DESDE 2013
https://repository.uaeh.edu.mx/revistas/index.php/icbi/issue/archive
Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI
ISSN: 2007-6363

Publicación Semestral Pädi Vol. 14 No. Especial (2026) 173-178

Asistente de inteligencia artificial por voz para apoyar la salud mental
Artificial intelligence assistant for voice to support at the mental health

D. Lopez-Salazar [a], M. Hernández-Chávez [b], J. D. Rivera-Fernández [b], K. Roa-Tort [b], D.A. Fabila-Bustos [b]

[a] *Escuela Superior Tlahuelilpan, Universidad Autónoma del Estado de Hidalgo, 42780, Hidalgo, México.*
[b] *Laboratorio de Optomecatrónica y Energías, UPIIH, Instituto Politécnico Nacional, Distrito de Educación, Salud, Ciencia, Tecnología e Innovación, San Agustín Tlaxiaca, 42162, Hidalgo, México.*

Resumen

Las alteraciones emocionales pueden impactar de manera significativa en el bienestar de las personas y muchas veces pasan desapercibidas o no se gestionan adecuadamente. Frente a esta necesidad, se ha buscado desarrollar tecnologías que permitan reconocer y responder a las emociones humanas en tiempo real. En este trabajo se presenta el diseño de un asistente inteligente que utiliza como base redes neuronales convolucionales y el modelo Local Binary Patterns Histograms (LBPH) entrenados con base de datos de imágenes faciales para identificar emociones básicas como enojo, felicidad, miedo y tristeza. El sistema capta los gestos faciales del usuario mediante una cámara y procesa las imágenes para clasificar la emoción detectada. Una vez identificada, el asistente emite una respuesta hablada, muestra visualmente la emoción reconocida y ofrece una recomendación emocional personalizada. Para enriquecer la interacción, se integra también un chatbot conversacional que permite al usuario dialogar en lenguaje natural con el sistema, fomentando así una experiencia más cercana y empática.

Palabras Clave: Inteligencia artificial, redes neuronales artificiales, asistente de emociones, modelo LBPH.

Abstract

Emotional disturbances can have a significant impact on individuals' well-being and often go unnoticed or are not properly addressed. To meet this need, technological development has aimed to recognize and respond to human emotions in real time. This work presents the development of an intelligent assistant that employs convolutional neuronal networks and the LBPH model, trained with facial image databases to identify basic emotions such as sadness, happiness, fear, and anger. The system captures the user's facial expression through a camera and processes it to classify the detected emotion. Once identified, the assistant provides a spoken response, visually displays the recognized emotion, and offers a personalized emotional recommendation. To enhance the user experience, a conversational chatbot is also integrated, allowing interaction in a natural language and promoting closer, more empathetic engagement.

Keywords: Artificial intelligence, artificial neuronal network, emotional assistant, LBPH model.

1. Introducción

According to the World Health Organization, adolescence, between the ages of 10 and 19, is a stage vulnerable to emotional health problems that can lead to disorders such as anxiety and stress. In addition, it is reported that 1 in 7 adolescents suffer from mental disorder at this stage (World Health Organization: WHO 2024). These problems often originate from emotional imbalances influenced by interpersonal relationships both at home and in the social environment, and even by the media. Recent studies suggest that mental health disorders appear to intensify with each generation; additionally, lack of education about emotional regulation in adolescents can lead to serious health risk factors, such as eating disorders, depression and even suicide in extreme cases (Nguyen et al., 2025). Since the COVID-19 pandemic declared in early 2020, significant negative effects have further impacted people's emotional well-being worldwide. The rapid spread of the virus, lockdown measures, social isolation, economic uncertainty, and misinformation have intensified stress, anxiety and depression and other emotional difficulties across populations (De la Torre & Gutiérrez, 2022).

Over the past decade, the application of artificial intelligence (AI) in various fields such as industry (Malik et al., 2024), environment (Konya & Nematzadeh, 2023),

*Autor para la correspondencia: dfabilab@ipn.mx
Correo electrónico: dfabilab@ipn.mx (Deyanira Lopez-Salazar), mhernandezch@ipn.mx (Macaria Hernández-Chávez), jdriveraf@ipn.mx (Josué Daniel Rivera-Fernández), kroat@ipn.mx (Karen Roa-Tort), dfabilab@ipn.mx (Diego Adrián Fabila-Bustos).

