4 Intermediate layer

The convolutional and intermediate layers of this type of neural network design are combined with other layers to enhance the network's robustness and prevent overfitting. These layers aim to introduce dimension-preserving non-linearity. ReLU activation units are foundational for these layers, as they are computationally more efficient compared to conventional sigmoid or tanh functions. ReLU activation layers use the function $f(x) = \max(0, x)$ to convert negative input values to zero.

In a neural network, the initial layers recognize more basic image structures in small regions, while deeper layers derive higher-level representations from the features identified in the earlier convolutional layers across much larger image regions. As you progress deeper into the network and pass through additional convolutional layers, the activation maps represent more complex features of larger and more significant areas of the image.

2.1.5 Exit layer

For classification tasks, the final layer of a convolutional network is called a fully connected layer. This layer produces an output vector of dimension N, where N is the number of classes for image classification, based on the input received from the previous convolutional layer. A Softmax layer is typically used for this purpose.

2.1.6 Stride and padding

Two essential factors, stride and padding, affect a convolutional layer's behavior in addition to filter size. The stride determines how the filter moves across the image. Increasing the stride size causes the features from the convolutional layer to be captured from more distant regions of the image, leading to a reduction in dimensionality. The formula

$$\left\lceil \frac{N-F}{S} \right\rceil + 1$$

, where N is the size of the input $N \times N, F$ is the size of the filter, and S is the stride size, can be used to determine the final size of the output (Simonyan and Zisserman, 2014).

The zero-padding method is widely used to preserve the dimensions of the input and output images by preventing dimensionality reduction. This method maintains the input dimensionality while "recovering" pixels lost during filter application by padding the borders of the output with zeros.

2.2 Development

2.2.1 Problem statement

Facial expressions are essential for communication, as they are the primary means of expressing complex mental states during interactions. The human face serves as the main channel for transmitting emotions in non-verbal communication. Six fundamental emotions are considered universal and innate, corresponding to distinct facial expressions: disgust, fear, joy, anger, surprise, and sadness (Ekman, 1994). Research conducted by Ekman and Friesen (Ekman, 1999) shows that while participants can accurately identify certain emotions, complete identification is still lacking.

Displaying an emotion involves complex muscle movements and the formation of identifiable patterns. Key facial regions that play a major role in expressing these emotions include the forehead and eyebrows, the eyes and eyelids, and the lower portion of the face (around the lips). Emotions fundamentally influence human decision-making.

The integration of new technologies, particularly those based on artificial intelligence, is becoming increasingly common. These technologies can adapt advertisements based on user preferences, past searches, and purchases; suggest movies and TV shows; and learn from everyday activities and user interactions with apps. Affective computing, which detects users' emotional states, enhances user-machine interaction by generating responses that align with users' emotional states.

2.2.1.1 Research question

How can we use computer vision techniques to recognize facial expressions in order to design software that recognizes emotions?

2.2.2 Objectives

Overarching goal: To create software for emotion recognition that uses computer vision algorithms to identify users' emotions based on their facial expressions.

Particular goals:

1. Create the software architecture that outlines an application's flow from beginning to end and enables the identification of users' emotions the means of a capture apparatus.
2. Put in place a graphical user interface that facilitates the application and enables the capture device's photographs to be seen in real time. The outcomes of processing the input photographs must also be shown on this graphical user interface.
3. Put in place a software element that uses an input image from the capture device to recognize faces of users. This part needs to be computationally capable enough to guarantee that the program runs correctly in real time.
4. Put in place a software element that uses a user's facial expressions to categorize their emotions. For the proper result, this component needs to be computationally efficient enough.
5. Verify the application that was obtained. Assessing the system's functionality will yield the data required to ascertain whether identifying users' facial expressions can be used to discern emotions.

