Real-Time Emotion Recognition in Psychological Intervention Methods

Sebastián Ramos-Cosi[1], Daniel Yupanqui-Lorenzo[2], Meyluz Paico-Campos[3], Claudia Marrujo-Ingunza[4], Ana Huamaní-Huaracca[5], Maycol Acuña-Diaz[6], Enrique Huamani-Uriarte[7]
Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, Lima, Perú[1, 4, 7]
E-Health Research Center, Universidad de Ciencias y Humanidades, Lima, Perú[2, 5]
Faculty of Engineering, Universidad de Ciencias y Humanidades (UCH), Lima, Perú[3, 6]

***Abstract*—In the context of mental health, this study aims to develop a real-time emotion-focused facial recognition system based on psychological intervention methods. It uses a convolutional neural network (CNN) base and is trained with the FER2013 dataset, which consists of 35,887 facial images classified into seven basic emotions. Through normalisation, data augmentation, and training in TensorFlow and Keras, the model achieved 92.3% accuracy in a pilot test with 1,000 images, achieving an F1 score of 0.92, precision of 0.93, and recall of 0.91. Subsequently, when scaled to 71,774 images, it maintained robust performance with an overall accuracy of 77.5%. Emotions such as happiness (0.83), surprise (0.80), and neutrality (0.85) were recognised with greater accuracy, while K-means analysis was applied to cluster emotional patterns in a visually interpretable way. Complementing the technical architecture, a user-friendly graphical interface was designed for psychology professionals, allowing clear visualisation of the detected emotions with a latency of just 150 milliseconds per image. Overall, this proposal represents a significant advance toward more interactive, personalised, and efficient therapies, without requiring a complex technological infrastructure. Future studies recommend exploring different multimodal signals and increasing the use of convolutional layers to improve the quality of results and data efficiency.**

***Keywords*—Facial recognition; real-time; methods; psychological interventions**

I. Introduction

In today's mental health context, understanding and responding appropriately to patients' emotions during psychological interventions has become key to achieving more effective and personalized care [1], [2]. In this sense, emotions have a direct impact on how people communicate, process their experiences, and react to therapeutic strategies [3]. Identifying these emotions in real time allows professionals to modify their approach during sessions [4], strengthening the therapeutic relationship and optimizing treatment outcomes.

Traditionally, psychologists have interpreted emotions through direct observation of gestures [5], tone of voice [6], and body language [7]. However, these methods can be limited by the observer's subjectivity, the practitioner's cognitive load, or environmental conditions, especially in online therapies [8]. Given this, these limitations have prompted the development and implementation of technological tools that enhance clinical work [9] through the automatic analysis of emotional signals.

The advent of Artificial Intelligence (AI) and Machine Learning (ML)-based technologies has enabled significant advances in automatic emotion recognition [10], [11], particularly through the study of facial expressions. These systems use advanced algorithms to identify visual patterns linked to different emotions, providing accurate, real-time assessments. Although these tools have been applied in other sectors, their integration into psychological contexts is still limited [12], despite their enormous potential to improve patient emotional understanding and support therapists during sessions.

Facial expressions are an essential source of emotional data in therapeutic contexts [13], as they facilitate the detection of internal states that are often not expressed verbally. In a psychological context, this detection of emotions facilitates the development of the relationship between the professional and the patient [14], which increases trust and the final outcome of the therapy.

The objective of this study is to develop a real-time emotion-focused facial recognition system for psychological intervention methods. This proposal seeks to contribute to the design of technological solutions that support clinical work, improve the patient experience, and promote more timely and effective interventions.

To carry out this study, a system based on convolutional neural networks (CNN) will be designed using the FER2013 dataset, which contains facial images labeled according to seven basic emotions [15]. Preprocessing techniques such as data normalization and augmentation will be applied, and the model was trained in an optimized environment with TensorFlow and Keras. Validation will be performed through a pilot test, evaluating its performance in real time. Finally, a graphical user interface (GUI) will be implemented, aimed at psychologists, allowing the visualization of emotions detected during clinical sessions for therapeutic and analytical purposes.

The importance of this system lies in its precision and ability to execute in real time, making it a suitable tool for application in clinical settings. By demonstrating the model's effectiveness in accurately recognizing different emotions during therapy sessions, it underscores AI's ability to modify psychological practices, enabling more personalized interventions tailored to the patient's emotional state.

The study is structured as follows: the Related Works section reviews and analyses the existing literature related to the topic of this study. The Methodology section then describes the pre-processing and processing steps, along with the architecture of the CNN model used. Following this, the Results section presents the analysis obtained after the necessary implementations and tests. The Discussion section presents a preliminary comparison with the literature to inform future studies. Finally, the Conclusion section summarises the findings and proposes new lines of future research.

II. Related Works

Despite significant progress in emotion recognition using AI techniques, most existing studies have focused on general applications such as human-computer interaction, education, security, and medical diagnosis. However, their limited exploration in the field of direct implementation in the clinical context of real-time psychological interventions remains a serious problem. This gap restricts the use of the therapeutic potential offered by these systems to strengthen the relationship between the professional and the patient, dynamically modify intervention tactics, and optimise treatment outcomes.

Facial emotion recognition has advanced significantly due to the integration of Artificial Intelligence (AI), especially with systems capable of processing data in real time. Hadjar et al. [16] developed TheraSense, designed to improve teleconsultation services through deep learning. This system demonstrated effectiveness in real-time emotion detection during video streams, standing out for its usability and integration into remote consultations. Similarly, Saadon et al. [17] proposed an alternative method, based on digital image correlation (DISC), which captured subtle variations in facial expressions with high accuracy and without racial or gender biases, compared to commercial systems such as Amazon Rekognition.