Historial del manuscrito: recibido el 15/07/2025, última versión-revisada recibida el 05/01/2026, aceptado el 12/02/2026, publicado el 25/03/2026. **DOI:** https://doi.org/10.29057/icbi.v14iEspecial.15511

education (Ogunleye et al., 2024) and medical (Kuwaiti et al., 2023) has had an exponential growth that has benefited society.

Additionally, AI has driven the development of voice assistant or chatbots, which are considered conversational agents that act to replicate human interaction through text, voice and visual forms of communication (Singh et al., 2023) and which usually have important applications in the management of daily routines, activities, and meetings. However, recent studies have demonstrated their possible application as a tool for the emotional and mental well-being of the population. An example of this data is back to Eliza, considered the first voice assistant, designed to simulate therapy by repeating users' responses as questions that encouraged conversation (Weizenbaum, 1966), to the voice assistant that are currently used every day by society, such as Siri, Alexa and Google Assistant. Recently, studies have been conducted on the behavior of this assistant to user questions about the safety and use of vaccines (Alagha & Helbing, 2019) and as a clinical counseling tool for postpartum depression (Yang et al., 2020).

On the other hand, ChatGPT has been evaluated to perform a timely intervention in mental health problems (Liu et al., 2024). Hume AI, which uses voice recognition to detect emotions through conversations (Berrebi, 2024). However, at present several of these tools are under evaluation and constant improvement.

In the field of artificial intelligence applied to emotion recognition, several models have been developed for the analysis of facial expressions. Classical methods such as Local Binary Patterns Histograms (LBPH) have been widely adopted due to their robustness to illumination changes and low computational cost, making them suitable for real-time and resource-constrained environments (Anohen et al., 2006; Shan et al., 2009). Recent studies confirm that LBPH remains a practical option for educational and applied research, achieving competitive performance in facial expression recognition. Based on this foundation, our work integrates LBPH with voice, facial expressions, and text analysis to propose a multimodal emotion recognition assistant.

Therefore, this work seeks to develop an assistant that detects emotions through the combination of three elements: Voice recognition, text analysis, and artificial vision, while the user interacts with the artificial intelligence system

2. Materials and Methods

2.1. Methodology

Figure 1 illustrates the development process of the proposed software system, which integrates an emotion recognition model and a chatbot interface.

The process is structured into five main phases: requirement analysis, system design, implementation and integration, and maintenance. Initially, the system requirements are defined, including the classification of emotions and creation of the dataset. This is followed by the design of the graphical user interface (GUI). In the implementation phase, the software components are developed, including the emotion recognition model, chatbot model, and GUI. The subsequent phase involves verifying the functionality and performance of each module through metrics evaluation and testing. Finally, the system enters a maintenance stage where future updates and improvements are planned to ensure the software's continuous evolution and stability.

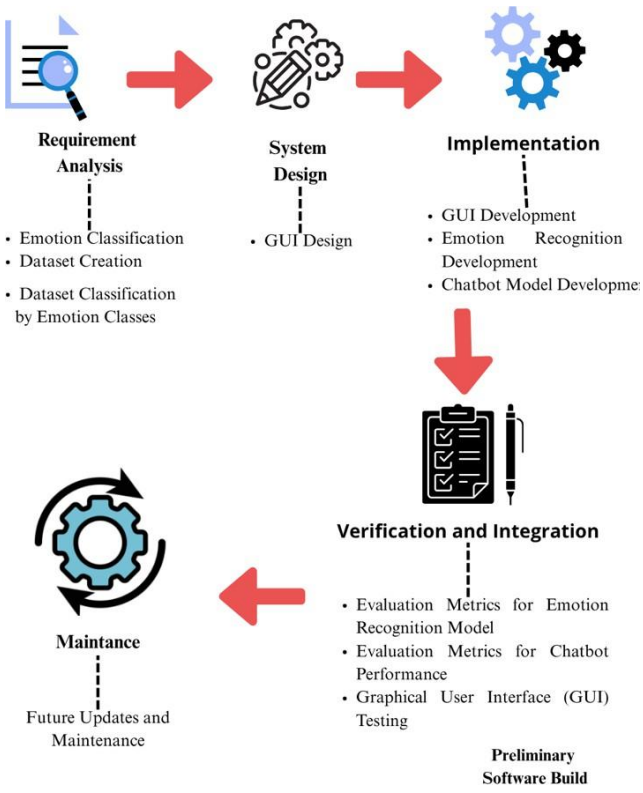

Figure 1: Methodology overview.