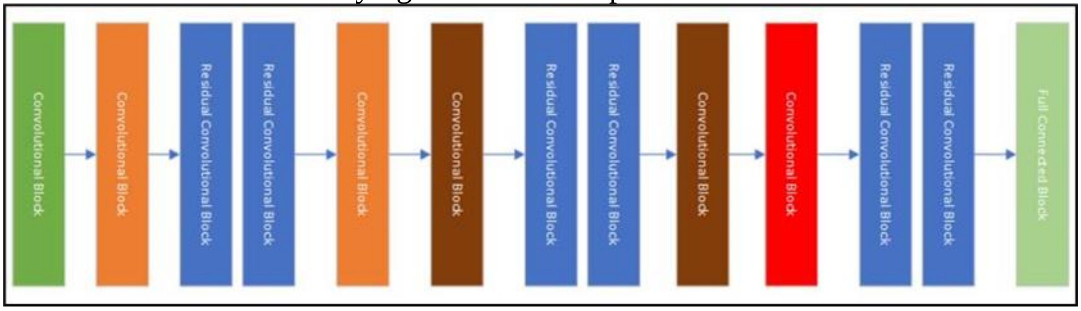

Convolutional Block
Convolutional Block
Residual Convolutional Block
Residual Convolutional Block
Convolutional Block
Convolutional Block
Residual Convolutional Block
Residual Convolutional Block
Convolutional Block
Convolutional Block
Residual Convolutional Block
Residual Convolutional Block
Full Connected Block

Figure 4: Process Chart

2.2.3 Methodology and the research approach

The aims of this work require a hybrid approach that integrates both qualitative and quantitative data within a descriptive applied research framework. To recognize emotions based on established emotional theories, video footage collected from user interactions will be classified and analyzed using a neural network and a classification model. This process will generate both qualitative and quantitative data, enabling a comprehensive study.

The software development process will follow the SCRUM methodology, an AGILE approach that allows for flexible adaptation to project needs to achieve the best possible outcomes. At the end of each Sprint—short, fixed-duration iterations in SCRUM—an intermediate product with some of the required functionality will be produced.

Functional requirements have been divided into modules and organized into three primary blocks based on their roles in the program: GUI, facial recognition, and emotion recognition. The GUI module handles the visualization of input data and the results of the emotion recognition process. The facial recognition module, which is the first step in the pipeline, identifies the user's face to process the input data. The final part of the pipeline, the Emotion Recognition module, analyzes the input data to determine the user's expressed emotion.

2.2.4 Software architecture

Designing systems that allow developers to expand and modify the software's capabilities is crucial for creating sophisticated software solutions. To achieve this, it is essential to design software architectures as distinct blocks, enabling the solution to be divided into separate, interconnected modules. This approach promotes system scalability and development efficiency. The primary benefit of modular structures is the reduction in development time, as they facilitate changes or expansions to pre-existing modules without affecting the entire system.

The architecture outlined for the proposed emotion recognition application, which utilizes computer vision techniques, follows a modular approach. This architecture consists of three main components: image capturing, processing pipeline, and output visualization. The architecture adheres to a producer-consumer paradigm with synchronized queues.

The image capturing process is responsible for initializing, gathering, processing, and making data available for the image processing pipeline. This pipeline is composed of three steps: emotion recognition, facial landmark detection, and face tracking. The results of the image processing are stored in the output synchronized queue. Result visualization involves collecting and displaying these outcomes in the graphical interface.

2.2.3 Methodology and the research approach

Based on their roles within the program, the functional requirements have been divided into modules and organized into three primary blocks: GUI, facial recognition, and emotion recognition.

- **GUI Module**: Handles the visualization of both the input data and the results from the emotion recognition process.
- **Facial Recognition Module**: The initial step in the pipeline, this module detects and recognizes the user's face to process the input data.
- **Emotion Recognition Module**: The final component of the pipeline, this module analyzes the processed input data to determine and interpret the user's expressed emotion.

2.2.4 Software architecture

Designing systems that allow developers to expand and modify the software's capabilities is essential for creating sophisticated software solutions. To achieve this, it is crucial to design architectures as modular blocks, enabling the software to be divided into separate, interconnected modules. This approach promotes scalability and development efficiency. The primary benefit of modular designs is the reduction in development time, as they facilitate the addition or modification of functionality in existing modules.

The architecture described for the proposed emotion recognition application, which utilizes computer vision techniques, consists of three main components: image capturing, the processing pipeline, and output visualization. This architecture adheres to a producer-consumer paradigm with synchronized queues.