Additionally, the review by Kaur et al. [18] emphasised the growing relevance of facial emotion recognition (FER) in various sectors, including medical diagnosis, vehicular automation, and educational assessment. This comprehensive analysis underlined the importance of FER in human-computer interaction, providing a broad context for future research and methodologies. In this direction, Elsheikh and Mohamed [19] presented an innovative model based on deep convolutional neural networks with anti-aliasing techniques (AA-DCN), highlighting its effectiveness in complex databases such as CK+, JAFFE, and RAF, achieving high accuracy rates against challenges such as lighting, occlusions, and cultural diversity.

From a clinical perspective, Rubin et al. [20] explored specific deficits in emotional facial recognition in patients with psychotic disorders such as schizophrenia and bipolar disorder, compared to healthy controls. They identified significant decreases in accuracy and speed of emotional identification, highlighting the importance of early personalized interventions for these patients. In parallel, Economou et al. [21] investigated the associations between emotional facial recognition and schizotypal traits in the general population, revealing specific difficulties related to certain dimensions of schizotypism and their direct impact on psychological well-being.

The implementation of multimodal methods has also brought important advances to emotional facial recognition. Ballesteros et al. [22] developed a system that combines convolutional neural networks (CNNs) with psychological theories, achieving satisfactory results and pointing out the need for additional training to improve accuracy in diverse contexts and similar emotions. In a similar approach, CNNs were integrated with visual transformers (CoAtNet) [23], adding facial key points through multimodal fusion, which considerably increased the accuracy of emotional recognition, being successfully implemented on hardware such as Raspberry Pi.

On the other hand, Hassouneh et al. [24] proposed a hybrid system that combines facial expressions and EEG signals for children with autism spectrum disorder (ASD), using advanced architectures such as Xception and IoT and fog technologies computing to reduce latency and improve quality of life, achieving high accuracy and sensitivity rates. In parallel, Talaat et al. [25] also demonstrated the effectiveness of emotional facial recognition in children with ASD, using DCNN and fog technology computing to provide fast and effective responses in real time.

Finally, in the area of security and privacy, Casaño et al. [26] implemented a facial recognition-based authentication system integrated with AI and blockchain for the company Dialyma. They used technologies such as Python, OpenCV, TensorFlow and biometric services such as Reniec, managing to guarantee security and privacy in the management of personal data and reducing the risks associated with traditional passwords. This background highlights the potential and growing importance of emotional recognition in clinical and technological contexts, establishing a solid foundation for further study of real-time emotion recognition applied specifically to psychological intervention methods.

In conclusion, the studies analysed show significant progress in the development of emotion recognition systems, highlighting their effectiveness in diverse contexts through the use of techniques such as convolutional neural networks, deep learning, and hybrid models. However, there remains limited integration of these technologies in real-life clinical settings, especially in real-time psychological interventions. This review identifies a need to refine and validate these tools in therapeutic contexts, taking into account both the specific characteristics of personal relationships and the ethical and practical challenges involved in their application. In this context, this analysis suggests a technological solution focused on psychological practice, with the aim of contributing to the development of more sensitive, personalised interventions based on precise emotional data.

III. Methodology

This section describes the method used in this study, including the dataset description, the preprocessing and processing steps, the CNN model architecture, the analysis techniques used, and the testing environment focused on psychological interventions. The goal is to ensure the development of an accurate and feasible system for use in real-time clinical situations.

A. Dataset Description

For training and evaluation of the model, the FER2013 dataset was used, a resource commonly used in research on emotion recognition through facial expressions [15]. This set contains features that are described in Table I.

TABLE I. FEATURES FOR MODEL TRAINING

Feature	Description
Set size	35,887 images
Resolution	48 x 48 pixels
Format	Grayscale images
Emotions labeled	Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral
Capture conditions	Diversity in lighting, facial positions and demographic conditions
Applications main	Evaluation of CNN models for facial emotion recognition

The diversity of demographics, lighting, and facial positions in FER2013 allows for the training of a model capable of generalising across diverse clinical contexts. Furthermore, its validation in previous studies and its accessibility facilitate the replication of experiments and the development of advanced AI techniques for real-time emotion recognition.

In addition to its technical advantages, the FER2013 suite has significant limitations, such as the lack of colour images, contextual information, and an exclusive focus on basic emotions. Therefore, we recognise the need to evaluate the suggested model in future research, complemented by more extensive and complex databases, such as AffectNet or RAF-DB, which include greater emotional variability. These evaluations could confirm and greatly improve the scalability and robustness of the system in diverse therapeutic contexts and clinical groups.

B. Data Pre-processing

The flowchart in Fig. 1 shows the steps involved in the process. The first pre-processing step was to normalise the images from the FER2013 set, ensuring that all samples were grayscale and had a uniform resolution of 48x48 pixels.

A data augmentation technique was applied by horizontally inverting each image along the Y axis, duplicating the original set and generating a mirrored version of each facial expression. This strategy preserved emotional characteristics while introducing spatial variability, improving the model's ability to generalise and reducing the risk of overfitting. As part of the experimental design, a pilot phase was conducted with 1,000 images to train a preliminary model. This allowed us to validate the pre-processing workflow and fine-tune hyperparameters before scaling the training to 71,774 images, ensuring efficiency and performance.

