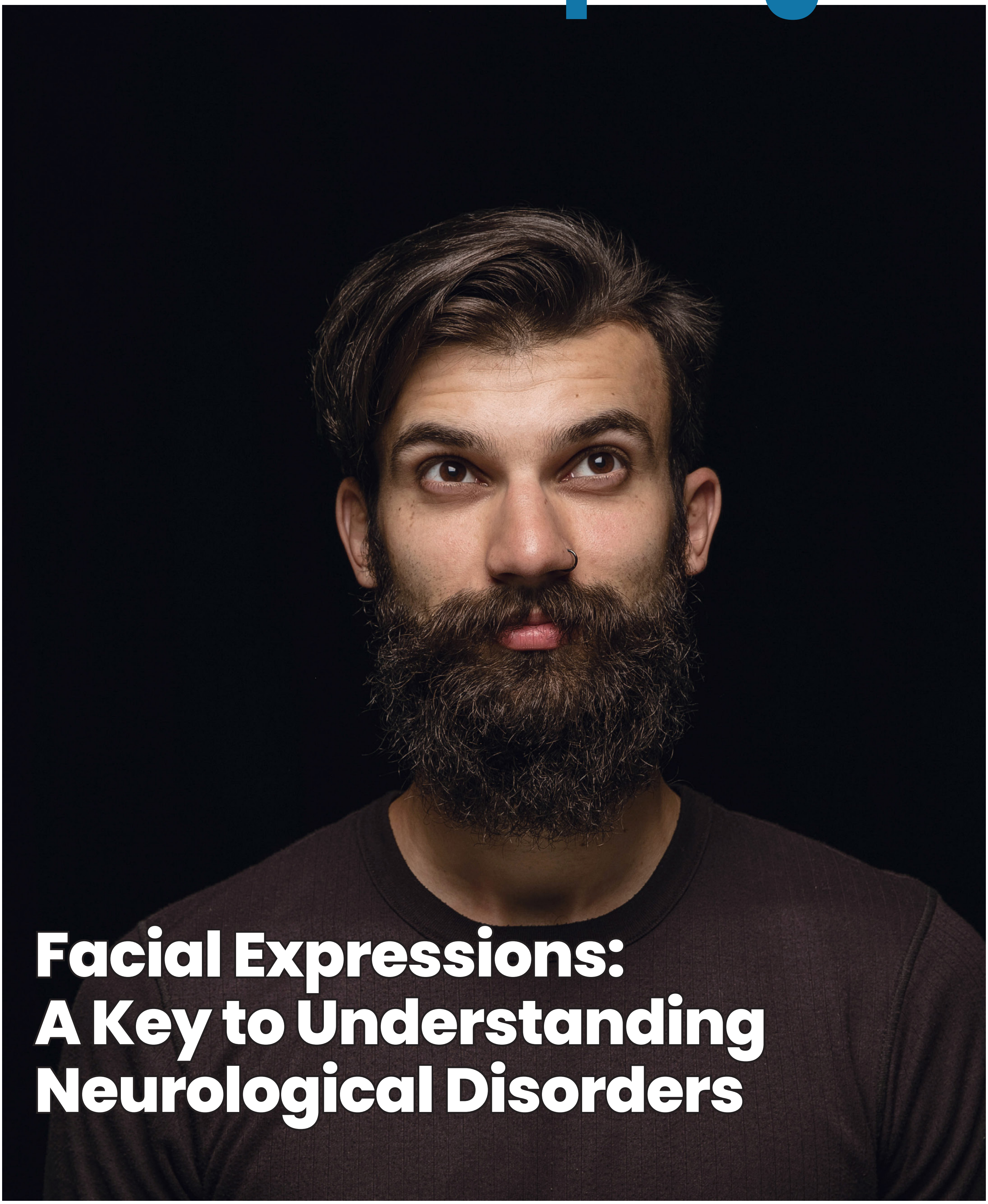
Public-Friendly Summaries
n 02

Innovation Spotlights



Facial Expressions: A Key to Understanding Neurological Disorders

Animi est enim omnis actio et imago animi vultus, indices oculi (Marco Tullio Cicerone, De Oratore). As the ancient Roman philosophers already realized, the relationship between visual communication and the transmission of emotions is a direct means of understanding human feelings, emotions and moods.