To verify the correct operation of the intelligent assistant, we developed the following tests. In Figure 2, a flow diagram of the general operation of the intelligent assistant is shown in the implementation phase:

- Unit: These tests consist of individually evaluating the software's functions and modules. In the case of assistant, the correct operation of each module was verified, from the startup module, the user interaction understanding module, to the information research and emotion recognition module.
- Integration: This type of test verifies that the different modules function correctly when integrated by combining the three modules. The developed assistant does not present errors during interaction between them.
- Functional: These tests focus on compliance with project requirements. They confirm that the developed project adequately meets the established requirements.
- Regressions: Illustrating the application of this test, by integrating the emotion recognition module with the rest of modules, it was verified that all modules maintain their functionality without errors

2.2. Emotion Recognition data set

A custom dataset was created, comprising 30 different individuals as reference for constructing four classes for emotion classification: sadness, happiness, fear and anger. From each reference image, 50 captures were taken, corresponding to different individuals between 20-30 years

old, male and female resulting in total of 6000 images. The pictures were taken with a white background and proper light exposure to ensure consistency for the emotion recognition model. All images were adapted to the facial region using Haar cascade classifier and then resized to 224 x 224 pixels in RGB format, which served as the standardized input dimension for both the LBPH and CNN models. As the dataset was generated exclusively for this study, it is not publicly available.

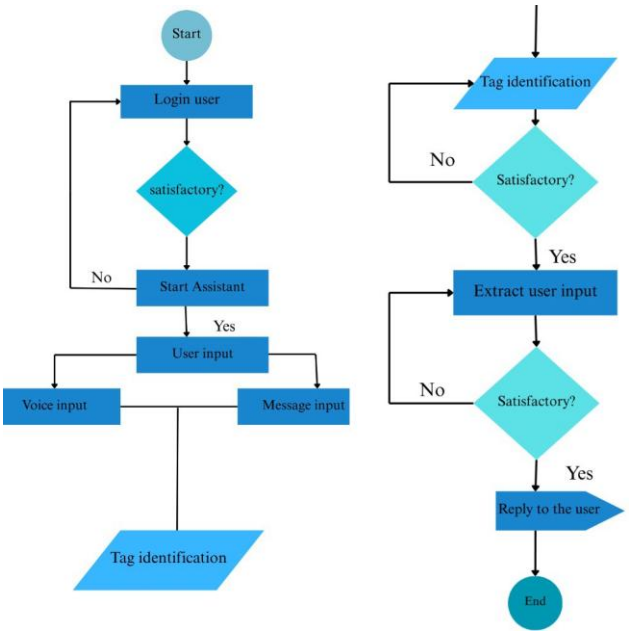

Figure 2: Flow diagram for general operation of the intelligent system.

LBPH is a classical method that encodes local pixels neighborhoods into histograms, providing a particularly under varying illumination. The model of emotion recognition was trained in two ways: first with the LBPH model combined with Haar cascade classifiers to detect facial regions, requiring only 9 seconds of training for the four emotion classes. Second, with a convolutional neural network (CNN) built using TensorFlow's Sequential API (version 2.13.0). The CNN architecture consists of three convolutional layers with ReLU activations and max pooling, followed by a flattening layer, a dropout of 50% and two fully connected dense layers, the final using the softmax activation function to perform multiclass classification. The model was compiled with the Adam optimizer (Kingma & Ba, 2025) and the sparse categorical cross-entropy loss function, trained. for 30 epochs with a batch size of 32 and evaluated on the test data. Prediction was generated on the test set and performance was analyzed using classification report and confusion matrix from scikit-learn library (version 1.3.2). Finally, a heatmap of the confusion matrix was visualized with Seaborn. For both models, 80% of the dataset was used for training and 20% for testing. In the CNN, hidden convolutional layers extracted hierarchical facial features, max pooling reduced dimensionality, and dense layers integrated these representations before, the softmax layer produced the final emotion class

For analytical support, the evaluation of metrics and methods employed are described as follows: The LBPH algorithm operates by comparing each pixel with neighbors within a defined radius, encoding these relationships into a binary pattern and constructing histograms of local features across image regions. These histograms are concatenated to form a global descriptor for classification, which makes LBPH computationally efficient and robust under varying illumination.