C. Training Tools

- FER2013: The FER2013 dataset is a set of facial images used to train emotion recognition models [27]. It was used in this study due to its ease of integration and immediate results.

- TensorFlow: It is an open-source library developed by Google for creating and training ML [28], [29]. Its use in this study lies in the construction of the Convolutional Neural Network (CNN) architecture and integration with FER2013 images. It also accelerates the implementation process and facilitates future integration of the model into clinical applications.

- Keras: It is a high-level interface for building and training neural networks, based on a TensorFlow API [30]. In this study, it was used to build the CNN architecture and develop rapid and efficient prototypes, which is essential during pilot testing.

- Google Colab Pro: It is a free cloud-based platform that allows you to run Python code, as well as access to 100 processing units and GPU access for computation-intensive tasks [31]. It is used to train the CNN model using the 1000-image subset of the FER2013 dataset. It was chosen for its accessibility, library integration, and accelerated processing capabilities without requiring local resources.

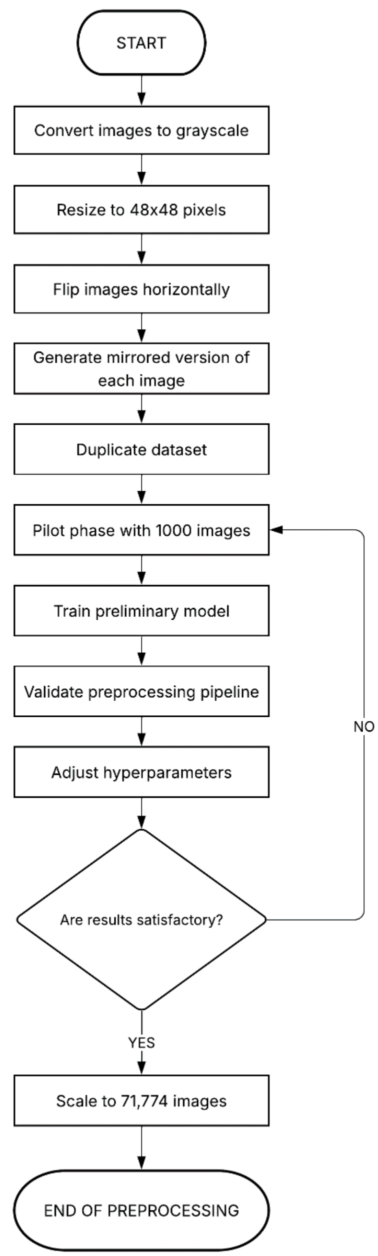

Fig. 1. Sampling design and implementation flowchart

D. CNN Architecture

In this study, the emotion recognition model developed is based on a CNN architecture, selected for its proven effectiveness in image classification tasks, particularly facial expression recognition. This design facilitates the gradual extraction of relevant patterns from the input images, with the goal of identifying basic emotions such as happiness, sadness, anger, fear, surprise, disgust, and neutrality in real time.

The model accepts grayscale images with input dimensions of 48x48, corresponding to the structure of the FER2013 dataset. These images are pre-processed to ensure normalisation and compatibility with the network.

The specific structure of the CNN model used for the pilot test with 1000 images is described below. This architecture, shown in Fig. 2 and Table II, was designed to be lightweight and efficient, suitable for training with small datasets, and oriented toward real-time applications.