Facial expressions (FE) are a natural and convincing way to communicate human emotions and can provide valuable insights to observers when assessing emotional states and mood. In healthcare, facial expressions of patients with neurological disorders such as **Parkinson's, Stroke, and Alzheimer's** can assist doctors in evaluating the physical condition of a patient as they can correlate with symptoms such as pain, fatigue or others. Monitoring mood over time, may uncover changes and anomalies in the functioning and behaviour of an individual, which might indicate significant preclinical decline. Traditional methods of diagnosing these conditions and assessing their progress over time involve proper observation and clinical tests, which can be invasive, expensive, and time-consuming. Thus, developing alternative approaches to monitor and analyze emotional wellbeing is essential.

Automatic facial expression recognition (FER) and mood estimation technologies can assist doctors in evaluating the overall behavior, emotional status and mood variations of

neurological patients. These technologies can efficiently differentiate and identify various facial expression features associated with specific clinical conditions. These features can be combined with traditional clinical outcomes as novel digital biomarkers to support the evaluation of therapeutic response.

To address these challenges, ALAMEDA has made two significant steps forward: our researchers have developed a novel **Mood Estimation Android App (MEAA)** and a **facial expression recognition (FER) AI model**.

The Mood Estimation Android App

The Mood Estimation Android App (MEAA) is a smartphone application, developed by our researchers from **Catalink**, that **monitors the face of the patients while they interact with other mobile applications, and estimates their mood** based on their facial expression. Specifically, the model is capable of differentiating between 'negative', 'neutral', 'positive' or 'other' mood states. The objective of MEAA is to give the doctors relevant insights of patients' mood during remote medical monitoring.

Our model has been **integrated with two other applications**, the Zenon ChatBot[1] , developed by University of Nicosia, and WellMojo[2], developed by Wellics, and it **operates as a background service** while patients respond to medical questionnaires through the two aforementioned applications. The app monitors the patient's facial expressions as they engage with these apps, enabling it to detect their mood. For the further improvement of the model's predictions, we combined our predictions with the text sentiment analysis output, which takes place while the patient interacts with the ChatBot. By fusing the predictions of two modalities, we aim to provide medical experts with a more robust and accurate mood analysis.

The main strength of our application lies in the **deep learning model** which processes the input video frames and outputs the mood analysis. The model chosen for this application is the mini-Xception[3] model, since it can be easily deployed and run on smartphones, and it demonstrates great prediction accuracy and negligible inference latency for mobile vision tasks. To test the model's accuracy under real conditions, we tested the model on real patients, where the results were encouraging since it scored 70% accuracy.

As soon as the application opens, the first step is the localization of the user's face within the video frame. This is realized through the exploitation of the ML Kit[4] tool, which is able to locate the face of the user and draw a bounding box around it. The region of interest is cropped and fed to the mood estimation model. The model outputs mood estimation analysis of the user ("negative", "positive", "neutral" or "other"), and each mood is accompanied with the prediction probability. The overall workflow of the application is depicted in Figure 1.