- **Image Capture Process**: Responsible for initializing, gathering, processing, and making data available for the image processing pipeline.
- **Processing Pipeline**: Comprises three steps: emotion recognition, facial landmark detection, and face tracking. The output synchronized queue is populated with the results of the image processing.
- **Result Visualization**: Involves collecting the processed data and displaying it in the graphical interface.

The class diagram illustrates the system's hierarchy based on its classes. The abstract class Abstracting Processor is implemented by the pipeline integration components to standardize the public methods required for pipeline execution. Unlike languages like Java, which use interfaces to define the requirements and behaviors for implementation, Python does not have a formal interface system.

2.2.5 Graphical interface

The primary module of the application is the graphical user interface (GUI). This module is responsible for orchestrating the application and managing the various components, as well as representing the elements that users interact with.

During the startup phase, the GUI module initializes and configures all components, plans asynchronous actions for the application's lifecycle, and sets up the capture device.

- **Viewer**: This visual component displays images from the capture device and shows the results of facial recognition from the input images.
- **Boxes**: The GUI includes two checkboxes by default. These checkboxes allow users to enable or disable the display of results from the facial recognition of the input photographs.

2.2.6 Image processing pipeline

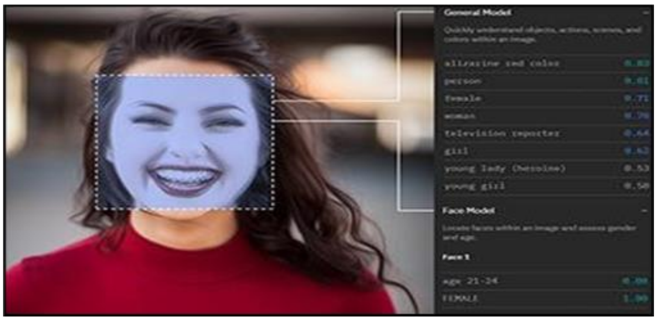
The processes required to identify emotions from the input data are outlined by the image processing pipeline. An asynchronous task consumes the data from the input line and executes this pipeline, while the results are sent to the output line for updating graphical components. The pipeline consists of three key stages for identifying the user's emotions:

1. **Face Tracking**: This phase involves detecting the face in the input image and providing the output for the next component in the image processing pipeline. It uses the Multitask Cascade Convolutional Networks (MTCNN) framework (Zhang et al., 2016) for face recognition. MTCNN is known for its high accuracy in facial recognition and performs exceptionally well for real-time applications, even on a CPU. The MTCNN architecture includes three cascaded convolutional networks: P-Net, R-Net, and O-Net, which work together to detect faces.
2. **Facial Landmarks Detection**: This stage identifies 64 key facial landmarks based on the outcomes of the facial recognition process. These landmarks help in precise face alignment and are crucial for accurate emotion recognition. This information is valuable for positioning the capture device correctly during the setup.
3. **Emotion Recognition**: The final stage analyzes the facial landmarks and other features to classify the emotional state of the user. This stage integrates the data from the previous steps to determine the user's expressed emotions accurately.

Table 1: Training Parameters

Training parameters	**Value**
Epoch number	50
Batch size	32
Learning rate	1.00E-01
Optimization	Stochastic Gradient Descent
Loss function	Cross Entry Loss

CNNs are utilized to identify the 64 crucial facial landmarks. The chosen network architecture for this task is Exception Net. This architecture is preferred due to its straightforward definition and ease of modification, which results from its design as a linear stack of depth-separable convolution layers with residual connections.



General Model

Quickly understand objects, actions, scenes, and colors within an image.

alizarine red color 0.83
person 0.81
female 0.71
woman 0.70
television reporter 0.64
girl 0.62
young lady (heroine) 0.53
young girl 0.50

Face Model

Locate faces within an image and assess gender and age.

Face 1

age 21-24 0.98
FEMALE 1.00

Figure 5: Emotion Recognition

Emotion Recognition: In this step, the facial expressions captured during Step 1: Face Tracking are used for emotion recognition. Convolutional Neural Networks (CNNs) are employed for classifying these facial expressions. Figure 9 illustrates the architecture developed for this classifier, which is an adaptation of the ResNet architecture (Centeno, 2021; Zhao et al., 2022).