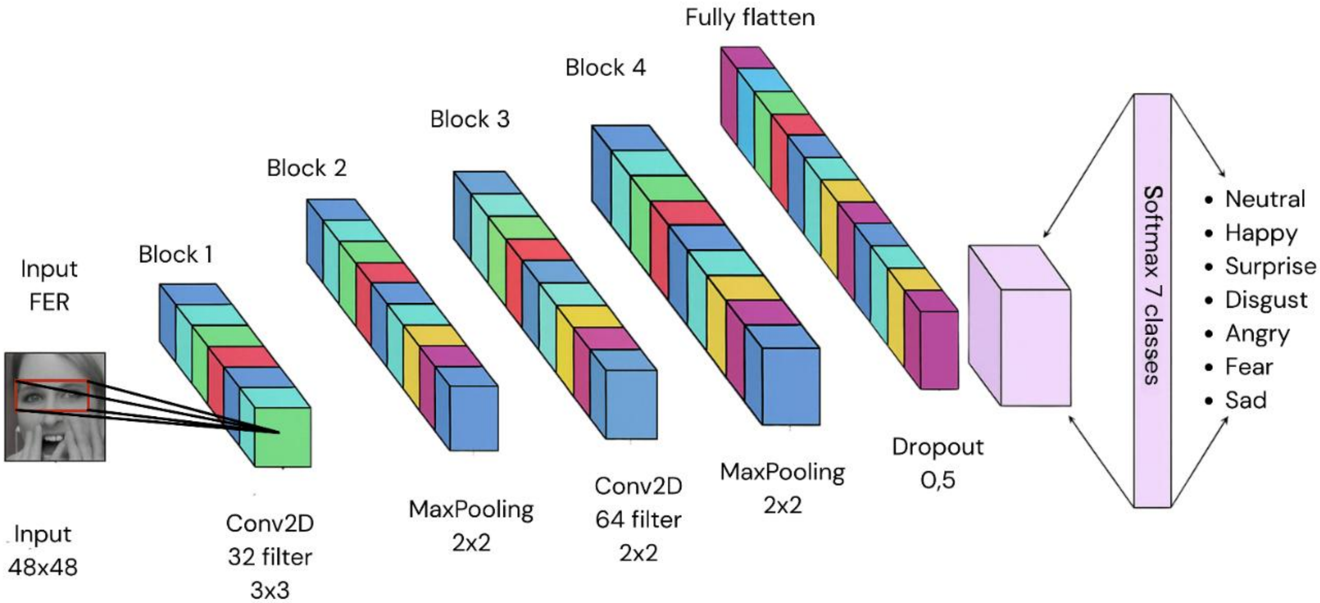

Fig. 2. CNN architecture

TABLE II. FEATURES OF CNN ARCHITECTURE

Layer	***Filters / Units***	***Kernel Size***	***Function***
Input layer	-	48x48x1	Receive grayscale image
Conv2D 1	32	3x3	Detect simple patterns (edges, lines)
MaxPooling 1	-	2x2	Reduce spatial dimension
Conv2D 2	64	3x3	Detect complex patterns (facial features)
MaxPooling 2	-	2x2	2° reduction of spatial dimension
Dropout	-	-	Avoid overfitting in training
Output layer (Softmax)	7	-	Classify among 7 basic emotions

1) Input layer (Input 48x48x1). This layer receives grayscale facial images with a resolution of 48x48 pixels and a single channel. This dimension is derived from the FER2013 dataset, ensuring direct compatibility without the need for additional transformations.

2) First convolutional layer (Conv2D – 32 filters, 3x3). This first layer applies 32 convolution filters of 3x3 pixel size, enabling the detection of basic features such as edges, corners, and texture transitions; and overall, improving learning capacity.

3) First pooling layer (MaxPooling 2x2). This layer reduces the dimensionality of the feature maps produced by the first convolution using a 2x2 pixel kernel. This improves computational efficiency and preserves the most relevant features.

4) Second convolutional layer (Conv2D – 64 filters, 3x3). This second convolutional layer increases the complexity of the analysis with 64 filters of 3x3 pixels, allowing the detection of more detailed patterns such as combinations of facial features (frown, smile, among others).

5) Second pooling layer (MaxPooling 2x2). This layer reduces the spatial dimensions of the already selected maps, in order to prepare the data for the following layers and to reduce the risk of overfitting.

6) Dropout layer (Dropout – 0.5). This layer randomly deactivates 50% of neurons during training, which forces the model to avoid overly relying on certain nodes. This leads to improved generalization capabilities, especially on specific datasets.

7) Output layer (Softmax – 7 classes). This last layer uses a softmax activation function, which converts the output into a distribution over 7 emotions (neutral, happy, surprise, disgust, angry, fear, sad). The model predicts the emotion with the highest probability as the final output.

E. Analysis Techniques

The analysis of the results obtained after real-time emotion detection is carried out using two main approaches: structured data collection in CSV files (Comma Separated Values) and the application of the K-means clustering algorithm. These techniques allow us to evaluate the performance of the CNN model and organize emotional patterns in greater depth.

During the development of the system, a mechanism for automatically recording the results of the emotional classification was implemented. Each time the model identifies an emotion in real time, the information is saved in a CSV file. This file includes important information such as the emotion detected, the model's confidence level, and the exact time of detection. This approach allows for the construction of a chronological database of emotional reactions, which can be used by clinical psychologists to examine how a person's emotional state changes during a session at specific times.

To identify shared emotional patterns in the recorded data, we used the unsupervised K-means clustering method. This method facilitates the separation of identified emotions into groups based on common aspects such as emotional intensity and similarity between facial expressions. Its use facilitates data exploration from a more visual and interpretive perspective and segments patients' emotional profiles.

F. Test Environment in Psychological Interventions

To evaluate the clinical applicability of the real-time emotion recognition model, a functional and user-friendly graphical user interface (GUI) was designed for psychology professionals. This interface allows users to upload facial images or activate the camera to process expressions and obtain the predominant emotion along with the prediction confidence level. Its user-centred design prioritises visual clarity, direct interpretation of results, and non-invasive integration during therapy sessions, allowing the specialist to observe the detected emotions in parallel.

In this study, the term "real-time" refers to the system's ability to process facial expressions with an average latency per image, ensuring an immediate response during the session. To improve visual stability without losing immediacy, temporal averaging over moving 1-second windows was incorporated, smoothing out abrupt variations without compromising instantaneous detection.

IV. Results

This section presents the main results obtained after the implementation and testing of the emotion recognition system. It describes the findings achieved during the pilot phase of training the CNN model, the real-time classification records, the emotion clustering analysis, and the behaviour of the graphical interface developed for clinical use of the system:

A. Data Analysis

In order to evaluate the feasibility and initial performance of the proposed model, a pilot test was developed using a subset of 1000 images from the FER2013 set. The data is shown in Table III. This representative sample included seven basic emotions distributed proportionally and allows validating the CNN model architecture and the most appropriate training configuration. Three iterations of hyperparameter tuning were performed, modifying the number of filters per layer, the batch size and the learning rate. Finally, a pipeline with two convolutional layers (32 and 64 filters), a grouping using MaxPooling, a dropout layer with a rate of 50% and a softmax output layer for multiclass classification was consolidated.

