

Article Type: Original Article
Theme: Artificial Intelligence
Recive: 07/10/2023 | Accepted: 08/11/2023

Monitoring Emotional Response During Mental Health Therapy

Monitoreo de la Respuesta Emocional Durante Terapias de Salud Mental

Jorge Félix Martínez Pazos ^{1*} <https://orcid.org/0009-0009-2477-8611>

Arturo Orellana García ² <https://orcid.org/0000-0002-3652-969X>

William Gómez Fernández ³ <https://orcid.org/0009-0006-3554-2506>

David Batard Lorenzo ⁴ <https://orcid.org/0009-0007-3555-2875>

¹ Medical Informatics Center. University of Informatics Sciences. Carretera a San Antonio de los Baños Km 2 1/2, Finca Torrens, La Lisa, Havana, Cuba. Jorgefmp.mle@gmail.com

² Medical Informatics Center. University of Informatics Sciences. Carretera a San Antonio de los Baños Km 2 1/2, Finca Torrens, La Lisa, Havana, Cuba. aorellana@uci.cu

³ University of Informatics Sciences. Carretera a San Antonio de los Baños Km 2 1/2, Finca Torrens, La Lisa, Havana, Cuba. wbilly.dev@gmail.com

⁴ Center for Computational Mathematics Studies. University of Informatics Sciences. Carretera a San Antonio de los Baños Km 2 1/2, Finca Torrens, La Lisa, Havana, Cuba. dbatardl@gmail.com

* Corresponding Author. (jorgefmp.mle@gmail.com)

ABSTRACT

Facial emotion recognition is one of the most complex problems in computer vision, due to multiple factors ranging from image brightness to the personality of the individual. This paper built and elucidates the implementation of facial expression recognition solutions, an open-source package called FFEM to easily perform this task, and an application that integrates the previous package, using state-of-the-art models and algorithms for facial detection and emotion recognition mainly coming from MediaPipe and DeepFace with the intention of addressing the challenge of recognizing patients' emotions during cognitive therapy sessions. However, the versatility of this approach allows it to be applied to different industries and tasks, highlighting its potential for diverse use cases.

Keywords: DeepFace; Emotional Response; Face Detection; Facial Emotion Recognition.

RESUMEN

El reconocimiento de emociones faciales es uno de los problemas más complejos en visión por computador, debido a múltiples factores que van desde la luminosidad de la imagen hasta la personalidad del individuo. Este trabajo construyó y elucidó la implementación de soluciones de reconocimiento de expresiones faciales, un paquete de código abierto llamado FFEM para realizar fácilmente esta tarea, y una aplicación que integra el paquete anterior, utilizando modelos y algoritmos de última generación para la detección facial y el reconocimiento de emociones provenientes principalmente de MediaPipe y DeepFace con la intención de abordar el desafío de reconocer las emociones de los pacientes durante las sesiones de terapia cognitiva. Sin embargo, la versatilidad de este enfoque permite su aplicación a diferentes industrias y tareas, destacando su potencial para diversos casos de uso.

Palabras clave: DeepFace; respuesta emocional; detección facial; reconocimiento de emociones faciales.

Introduction

Facial emotion recognition (FER) is often considered one of the most challenging tasks in computer vision. As a cornerstone of human-computer interaction, FER has profound implications for disciplines such as human behavior analysis and healthcare (Khairuddin & Chen, 2021). The core of FER is the identification and classification of human emotions inferred from facial expressions, by examining facial patterns and features, computational systems are empowered to make educated guesses about a person's emotional state. This process underscores the transformative potential of FER in advancing our understanding of human emotions and the ability of machines to interpret them (Durai, 2023).

The performance of facial emotion recognition (FER) is accelerated by the application of sophisticated methods, primarily from the fields of deep learning and computer vision. Among these techniques, the use of convolutional neural networks (CNNs) has proved to be a powerful tool for emotion recognition tasks. An innovative approach called Facial Emotion Recognition via Convolutional Neural Networks (FERC) has been proposed, which uses a two-component CNN: the first component is tasked with eliminating the background from the image, while the second component focuses on deriving facial feature vectors. This unique structure allows for a more accurate and efficient emotion recognition process (Mehendale, 2020).