In the CNN, the Rectified Linear Unit (ReLU) activation, was used to improve training efficiency and avoid vanishing gradients. Model performance was assessed with accuracy, precision, indicates the proportion of correctly classified positive instances among all predicted positives and the F1-score; represents the harmonic mean of precision and recall, offering a balanced measure of the model's effectiveness.

Additionally, confusion matrices generated for both LBPH and CNN, providing a clearer view of correct and incorrect classifications across all emotion classes, complementing the numerical metrics with a visual analysis of model behavior.

2.3. Chatbot development

The chatbot was developed and trained using neural network based on an intent classification model, employing Natural Language Processing (NLP) techniques and the Keras library (version 2.13.1), Keras is an open-source deep learning library initially developed by Francois Chollet. A .json file was created to define a set of user input patterns along with their corresponding intent tags. In total, the dataset contained 10 intent categories, with 50 training patterns (five per intent) and 30 predefined responses (three per intent). During preprocessing, each pattern was tokenized and lemmatized, generating a unique lemmatized vocabulary and a distinct set of intent classes. Each pattern was then transformed into a binary vector using the Bag of Words model, indicating the presence of known vocabulary terms. In this case, the final vocabulary consisted of 68 unique words, resulting in input vectors of 10 dimensions. Each intent tag was encoded into a one-hot vector of 10 dimensions, corresponding to the total number of emotional categories defined in the .json file.

To prepare the training data, the dataset was shuffled and converted into NumPy arrays, train_x and train_y. The train_x array contains the binary feature vectors, while train_y contains the one-hot encoded intent labels.

The neural network architecture consists of a sequential model with an input layer of 128 neurons using ReLU activation function, followed by dropout layers regularization, and an output layer with a softmax activation function for multiclass classification. The model was compiled using the Stochastic Gradient Descent (SGD) optimizer and trained over 200 epochs with a batch size of 5. After training, the model was saved as chatbot_model.h5. Additionally, the vocabulary and intent tags were stored using the pickle library to ensure consistent preprocessing during future predictions.

2.4. Voice recognition for the chatbot

To address the assistant's voice recognition functionality, an API was integrated to convert audio into text using speech recognition (Speech Recognition, 2025) technology (version 3.10.4), developed by Francois Chollet. This allows the assistant to effectively support voice-based user interaction. Additionally, the assistant can maintain a conversation through

voice and deliver recommendations verbally, providing a more natural and accessible user experience.

2.5. Personal computer Specifications (PC)

In the development of any software, it is necessary to consider the specifications of the computer equipment used, in this sense in Table 1 we summarize the PC specifications used for the assistant development.

Table 1: Personal Computer Specifications (PC)

Processor	AMD Ryzen 7 8840HSw
RAM	16.0 GB (15.3 GB usable)
System	64-bit operating system, x64-based processor

3. Results and Discussion

The algorithm was trained in two ways: the LBPH model and CNN. To achieve this, it used a dataset with facial images of the people it wanted to recognize. Each image must be assigned an identifier (in this case, called a "face") so that the algorithm can use this information to recognize each input image and return the expected result. Figure 3 show examples of the four emotion classes used in the dataset: A) anger, B) happiness, C) sadness and D) fear. Each class was captured in separate sessions, where the target emotion was selected in advance and stored in its corresponding folder, ensuring consistent labeling during acquisition.