The model was trained for 50 epochs on the Google Colab platform, achieving a total accuracy of 92.3% on the validation set. The average performance metrics per class showed a precision of 0.93, a recall of 0.91, and an F1-score of 0.92, demonstrating a robust ability to discriminate between different emotional expressions, even with slight variations in lighting or posture. This configuration proved to be optimal in terms of both accuracy and computational efficiency, validating its appropriate capacity for real-time implementation within controlled clinical environments.

TABLE III. CNN Model Configuration and Results

Iteration	Filters	Batch Size	Learning Rate	Accuracy (%)	Precision	Recall	F1 score	Eras	Time / epoch
1	32 / 64	64	0.001	85.6	0.86	0.84	0.85	50	5.1 s
2	32 / 64 / 128	32	0.0005	89.8	0.90	0.89	0.89	50	5.3 s
3	32 / 64 / 128	64	0.0003	92.3	0.93	0.91	0.92	50	5.2 s

B. CNN Model

In this section, Fig. 3, illustrates the normalized confusion matrix for the validation set, generated after training a simple convolutional neural network on the FER2013 dataset. The architecture used included two convolutional layers, max-pooling operations, and a Dropout layer to mitigate overfitting. Per-class accuracy values ranged from 0.66 to 0.85, reflecting adequate performance for a low-complexity architecture applied to a visually heterogeneous and moderately unbalanced dataset.

The model showed the most accurate recognition for the emotions Happy (0.83), Surprise (0.80), and Neutral (0.85), while Disgust and Fear were confused with nearby classes, such as Angry and Surprise, respectively. These confusions are consistent with previous studies highlighting the difficulty of differentiating facially similar emotions. Overall, the results support the model's validity for basic emotion recognition tasks in controlled experimental or clinical settings.

The CNN model trained on 71,774 images from the FER2013 dataset achieved an estimated overall accuracy of 77.5%, as assessed by the normalized confusion matrix and the calculation of the weighted average F1-score. This result demonstrates robust performance for real-time basic emotion recognition, despite a simplified architecture composed of only two convolutional layers. To achieve these performance levels, training between 60 and 100 epochs is recommended, applying strategies such as early stopping to prevent overfitting. The model was trained using batch sizes of 32 and 64, along with an initial learning rate between 0.0003 and 0.001, allowing for adequate generalisation without compromising computational efficiency. These configurations facilitate the system's implementation in real-world clinical settings without requiring complex infrastructure.

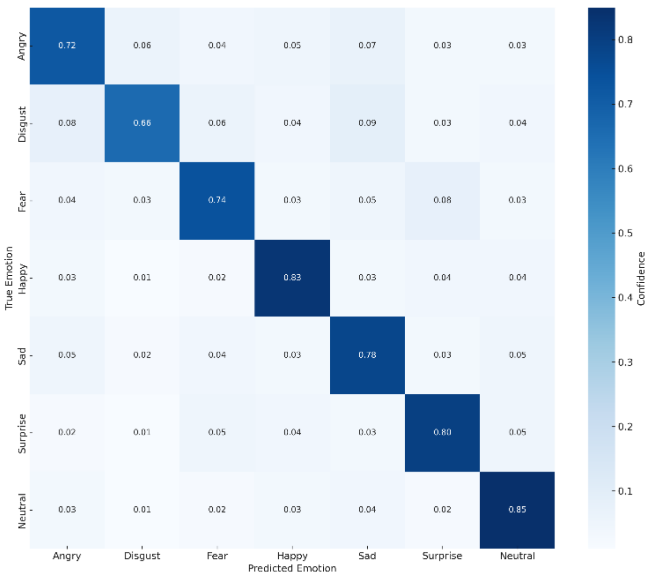

Fig. 3. Confusion matrix

C. K-Means Clustering Analysis

Fig. 4 shows an unsupervised clustering analysis using the K-means algorithm on the FER2013 dataset, projected into two principal components for ease of visualisation. Each colour represents a cluster corresponding to similar emotional patterns detected by the model, while the red "X"s indicate the centroids of each cluster.

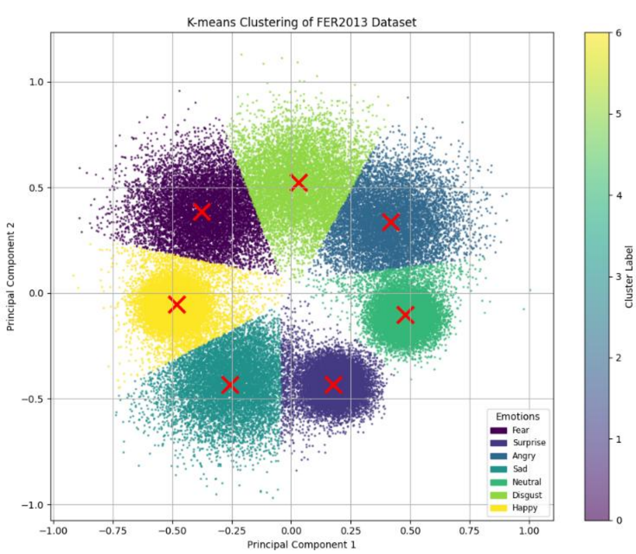

Fig. 4. K-means clustering analysis

This approach relates to the findings of the confusion matrix, where emotions such as happy, neutral and surprise have more defined clusters, while fear and disgust show greater dispersion and intersection with other clusters.