Despite the progress that has been made, the FER endeavor is fraught with complications that arise from several aspects. First and foremost, the considerable variation in facial configurations among individuals contributes to the complexity of accurately identifying emotions. In addition, there is an inherent element of ambiguity in the emotions displayed by an individual (Saroop et al., 2021). The efficacy of FER is tested through image changes that include variations in lighting conditions and facial orientations. These factors pose significant challenges to the robustness of emotion recognition systems (Khairuddin & Chen, 2021). Beyond personal experience, emotions are shaped by historical and cultural factors, resulting in differences in classifying, understanding, or even labeling emotions. This cultural diversity suggests that while certain emotional experiences may be common, their interpretation and manifestation may vary considerably (Durai, 2023). Several authors have proposed novel approaches to the difficult task of Facial Emotion Recognition

(FER), while others have undertaken comprehensive reviews and comparisons of existing Deep Neural Network architectures, with the aim of delineating the most optimal approach to tackle this complex task (Mehendale, 2020; Saroop et al., 2021; Dalvi et al., 2021; Özkara & Ekim, 2022).

Fei et al. (2020) describe an innovative framework for facial expression recognition, specifically tailored to support mental health care. The proposed system leverages the power of deep learning by extracting salient features from the Fully Connected Layer 6 of the renowned AlexNet architecture. These features are then trained using a standard linear discriminant analysis classifier. The authors postulate that this system has the potential for early detection of cognitive impairment, as patients with such conditions may exhibit abnormal facial expressions when exposed to visual and emotional stimuli. Jain et al. (2019) proposed a novel methodology for facial emotion recognition using a model based on single deep convolutional neural networks (DNNs) that incorporate convolutional layers and deep residual blocks outperforming current state-of-the-art methods in emotion recognition.

Based on the previous affirmations, FER allows healthcare professionals to gain a deeper understanding of their patients' emotional responses, which can be particularly useful in fields such as psychotherapy and psychiatry. During psychological sessions, the doctor or psychoanalyst can conduct patient monitoring, which is a continuous and objective tracking of the patient's emotions. This can be particularly useful for identifying subtle or gradual changes in emotional state that are difficult to capture in a standard therapeutic session. It can also be applied to real-time stress monitoring by measuring facial reactions, providing an objective indication of the patient's stress level, which is particularly useful in situations where the patient has difficulty expressing or recognizing their own emotions. In addition, by providing an objective and quantifiable measure of emotional responses, it aids in the understanding, identification and characterization of deficits in various neurodevelopmental and psychiatric disorders. Of course, as a result of the above, it provides a deeper understanding of the patient's emotional responses, helping clinicians to tailor their treatments to the specific needs of the patient (Gao et al., 2021; Kyranides et al., 2022).

The objective of this research is to use DeepFace, one of the most powerful solutions for FER developed by the team of Facebook AI Research, in a software application for real-time detection of patients' emotions during mental health therapy sessions (Taigman et al., 2014). This application will facilitate and support the

work of psychologists and psychiatrists by providing a deeper understanding of their patients and their emotional responses to specific stimuli.

The following is an overview of the contributions that have been provided in order to draw attention to the relevance of the work that will be presented in this study:

- FFEM package: An open-source package, licensed under MIT and available on PyPi, that simplifies the process of performing facial emotion recognition.
- Two Streamlit applications: One that requires an Internet connection, and the other that can be run offline, that use the FFEM package to apply facial emotion recognition in the field of mental health, helping psychologists and psychiatrists make decisions and treat patients.
- The use of state-of-the-art algorithms and models to achieve optimal performance in Facial Detection and Facial Emotion Recognition.

Materials and Methods

The proposed solution will be developed utilizing the Python programming language, adhering to the PEP 8 coding standard. This approach ensures consistency and readability in the codebase. The solution will leverage key libraries for computer vision, including OpenCV and MediaPipe, which provide a robust foundation for image processing tasks. Additionally, DeepFace will be employed for its advanced facial emotion recognition capabilities. For the construction and deployment of the application, Streamlit will be utilized due to its simplicity and efficiency in creating data-driven web applications. This comprehensive approach combines industry-standard tools and practices to ensure an effective and efficient solution for real-time facial emotion recognition. DeepFace is reported by the author to have a robust 97% accuracy in the FER task, which makes it so affordable for our software solution (Taigman et al., 2014). The entirety of the development and experimentation was conducted on a computer equipped with an Intel i5 7th generation processor, 8GB of RAM, and a GeForce GTX 1050 graphics card. This hardware configuration provided a robust platform for the computational demands of the project.