The facial expression classification algorithm is trained using the FER-2013 public dataset. This dataset consists of grayscale images of faces, each 48 × 48 pixels in size. The FER-2013 dataset includes 28,000 labeled images for training and 3,500 images for validation. Each image in the dataset is categorized into one of seven emotions: 0 for angry, 1 for disgust, 2 for fear, 3 for joyful, 4 for sad, 5 for surprise, and 6 for neutral (Kaggle, 2019).

3. Outcomes

Initially, the emotion categorization and detection software was tested using pictures to evaluate its performance in identifying emotions such as sadness, fear, anger, surprise, disgust, and happiness. Subsequently, real-time user photos and multimedia content were used for classification and emotion detection tests, alongside a control group to compare results and determine the accuracy of emotion detection.

The experiments to assess the application's performance will be detailed in this chapter, focusing on the software tool for emotion recognition. The tests were conducted in two distinct scenarios: with images and with user-generated content.

In the image-based tests, the application processes the images, holding each in front of the capture device until it produces a result. If the application fails to generate a result for any image, it is considered unsuccessful. For instance, the tests on images categorized as "Angry" successfully highlighted the subject's distinguishing features. Based on the interface data, the system's classification performance can be evaluated. It is crucial to note that the selection of images introduces subjectivity; some images may belong to more than one category due to their contextual nature.

A confusion matrix based on these image test results is provided. This matrix includes an additional "Neutral/No Clear" row to capture outcomes where the algorithm fluctuated between categories without producing a definitive result.

The confusion matrix helps visualize the classification algorithm's performance during testing. Each column represents the number of predictions for each class, while the rows indicate the actual number of instances in each class. By analyzing the matrix, insights can be drawn about the types of errors made by the model during image testing.

Table 2: Emotion Category

Category	Sad	Fear	Angry	Surprise	Disgust	Happy
Sad	9	1	1	-	-	6
Fear	-	8	-	1	-	-
Angry	1	1	12	-	2	-
Surprise	-	2	-	14	-	-
Disgust	1	-	-	-	8	-
Happy	-	1	-	2	-	18
Neutral/No Clear	8	1	6	2	2	-
Precision	0.47	0.44	0.63	0.73	0.42	1
Accuracy	0.83	0.93	0.9	0.93	0.89	0.95

The accuracy metric, which measures the proportion of correct positive predictions, is shown in Table 2. It reveals that all classes achieved an accuracy rating greater than 82%. This high accuracy suggests that the model has strong predictive capabilities. Specifically, the accuracy values for the "angry," "surprised," and "happy" categories were above 60%, while other categories had values below 50%. This metric helps estimate the model's overall predictive performance.

Table 3: Accuracy of different emotions

Category	Precision	Recall	F1-score	Support
Angry	0.49	0.43	0.46	958
Disgust	0.49	0.55	0.52	111

Fear	0.42	0.25	0.32	1,024
Happy	0.98	0.9	0.87	1,774
Neutral	0.39	0.78	0.52	1,233
Sad	0.54	0.18	0.27	1,247
Surprise	0.83	0.79	0.75	831

The findings reinforced the idea that emotions are multidimensional, with certain images previously classified by the author fitting into multiple emotional categories. This overlap occurs because two emotions may exhibit similar physiological activation patterns. Comparing the results with the model evaluation metrics shown in Table 3, it is evident that, except for the emotions "angry," "surprised," and "happy," which demonstrate a slight improvement in model validation, the accuracy metric follows a similar trend across test results.

Russell's Circumplex Model (Figure 3) helps explain the lower hit rates observed in some emotional categories (Russell, 1980). The model shows that most of the identified emotional categories are concentrated in Quadrants I and II.

In Quadrant I, "happy" and "surprise" are emotions that represent varying degrees of pleasure and stimulation. While "happy" denotes high pleasure and lower stimulation, "surprise" has high activation but low pleasure, making them complementary feelings within the same quadrant.

In Quadrant II, the emotions "disgust," "anger," and "fear" are positioned close to each other. This quadrant reflects similar levels of unpleasantness activation among these emotions. "Anger" shows the highest hit rate in this quadrant, aligning with its intermediate position between "disgust" and "fear" in Russell's circumplex model.