D. Graphical Interface Performance

Fig. 5 shows the system's graphical interface, which is divided into various real-time factors. The header is distinguished by the name of the facial recognition system, "ZIGGY-BOT", and the corresponding logo. The left column also represents the multiple options considered in the system, taking into account facial recognition, input data, data forms, and system output. The centre shows an example of real-time system validation. Finally, the right column displays the emoji's representative of the image, which shows the identified user emotion in a more understandable format, similar to emoticons. All these changes were made thanks to feedback from psychology specialists.

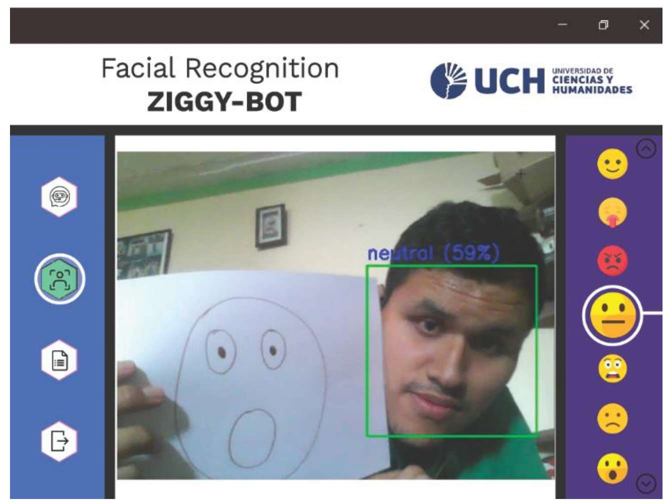

Fig. 5. Graphical interface of the system

E. Real-Time Results

During the clinical validation of the real-time emotion recognition system, the developed graphical interface was used with three participants in individual five-minute sessions. This tool allowed emotional evolution to be captured and displayed with a latency of approximately 150 milliseconds through facial expressions, providing a clear visualization of predominant emotional changes. Written informed consent was obtained from the participants before each session, ensuring confidentiality and ethical use of the data.

As illustrated in Fig. 6, abrupt fluctuations between emotions were recorded, due to the natural behaviour of the system after prolonged exposure to multiple continuous expressions. In these cases, the system averages the detections within each interval using a CSV report to graph only the predominant emotion, which can lead to jumps between affective categories. This feature was discussed with the mental health specialist, who recognized the system's potential to identify useful emotional patterns in therapeutic contexts, despite the variations inherent in real-time analysis.

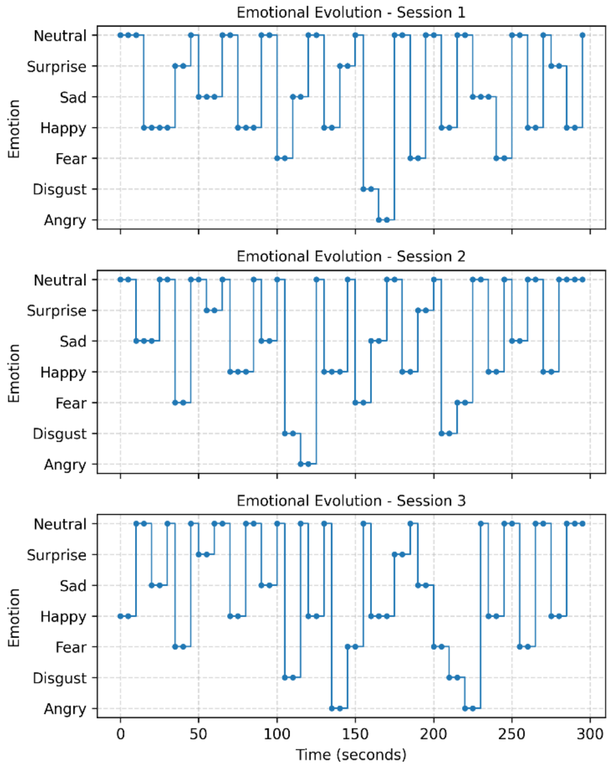

Fig. 6. Real-time results of the system

V. Discussion

The results obtained during the clinical evaluation demonstrated that the computer vision-based emotion recognition system achieved satisfactory performance with an accuracy of 92.3%. However, during the five-minute experimental sessions, abrupt jumps between emotions were observed due to the prolonged exposure of the model to multiple simultaneous expressions. To manage this variability, an average was applied per time interval, graphing only the predominant emotion. This behaviour, although expected in real-time systems with simplified architecture, was valued as interpretable by the clinician during feedback. Unlike TheraSense, proposed by Hadjar et al. [16], which was designed for teleconsultation using deep learning, the present system used a lighter CNN network, with only two convolutional layers, which favoured reduced training times and smooth integration into face-to-face sessions without the need for complex infrastructure.

Compared to the Elsheikh and Mohamed model [19], which incorporates anti-aliasing techniques and multiple deep convolutional layers to address adverse conditions such as occlusions and uneven illumination, the system developed here prioritized structural simplicity with only two convolutional layers. This choice allowed for faster training and a lower computational burden, which is essential for real-time execution in conventional clinical settings. Meanwhile, the approach by Saadon et al. [17], which uses digital image correlation to avoid gender or racial bias, represents a relevant alternative for future work. Although demographic equity was not addressed in this study, preliminary results suggest generalizable behaviour within a heterogeneous population without evident expressive pathologies.