Figure 1 outlines the entire process: first, the input video is segmented into individual frames, then each frame is processed via MediaPipe for face detection, next, DeepFace is used for emotion recognition, then OpenCV is used to apply the identified emotion to the corresponding frame. This whole loop runs in real time, allowing real-time visualization of the processed video with the applied face detection and emotion recognition modifications.

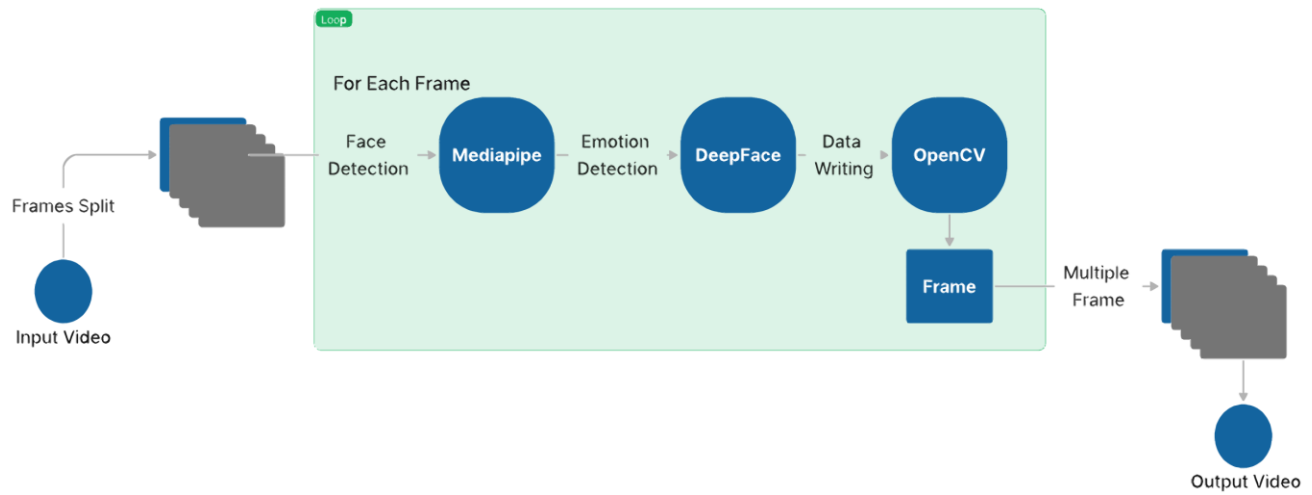


Fig. 1 – Video processing flow. Source: Authors

It is important to note that the processing power of the computer or server running the software plays a significant role in its performance. The frame rate (fps) may increase or decrease depending on the processing power available. This is due to the high computational cost associated with real-time tasks such as this. Therefore, for optimal performance, it is recommended to run the software on a machine with sufficient processing power.

Results and Discussions

The proposed solution is capable of face recognition and emotion recognition of either a single individual or multiple individuals. However, it is recommended to use the system on a single individual, as the inclusion of multiple individuals would significantly increase the processing level and require more computational

power. The following figures 2 and 3 illustrate how the face is detected and cropped using its bounding box for both scenarios: a single individual and multiple individuals.



Fig. 2- Face Detection of Single Individual. Source: The Authors





Fig. 3- Face Detection of Multiple Individuals. Source: The Authors

As observed in the previous figures, the faces are positioned at different angles and some individuals are wearing glasses, which does not represent a challenge; the system is proficient in face recognition as long as the faces are complete. This means that faces that are cut off or partially outside the image or video may not be satisfactorily recognized, as can be seen with the individual on the far right.

The videos used for the analysis presented in figure 4 were obtained from pexels.com, which operates under the CC0 license (Pexels, 2023). This license waives all copyright restrictions and allows unrestricted use and modification without legal consequences (Creative Commons, 2023).