This analysis highlights how the model's performance correlates with the theoretical understanding of emotional dimensions and their overlaps.

Figure 6(a): Different Expressions

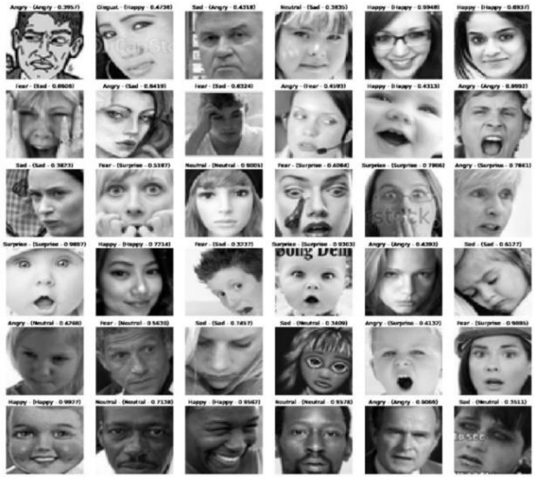


Figure 6(b): Different Expressions

4. Discussions

Building a system based on computer vision techniques to accurately categorize the full spectrum of user emotions is challenging, as evidenced by the volunteer tests detailed in the paper and the inherent complexities in precise emotion classification (Monteith et al., 2022). This essay underscores the difficulty of defining emotions accurately and argues for a comprehensive understanding of emotions to address the challenges of classification. The similarities in physiological activation patterns among different emotions complicate the process of identification.

Artificial intelligence (AI) holds significant potential in emotion recognition, offering benefits across various domains such as mental health support, personalized user experiences, education, security, entertainment, and the military (Chollet, 2017; Lu, 2022). Emotion-aware systems can facilitate the development of compassionate AI applications, from virtual assistants that respond to users' emotional states to mental health tools providing timely, context-sensitive interventions. Incorporating emotion recognition into AI aligns with the growing emphasis on human-centric technology, fostering more sophisticated and flexible interactions between people and digital systems (Lee and Park, 2022).

During the tool's development and evaluation, it became evident that additional contextual information was necessary for accurate emotion determination. While some facial expressions, such as joy or surprise, can be readily identified without background information, this is not always the case. Russell's Circumplex Model illustrates the study's inherent complexity, with Quadrant I including emotions like "Happy" and "Surprise," and Quadrant II encompassing more than 57% of the identified categories: "Disgust," "Anger," "Fear," and, to a lesser extent, "Sad." The significant variations within these groups, which represent 25% of Russell's model, make accurate identification challenging.

Despite this complexity, the potential to develop effective emotion recognition systems using computer vision remains promising. To enhance the accuracy of emotion detection, future strategies should incorporate additional features. This may involve retraining models with diverse datasets, improving image quality and resolution, and expanding testing to include a broader range of users and scenarios.

The main objective of developing a tool capable of identifying user emotions through computer vision techniques has been achieved. The application successfully demonstrates real-time emotion identification, with convolutional neural networks underpinning the facial emotion categorization, face detection, and facial landmark recognition components. Three different convolutional network architectures, chosen for their computational efficiency, were employed in this study.

Future research should focus on refining the accuracy of facial expression classification. Precise emotion categorization necessitates understanding the context of detected facial expressions. Although significant research remains, affective computing holds great promise for societal impact. As affective computing solutions, such as virtual assistants, become more integrated into daily life, they will offer users highly personalized experiences.

This presentation provides a glimpse into ongoing research in affective computing aimed at accurately detecting user emotions during system interactions. It highlights the complexity of emotion identification, emphasizing the need for continued refinement of emotion recognition tools. The challenges encountered underscore the intricate nature of emotion classification, a task that, despite appearing straightforward to humans, presents significant challenges in computational systems.

5. Conclusions

In order to acquire contextual Information on emotional Classification, the Facial Recognition feature has become highly inevitable in the current scenario. Suggested improvements include using larger, higher-resolution images and expanding testing to include more users and videos. Incorporating temporal series of facial expressions could further enhance classification accuracy. In spite of the dependency on Single Snapshots, the Training emotion classifiers with Long Short-Term Memory (LSTM) networks, which analyse sequences of images, could provide context by capturing the progression of facial expressions over time.