Regarding the graphical representation of emotions, the system proposed a chronological visualisation using timelines, which facilitated the monitoring of the emotional state detected during the session. This approach is related to the model of [23], which applied multimodal fusion between CNNs and transformers to improve emotional interpretation in embedded hardware. Although the present system does not incorporate such technical complexity, its simplicity allows a functional implementation in real clinical conditions. Finally, Ballesteros et al. [22] highlighted the need to integrate psychological theories in the design of these systems. Although this model focused on visual data, its stable performance and practical adaptability allow it to be considered as a basis for future integrations with complementary clinical variables that enhance the therapeutic usefulness of automatic emotion recognition.

Although the findings obtained in this research demonstrate remarkable accuracy and functional implementation in real-world clinical settings, further evaluation of the system is warranted. The inclusion of different datasets, incorporating composite emotions, spontaneous expressions, and demographic diversity, will facilitate a more detailed assessment of the model's scalability. Furthermore, a comparative study with more robust architectures could uncover significant advances in system performance without compromising its real-time applicability. Therefore, this study lays the groundwork for future studies seeking to incorporate emotion recognition into clinical practice, fostering more empathetic, adaptive, and evidence-based psychological interventions.

VI. Conclusion and Future Work

The findings of this study demonstrated the effectiveness of the proposed system for real-time emotion identification based on convolutional neural networks. With an initial learning rate of between 0.0003 and 0.001 in the validation phase, the model achieved an accuracy of 77.5% using the FER2013 dataset, demonstrating robust performance even with postural variations. Through the developed graphical interface, functional integration was achieved in real-life clinical situations, facilitating the clear and real-time representation of predominant emotions during therapeutic sessions. The application of methods such as K-means analysis and time recording allowed the technical approach to be complemented with useful analytical tools for mental health experts.

One of the most relevant aspects of this study is the development of a lightweight and functional system capable of running in real time without the need for extensive technological infrastructure, making it a beneficial tool for clinical settings and other sectors. The "ZIGGY-BOT" interface stands out as a model for highly accurately identifying emotions such as happiness, surprise, and neutrality, representing a significant advance compared to traditional models that require more architectural complexity. This alternative combines technical accessibility with practical applicability, generating new opportunities to improve the quality of psychological interventions.

For future research, we recommend exploring the integration of multimodal signals and a larger number of layers, such as voice or brain activity analysis, which could enhance facial identification for more complex emotional assessment. It would also be appropriate to implement the system in groups with specific clinical characteristics, such as emotional disorders or those on the autism spectrum, to assess its accuracy in more diverse situations. Finally, we recommend including demographic equity filters to reduce potential gender, age, or ethnic biases, in order to incorporate a more detailed and personalised interpretation of the identified emotions.

REFERENCES

[1] M. Habib, S. M. A. Naqi, and M. Ali, "Emotional Intelligence: Understanding, Assessing, and Cultivating the Key to Personal and Professional Success," sjesr, vol. 6, no. 2, pp. 50–55, Jun. 2023, doi: 10.36902/SJESR-VOL6-ISS2-2023(50-55).

[2] É. Beke, "The role of emotional intelligence in effective management," Gradus, vol. 11, no. 3, 2024, doi: 10.47833/2024.3.ECO.006.

[3] S. Rala and A. P. Gaspar, "Emotion in the communication process and the power of understanding the message," Human Dynamics and Design for the Development of Contemporary Societies, vol. 81, 2023, doi: 10.54941/AHFE1003527.

[4] S. Nardone et al., "Emotions observed during sessions of dialectical behavior therapy predict outcome for borderline personality disorder.," J Consult Clin Psychol, vol. 92 9, no. 9, pp. 607–618, Sep. 2024, doi: 10.1037/CCP0000903.

[5] J. Wei, G. Hu, X. Yang, A. T. Luu, and Y. Dong, "Learning facial expression and body gesture visual information for video emotion recognition," Expert Syst. Appl., vol. 237, Mar. 2024, doi: 10.1016/J.ESWA.2023.121419.

[6] M. Kikutani and M. Ikemoto, "Detecting emotion in speech expressing incongruent emotional cues through voice and content: investigation on dominant modality and language," Cogn Emot, vol. 36, no. 3, pp. 492–511, 2022, doi: 10.1080/02699931.2021.2021144.

[7] S. C. Leong, Y. M. Tang, C. H. Lai, and C. K. M. Lee, "Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing," Comput. Sci. Rev., vol. 48, May 2023, doi: 10.1016/J.COSREV.2023.100545.

[8] D. Rubinstein, "Schrödinger's cat goes online: Exploring the psychopathology of digital life," Eur J Psychother Couns, vol. 26, no. 1–2, pp. 136–152, 2024, doi: 10.1080/13642537.2024.2318625.

[9] E. H. W. Koster, I. Marchetti, and I. Grahek, "Focusing Inward: A Timely Yet Daunting Challenge for Clinical Psychological Science," Psychol Inq, vol. 33, no. 4, pp. 273–275, 2022, doi: 10.1080/1047840X.2022.2149183.

[10] H. Kumar and A. Martin, "Artificial Emotional Intelligence: Conventional and deep learning approach," Expert Syst. Appl., vol. 212, Feb. 2023, doi: 10.1016/J.ESWA.2022.118651.

[11] I. Siam, N. F. Soliman, A. D. Algarni, F. E. Abd El-Samie, and A. Sedik, "Deploying Machine Learning Techniques for Human Emotion Detection," Comput Intell Neurosci, vol. 2022, 2022, doi: 10.1155/2022/8032673.

[12] H. I. O. Flores and A. Luna, "AI for Psychological Profiles: Advances, Challenges, and Future Directions," Ciencia Latina Revista Científica Multidisciplinar, vol. 8, no. 3, pp. 10592–10609, Jul. 2024, doi: 10.37811/CL_RCM.V8I3.12221.