Fig. 4- Test of the proposal over multiple videos with different characteristics. Source: Authors

Throughout the experiments, the model demonstrated exceptional performance in both the face and emotion recognition tasks. This high level of accuracy was maintained even under challenging conditions, such as when people were wearing hats or in scenes with extreme lighting conditions. Through several experiments, it was observed that the processing on the computer used took place at about 8 frames per second (fps), a very good value since 3-5 fps is a good rate for this type of task where the speed of movement is not so high. However, it is recommended to use more powerful computers when using the application in a production environment to ensure optimal performance (Lee & Hwang, 2022). The solution is divided into two variants,

the first is the Fast Facial Emotion Monitoring (FFEM) package, which is deployed on the Python Package Install under the MIT License, an open-source license, and the second is the integration of this package into the Streamlit application, which can be directly used for the task of monitoring patients' emotions during psychotherapy sessions.

The FFEM package allows fast execution of the whole process using only the *MonitorEmotion_From_Video* function. This function accepts a video path or an available webcam and performs processing, in addition to saving the resulting video to a destination folder. The main objective is to have an accessible module for future integration into health IT solutions, including its future integration into solutions for Xavia Pacs, a medical imaging system deployed in Cuban health centers by the Medical Informatics Center (CESIM). The FFEM package can be installed in a Python environment using the command ``pip install FFEM``, otherwise it can be obtained directly from its corresponding GitHub repository: <https://github.com/WiseGeorge/Fast-Facial-Emotion-Monitoring-FFEM-Package> or PyPi repository: <https://pypi.org/project/FFEM/>.

Two versions of the Streamlit application have been developed to address different scenarios. The first version is designed to run in an environment with an Internet connection, as it uses a video streaming library to optimize the process. On the other hand, the second version is designed for scenarios where there is no Internet connection, as it uses manual video streaming that does not require Internet access. In this way, we are able to cover any possible scenario for the deployment of the proposed application. The graphical user interface of version 1 of the application is shown in figure 5 below. The Streamlit application can be accessed through Streamlit Cloud: <https://monitoring-emotional-response-during-mental-health-therapy.streamlit.app/>, and its source code is available on the GitHub repository under the MIT license: <https://github.com/WiseGeorge/Monitoring-Emotional-Response-During-Mental-Health-Therapy> to serve as the backbone for future solutions.

The person depicted in Figure 3 was randomly selected from a pool of candidates, none of whom were authors of the paper. The image capture was done with their consent, thus eliminating any legal issues involved. This experiment shows that the use of eyeglasses does not pose a challenge to the proposed system. Although it

may affect classifications, the performance of the system remains largely consistent with previous experiments, shown in Figure 2, where subjects do not wear glasses.