The results from the current tests, particularly with unclear or challenging expressions, indicate that the model could benefit from retraining with a more diverse dataset containing additional images. The accuracy and certainty of emotional classifications can be improved by increasing image resolution and dimensions as well as more user videos in testing. Additionally, investigating micro-expressions in brief intervals may reveal further insights and enhance the model's precision.

References

[1] Darwin, C., and Prodger, P. (1996). The Expression of the Emotions in Man and Animals. Oxford: Oxford University Press.

[2] Ekman, P. (1994). Strong evidence for universals in facial expressions: a reply to Russell’s mistaken critique. Psychol. Bull. 115, 268–287. doi: 10.1037/0033-2909.115.2.268

[3] Ekman, P. (1999). Basic emotions. Handb. Cogn. Emot. 3, 45–60. doi: 10.1002/0470013494.ch3

[4] Ekman, P., Sorenson, E., and Friesen, W. (1969). Pan-cultural elements in facial displays of emotion. Science 164, 86–88. doi: 10.1126/science.164.3875.86

[5] Frijda, N. H. (2017). The Laws of Emotion. London: Psychology Press. García, A. R. (2013). La educación emocional, el autoconcepto, la autoestima y su importancia en la infancia. Estudios y propuestas socioeducativas. 44, 241–257.

[6] Ghotbi, N. (2023). The ethics of emotional artificial intelligence: a mixed method analysis. Asian Bioethics Rev. 15, 417–430. doi: 10.1007/s41649-022-00237-y

[7] Hernández Sampieri, R., Fernández, C., and Baptista, L. C. (2003). Metodología de la Investigación. Chile: McGraw Hill. Kaggle (2019). FER−2013. Available online at: https://www.kaggle.com/ (accessed October 5, 2023).

[8] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. doi: 10.1145/3065386

[9] Lee, Y. S., and Park, W. H. (2022). Diagnosis of depressive disorder model on facial expression based on fast R-CNN. Diagnostics 12:317. doi: 10.3390/diagnostics12020317

[10] Lu, X. (2022). Deep learning based emotion recognition and visualization of figural representation. Front. Psychol. 12:818833. doi: 10.3389/fpsyg.2021.818833 Mathworks (2023).

[11] Monteith, S., Glenn, T., Geddes, J., Whybrow, P. C., and Bauer, M. (2022).Integral Image. Available online at: https://www.mathworks. com/help/images/integral-image.html (accessed October 16, 2023).

[12] Commercial use of emotion artificial intelligence (AI): implications for psychiatry. Curr. Psychiatr. Rep. 24, 203–211. doi: 10.1007/s11920-022-01330-7

[13] Plutchik, R. (2001). The nature of emotions. Am. Scientist 89, 334–350. doi: 10.1511/2001.28.334

[14] Plutchik, R. E., and Conte, H. R. (1997). Circumplex Models of Personality and Emotions. Washington, DC: American Psychological Association.

[15] Russell, J. A. (1980). A circumplex model of effect. J. Personal. Soc. Psychol. 39:1161. doi: 10.1037/h0077714

[16] Russell, J. A. (1997). "Reading emotions from and into faces: resurrecting a dimensional-contextual perspective," in The Psychology of Facial Expression, eds J.

[17] A. Russell and J. M. Fernández-Dols (Cambridge University Press; Editions de la Maison des Sciences de l'Homme), 295–320.

[18] Salovey, P., and Mayer, J. (1990). Emotional Intelligence. Imag. Cogn. Personal. 9, 185–211. doi: 10.2190/DUGG-P24E-52WK-6CDG

[19] Sambare, M. (2023). Kraggle. FER-013. Learn Facial Expresions From a Image. Available online at: https://www.kaggle.com/datasets/msambare/fer2013 (accessed October 16, 2023).

[20] Schapire, R. E. (2013). "Explaining adaboost," in Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik (Berlin; Heidelberg: Springer), 37–52. doi: 10.1007/978-3-642-41136-6_5