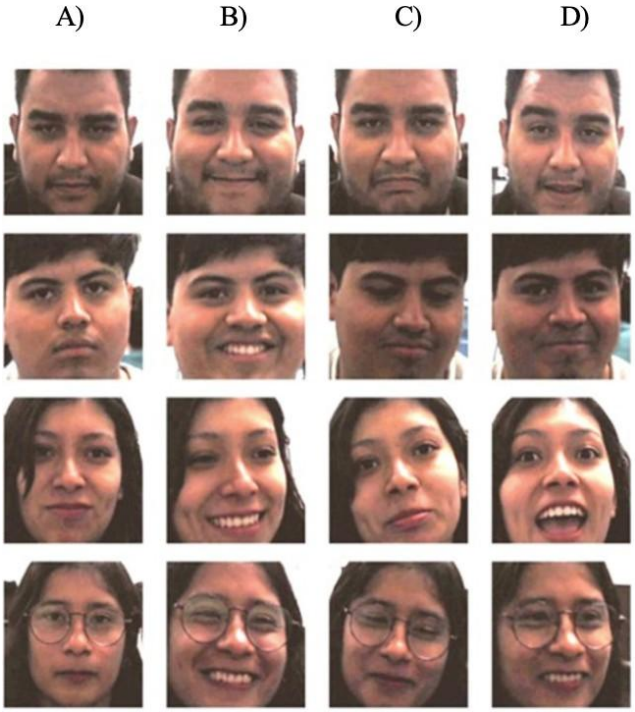

Figure 3: Emotion class: A) anger, B) happiness, C) sadness, D) fear.

During the training phase for emotion recognition, a comparison was conducted between the LBPH model and CNN. As summarized in Table 2, the LBPH model demonstrated superior performance across all metrics, achieving a global accuracy of 0.87. This table is particularly relevant because it shows that LBPH not only reached performance in precision, recall and F1-score across the different emotional categories.

Table 2 provides the empirical basis for selecting LBPH as the optimal model in this preliminary stage.

Table 2 Evaluation metrics: LBPH model emotion classification.

	Precision	Recall	F1-Score
Anger	0.86	0.85	0.85
Happiness	0.88	0.90	0.89
Fear	0.85	0.86	0.85
Sadness	0.87	0.86	0.86
accuracy			0.87
Macro avg	0.87	0.87	0.86
Weight avg	0.87	0.87	0.87

Nevertheless, despite having obtained an optimal model with LBPH model, the classification of emotions was also evaluated using a convolutional neuronal network (CNN). This comparison was performed using the same metrics as for the LBPH model. The results are presented in Table 3.

Table 3: Evaluation metrics of CNN model emotion recognition

	Precision	Recall	F1-Score
Anger	0.78	0.80	0.79
Happiness	0.83	0.85	0.84
Fear	0.79	0.77	0.78
Sadness	0.80	0.78	0.79
accuracy			0.80
Macro avg	0.80	0.80	0.80
Weight	0.80	0.80	0.80

To analyze the confusion between classes, a confusion matrix was applied to each class. The results of the LBPH model are shown in Figure 4, while the results obtained with the CNN are presented in Figure 5.

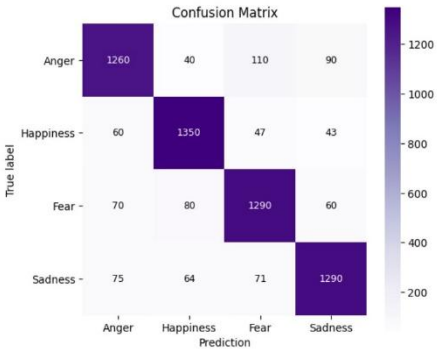

Figure 4: LBPH model emotion classification, confusion matrix.

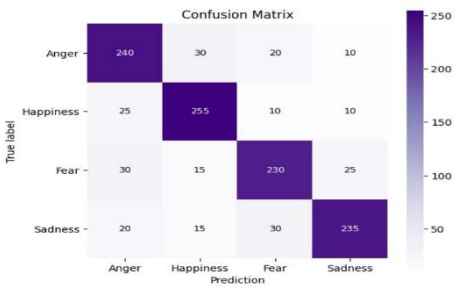

Figure 5: Confusion matrix of the CNN emotion classification model.

For the chatbot design, the trained model yielded the following results, considering each tag based on the four emotional classes shows in Figure 6.