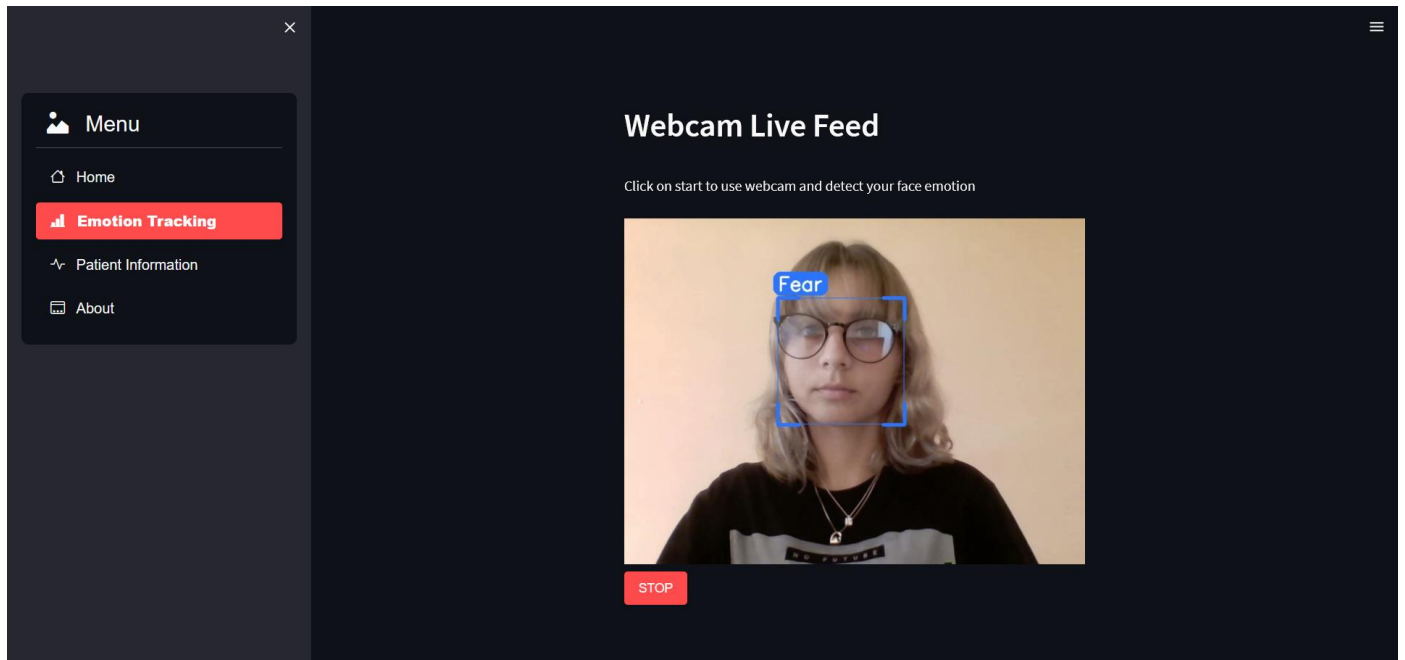


Fig. 5- Graphical user interface of the streamlit application (Online Version). Source: Authors

The proposed solution has the potential to revolutionize the field of mental health care: by providing real-time emotion monitoring during therapy sessions, it can offer valuable insights to psychologists and psychiatrists, improving their understanding of patients' emotional states and the effectiveness of therapeutic interventions. In addition, this solution could significantly contribute to the digitalization of healthcare in Cuba; by integrating advanced technology into mental healthcare, it could help bridge the gap between traditional practices and modern, data-driven approaches, which could lead to more accurate diagnoses, more effective treatments, and ultimately better patient outcomes.

In addition, the development of this solution underscores the potential of computational diagnostics in healthcare through the use of advanced algorithms and machine learning techniques to extract meaningful information from complex data sets, such as video feeds of therapy sessions. This could pave the way for

more sophisticated diagnostic tools in the future, changing the landscape of healthcare not only in Cuba but globally.

The clinical implications of a proposed solution in psychotherapy sessions are profound. This solution provides a novel approach to understanding the patient's emotional state, which is critical to the therapeutic process. By monitoring emotional responses, therapists can tailor their approach based on the patient's real-time emotional state, allowing for more personalized therapy, potentially leading to more effective outcomes, thereby improving therapeutic accuracy (Soloski & Deitz, 2016). It can also lead to improved patient engagement. When patients are aware that their emotional responses are being monitored, they may feel more involved in the therapy process, which can improve the therapeutic alliance, a critical factor in treatment success (Barlow et al., 2010; DeAngelis, 2019). The ability to monitor emotional responses can provide valuable data for research. Data that can be used to study the effectiveness of different therapeutic interventions and contribute to evidence-based practice (Soloski & Deitz, 2016).

Future research aims to store patients' emotions in a database in real-time, along with the timestamp of their occurrence. This would allow identification of the prevailing emotion at any given time during a session. In addition, the inclusion of a dashboard to visualize data analysis related to each patient's emotions would enhance the insights that specialists can gain and potentially improve patient outcomes.

Conclusions

In this study, a real-time facial emotion recognition (FER) solution has been successfully developed and implemented. This solution uses state-of-the-art models and algorithms, demonstrating the potential of advanced computational techniques in emotion recognition tasks. The solution was encapsulated in an open source package, FFEM, and integrated into a Streamlit application, demonstrating its applicability to various use cases and facilitating the development of future solutions based on the study and the open source tools provided. The project plays an important role in advancing digital transformation in Cuba, with a particular

focus on digital health. By harnessing the power of advanced technologies such as FER, this study contributes to ongoing efforts to modernize healthcare practices.