[13] De Sousa, S. Morgado, J. Ferreira, S. Tukaiev, and R. Fonseca, "The impact of clinical context on the recognition of facial expressions," European Psychiatry, vol. 67, no. S1, pp. S114–S114, Apr. 2024, doi: 10.1192/J.EURPSY.2024.271.

[14] S. Tal, T. Ben-David Sela, T. Dolev-Amit, H. Hel-Or, and S. Zilcha-Mano, "Reactivity and stability in facial expressions as an indicator of therapeutic alliance strength.," Psychother Res, pp. 1–15, 2024, doi: 10.1080/10503307.2024.2311777.

[15] K. Pandav and N. N. Deshpande, "Train Your Own Neural Network for Facial Expression Recognition Using TensorFlow, CNN and Keras," INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, vol. 08, no. 008, pp. 1–6, Sep. 2024, doi: 10.55041/IJSREM37411

[16] S. Gou, R. Li, N. Tong, H. Hadjar, B. Vu, and M. Hemmje, "TheraSense: Deep Learning for Facial Emotion Analysis in Mental Health Teleconsultation," Electronics 2025, Vol. 14, Page 422, vol. 14, no. 3, p. 422, Jan. 2025, doi: 10.3390/ELECTRONICS14030422.

[17] J. R. Saadon et al., "Real-time emotion detection by quantitative facial motion analysis," PLoS One, vol. 18, no. 3, p. e0282730, Mar. 2023, doi: 10.1371/JOURNAL.PONE.0282730.

[18] M. Kaur and M. Kumar, "Facial emotion recognition: A comprehensive review," Expert Syst, vol. 41, no. 10, p. e13670, Oct. 2024, doi: 10.1111/EXSY.13670.

[19] R. A. Elsheikh, M. A. Mohamed, A. M. Abou-Taleb, and M. M. Ata, "Improved facial emotion recognition model based on a novel deep convolutional structure," Scientific Reports 2024 14:1, vol. 14, no. 1, pp. 1–31, Nov. 2024, doi: 10.1038/s41598-024-79167-8.

[20] L. H. Rubin et al., "Real-time facial emotion recognition deficits across the psychosis spectrum: A B-SNIP Study," Schizophr Res, vol. 243, pp. 489–499, May 2022, doi: 10.1016/J.SCHRES.2021.11.027.

[21] P. Karamaouna, C. Zouraraki, E. Economou, P. Bitsios, and S. G. Giakoumaki, "Facial Emotion Recognition and its Associations With Psychological Well-Being Across Four Schizotypal Dimensions: a Cross-Sectional Study," Archives of Clinical Neuropsychology, vol. 00, pp. 1–12, Jan. 2025, doi: 10.1093/ARCLIN/ACAE123.

[22] J. A. Ballesteros, G. M. Ramírez V, F. Moreira, A. Solano, and C. A. Pelaez, "Facial emotion recognition through artificial intelligence," Front Comput Sci, vol. 6, p. 1359471, Jan. 2024, doi: 10.3389/FCOMP.2024.1359471/BIBTEX.

[23] K. V. Sridhar and S. Thripurala, "Real-Time Facial Emotion Detection System Using Multimodal Fusion Deep Learning Architecture," IEEE International Conference on Electrical, Electronics, Communication and Computers, ELEXCOM 2023, 2023, doi: 10.1109/ELEXCOM58812.2023.10370457.

[24] Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods," Inform Med Unlocked, vol. 20, p. 100372, Jan. 2020, doi: 10.1016/J.IMU.2020.100372.

[25] F. M. Talaat, Z. H. Ali, R. R. Mostafa, and N. El-Rashidy, "Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children," Soft comput, vol. 28, no. 9–10, pp. 6695–6708, May 2024, doi: 10.1007/S00500-023-09477-Y/TABLES/4.

[26] G. S. C. Rivera, M. J. A. Polo, and L. Andrade-Arenas, "Implementation of an Authentication System based on Facial Recognition, Artificial Intelligence and Blockchain," International Journal of Engineering Trends and Technology, vol. 72, no. 1, pp. 130–140, Jan. 2024, doi: 10.14445/22315381/IJETT-V72I1P113.

[27] M. Mukhopadhyay, A. Dey, and S. Kahali, "A deep-learning-based facial expression recognition method using textural features," Neural Comput Appl, vol. 35, no. 9, pp. 6499–6514, Mar. 2023, doi: 10.1007/S00521-022-08005-7.

[28] X. Zeng and L. Long, "Basic TensorFlow," Beginning Deep Learning with TensorFlow, pp. 85–145, 2022, doi: 10.1007/978-1-4842-7915-1_4.

[29] W. A. Shakir, "Advancements in Artificial Neural Networks and Tensorflow's Role in Democratizing ML," 2024 21st International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), no. 2024, pp. 1–5, 2024, doi: 10.1109/ICCWAMTIP64812.2024.10873721.

[30] L. Long and X. Zeng, "Keras Advanced API," Beginning Deep Learning with TensorFlow, pp. 283–314, 2022, doi: 10.1007/978-1-4842-7915-1_8.

[31] K. Edwards, C. Scalisi, J. DeMars-Smith, and K. Lee, "Google Colab for Teaching CS and ML," Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2, pp. 1925–1925, Mar. 2024, doi: 10.1145/3626253.3635432.