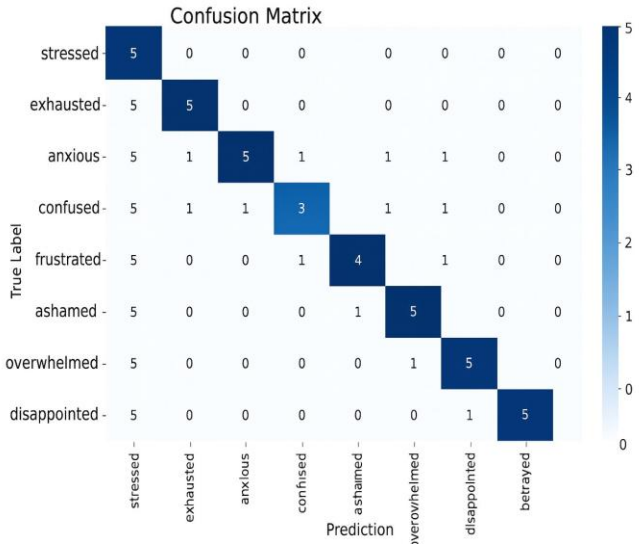

Figure 6: Confusion matrix chatbot emotion classes

Althought both LBPH and CNN achieved high accuracies in this study (87% and 80%, respectively), these results should be interpreted considering the limited size of the dataset used at this preliminary stage. For comparison, (Akhand, Roy, Siddique, Kamal, & Shimamura, 2021), reported an accuracy of around 92% using a deep CNN with transfer learning, which illustrates that even advanced deep-learning methods rarely reach perfect performance in more complex datasets. Therefore, while the obtained results are promising for validating the proof of concept, they should be interpreted with caution.

Future work will address this limitation by expanding the dataset and conducting validation with more diverse population, enabling a more realistic estimation of the system's generalization capacity, but not all people wants to participate in the capture of their faces.

Finally, once all the previously developed functionalities were integrated, a graphical user interface was implemented using Streamlit (version 1.38.0, Snowflake Inc). Figure 7 shows the assistant's interface during an interaction with a user.

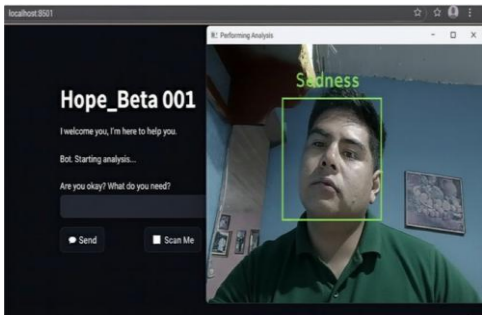

Figure: 7 Assistant interface.

The graphical user interface (GUI) allows interaction either through speech or text. A dedicated analysis button enables real-time facial expression recognition, while text and voice inputs are processed in parallel. The system then outputs one of the four emotions and provides a personalized recommendation according to the user's current emotional state, which is displayed directly on the interface.

Some artificial intelligence systems use voice recognition or detect emotions through textual conversations. In contrast, the present work integrates computer vision for facial emotion detection together with voice recognition, allowing the system not only to recognize emotions but also to provide recommendations aimed at supporting users in regulating them. Unlike many studies that rely on pre-existing datasets for training emotion recognition algorithms, this project developed a custom image database tailored to the application. Furthermore, while most prior research has focused on analyzing the interaction between users and commercial assistants such as Alexa (Yan, Johnson, & Jones, 2024), our approach emphasizes the design of an assistant capable of real-time multimodal emotion recognition. Similarly, although some works provide recommendations based on emotional states (Ghatak, Paul, & Ghosh, 2023), the distinguishing contribution of this project is the combination of computer vision, voice recognition, and pattern analysis to enable real-time, adaptive responses.

4. Conclusions

The assistant demonstrated an effective ability to detect emotions from facial expressions, achieving 87% accuracy with the LBPH model, with the CNN achieving 80% accuracy, dataset, though these results should be interpreted cautiously due to the limited dataset size. At the conversational level, the chatbot achieved 98% accuracy when responding to user queries, although this performance is influenced by the small json dataset used for training.