References

- Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. T. (2010). Module 2: Recognizing and Tracking Emotional Responses. In Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide. Oxford University Press <https://doi.org/10.1093/med:psych/9780199772667.003.0007>
- Creative Commons. CC0. Available at <https://creativecommons.org/public-domain/cc0/>. Last Accessed 27/10/2023
- Dalvi, C., Rathod, M., Patil, S., Gite, S., & Kotecha, K. (2021). "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions," in IEEE Access, vol. 9, pp. 165806-165840. <https://doi.org/10.1109/ACCESS.2021.3131733>
- DeAngelis, T. (2019). Better relationships with patients lead to better outcomes. Monitor on Psychology, 50(10). <https://www.apa.org/monitor/2019/11/ce-corner-relationships>
- Durai, P. (2023). Facial Emotion Recognition: Decoding Expressions. Available at <https://learnopencv.com/facial-emotion-recognition/>
- Fei, Z., Yang, E., Li, D.D.-U., Butler, S., Ijomah, W., Li, X., Zhou, H., (2020). Deep convolution network based emotion analysis towards mental health care. Neurocomputing 388, 212–227. <https://doi.org/10.1016/j.neucom.2020.01.034>
- Gao, Z., Zhao, W., Liu, S., Liu, Z., Yang, C., & Xu, Y. (2021). Facial Emotion Recognition in Schizophrenia. Frontiers in Psychiatry. <https://www.frontiersin.org/articles/10.3389/fpsyt.2021.633717>
- Jain, D.K., Shamsolmoali,P., & Sehdev,P.(2019). Extended Deep Neural Network for Facial Emotion Recognition.Pattern Recognition Letters. <https://doi.org/10.1016/j.patrec.2019.01.008>

- Khairuddin, Y. & Chen, Z. (2021). Facial Emotion Recognition: State of the Art Performance on FER2013 <https://doi.org/10.48550/arXiv.2105.03588>
- Kyranides, M. N., Christofides, D., & Çetin, M. (2022). Difficulties in facial emotion recognition: Taking psychopathic and alexithymic traits into account. *BMC Psychology*, 10(1), 239. <https://doi.org/10.1186/s40359-022-00946-x>
- Lee, J., Hwang, K.-I., 2022. YOLO with adaptive frame control for real-time object detection applications. *Multimedia Tools and Applications* 81, 36375–36396. <https://doi.org/10.1007/s11042-021-11480-0>
- Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Appl. Sci.* 2, 446. <https://doi.org/10.1007/s42452-020-2234-1>
- Pexels. Legal Simplicity. Available at <https://www.pexels.com/license/>. Last Accessed 27/10/2023
- Saroop, A., Ghugare, P., Mathamsetty, S., & Vasani, V. (2021). Facial Emotion Recognition: A multi-task approach using deep learning. <https://doi.org/10.48550/arXiv.2110.15028>
- Soloski, K.L., Deitz, S.L. Managing Emotional Responses in Therapy: An Adapted EFT Supervision Approach. *Contemp Fam Ther* 38, 361–372 (2016). <https://doi.org/10.1007/s10591-016-9392-8>
- Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In 2014 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1701-1708). Columbus, OH, USA. <https://doi.org/10.1109/CVPR.2014.220>
- Özkara, C., & Ekim, P. O. (2022). Real-Time Facial Emotion Recognition for Visualization Systems, 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), Antalya, Turkey, pp. 1-5 <https://easychair.org/publications/preprint/Z9H9>

Conflict of interest

The author authorizes the distribution and use of his article.

Author's contributions

1. Conceptualization: Jorge Félix Martínez Pazos, Arturo Orellana García
2. Data Curation: Jorge Félix Martínez Pazos
3. Formal analysis: Jorge Félix Martínez Pazos, Arturo Orellana García

4. Acquisition of funds: -
5. Research: Jorge Felix Martinez Pazos, Arturo Orellana García, William Gómez Fernández.
6. Methodology: Jorge Félix Martínez Pazos, Arturo Orellana García
7. Project Administration: Jorge Félix Martínez Pazos, Arturo Orellana García, Arturo Orellana García
8. Resources: Jorge Félix Martínez Pazos
9. Software: Jorge Félix Martínez Pazos, William Gómez Fernández
10. Supervision: Arturo Orellana García
11. Validation: Arturo Orellana García
12. Visualization: Jorge Félix Martínez Pazos
13. Editing: Jorge Félix Martínez Pazos, Arturo Orellana García. William Gómez Fernández
14. Editorial staff: Jorge Félix Martínez Pazos, Arturo Orellana García.