The system integrates facial expression recognition with voice-based emotion recognition. Future improvements will focus on expanding both the image and text datasets to strengthen robustness and generalization. By combining computer vision and voice recognition, the assistant provides a more comprehensive emotional understanding, reducing the subjectivity of visual analysis and contributing to more precise and responsive interactions

Acknowledgments

This work was supported by a grant from the Instituto Politécnico Nacional (IPN), SIP-20254769, awarded to Diego Adrián Fabila Bustos

References

Alagha, E. C., & Helbing, R. R. (2019). Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health & Care Informatics, 26(1), e100075. https://doi.org/10.1136/bmjhci-2019-100075

Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using Transfer Learning in Deep CNN. Electronics, 10(9), 1036. doi:10.3390/electronics10091036

Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12), 2037–2041. doi:10.1109/TPAMI.2006.244

Berrebi, M. (2024, 3 mayo). L'IA empathique : explorer la reconnaissance vocale émotionnelle de Hume AI. Unlock AI Tools & Expert Guides | Tips

& Tricks For AI Enthusiasts. https://pwraitools.com/generative-ai-tools/empathic-ai-and- emotional-voice-recognition-with-hume-ai/

De la Torre, J. C. P., & Gutiérrez, C. L. C. (2022). Estados emocionales agudos en pobladores mexicanos durante la pandemia de COVID-19. Revista de Psicología, 41(1), 377-400. https://doi.org/10.18800/psico.202301.014

Ghatak, S., Paul, H., & Ghosh, D. (2023). Voicebot For Mental Disease Prediction And Treatment Recommendation Using MachineLearning.TechRxiv.https://doi.org/10.36227/techrxiv.258239.v1

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1412.6980

Konya, A., & Nematzadeh, P. (2023). Recent applications of AI to environmental disciplines: A review. The Science Of The Total Environment, 906, 167705. https://doi.org/10.1016/j.scitotenv.2023.167705

Kuwaiti, A. A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., Muhanna, D. A., & Al-Muhanna, F. A. (2023). A Review of the Role of Artificial Intelligence in Healthcare. Journal Of Personalized Medicine, 13(6), 951. https://doi.org/10.3390/jpm13060951

Liu, Y., Ding, X., Peng, S., & Zhang, C. (2024). Leveraging ChatGPT to optimize depression intervention through explainable deep learning. Frontiers In Psychiatry,15. https://doi.org/10.3389/fpsyt.2024.1383648

Malik, S., Muhammad, K., & Waheed, Y. (2024). Artificial intelligence and industrial applications-A revolution in modern industries. Ain Shams Engineering Journal, 15(9), 102886. https://doi.org/10.1016/j.asej.2024.102886.

Nguyen, A., Grummitt, L., Barrett, E. L., Bailey, S., Gardner, L. A., Champion, K. E., … Birrell, L. (2025). The relationship between emotion regulation and mental health in adolescents: Self-compassion as a moderator. Mental Health & Prevention, 38(200430), 200430. doi:10.1016/j.mhp.2025.200430

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma,H. (2024). A Systematic Review of Generative AI for Teaching and Learning Practice. Education Sciences, 14(6), 636. https://doi.org/10.3390/educsci14060636

Singh, B., Olds, T., Brinsley, J., Dumuid, D., Virgara, R., Matricciani, L., Watson, A., Szeto, K., Eglitis, E., Miatke, A., Simpson, C. E. M., Vandelanotte, C., & Maher, C. (2023). Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours.Npi Digital Medicine, 6(1). https://doi.org/10.1038/s41746-023-00856-1

Shan, C., Gong, S., & McOwan, P. W. (2009). Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing, 27(6), 803–816. doi:10.1016/j.imavis.2008.08.005

SpeechRecognition.(2025,12mayo).PyPI.https://pypi.org/project/SpeechRecognition/

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications Of The ACM, 9(1), 36-45. https://doi.org/10.1145/365153.365168

World Health Organization: WHO. (2024, 10 octubre). La saludmentalde los adolescentes. https://www.who.int/es/newsroom/factsheets/detail/adolescent-mental-health

Yan, C., Johnson, K., & Jones, V. K. (2024). The Impact of Interaction Time and Verbal Engagement with Personal Voice Assistants on Alleviating Loneliness among Older Adults: An Exploratory Study. International Journal Of Environmental Research and Public Health, 21(1), 100. https://doi.org/10.3390/ijerph21010100

Yang, S., Lee, J., Sezgin, E., Bridge, J., & Lin, S. (2020). Clinical Advice by Voice Assistants on Postpartum Depression: Cross- Sectional Investigation Using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. JMIR mhealth and uhealth, 9(1), e24045. https://doi.org/10.2196/2405