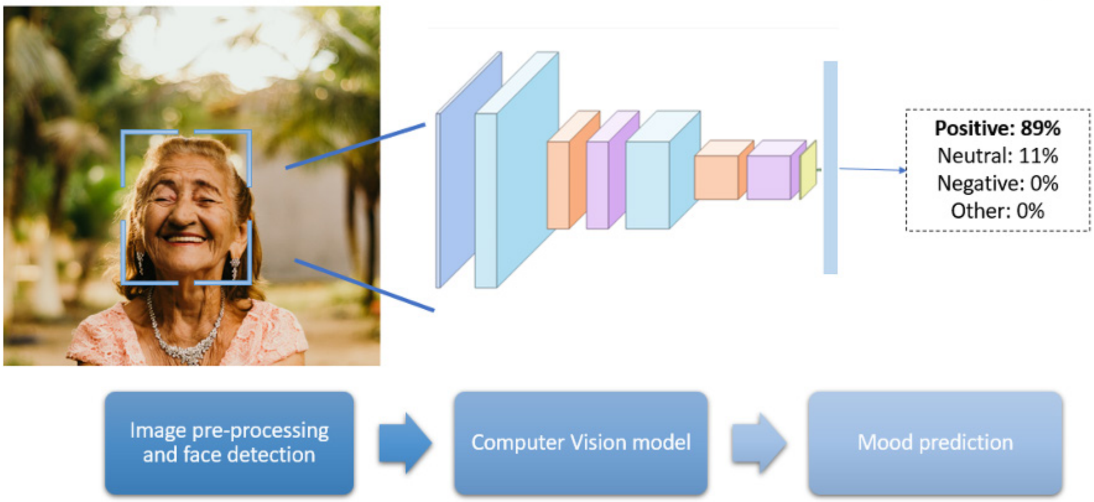


Figure 1: *The workflow of the Mood Estimation Android App. First the input image is pre-processed and the face of the user is detected and cropped. Afterwards, the cropped image is fed into the Computer Vision model which finally outputs the mood estimation result.*

The FER AI Toolkit

[1] Chira, C. et al. (2023). An Affective Multi-modal Conversational Agent for Non Intrusive Data Collection from Patients with Brain Diseases. In: Følstad, A., et al. Chatbot Research and Design. CONVERSATIONS 2022. Lecture Notes in Computer Science, vol 13815. Springer, Cham. **https://doi.org/10.1007/978-3-031-25581-6_9**

[2] Sorici, A., et al. (2023). Monitoring and Predicting Health Status in Neurological Patients: The ALAMEDA Data Collection Protocol. Healthcare. 2023; 11(19):2656. **https://doi.org/10.3390/healthcare11192656**

[3] Arriaga, O., et al. (2017). Real-time Convolutional Neural Networks for Emotion and Gender Classification. arXiv:1710.07557. Retrieved from **https://arxiv.org/abs/1710.07557**

[4] **https://developers.google.com/ml-kit**

The **facial expression recognition (FER) AI toolkit**[5] for **monitoring neurological disorder patients via deep learning based facial expression analysis** developed by the ALAMEDA researchers from **Norwegian University of Science and Technology (NTNU)** proposes an innovative deep learning-based facial expression analysis framework that can monitor early-stage neurological disorder patients. The result is a model deployable on resource-constrained devices such as smartphones and tablets and achieving high accuracy on data collected from YouTube containing faces with numerous expressions belonging to different genders and ages. Specifically, the dataset used to train the framework is the Karolinska Directed Emotional Faces (KDEF), a set of 4900 pictures of human facial expressions. KDEF is a publicly available dataset, produced in 1998, developed by the psychological section of the Department of clinical neuroscience. It contains images of human facial expressions taken from five different angles with various cameras.

The arranged data consists of 900 RGB images in each of the four classes (normal, happy, anger, and sad) split between the training and validation set. 80% of the data is used for training and 20% for evaluation. For testing on real patients, the researchers collected full-length videos from YouTube by searching in different well-known channels like Michigan Medicine and 60min Australia BAYSTATEHEATH. They extracted frames from each video and selected frames or parts of the video to pass through the trained model for real-time evaluation based on the expression and age of the patients.

Preprocessing is a critical step to improve the learning capabilities of the model during training. The aim is to remove unessential pixels from raw images and keep only the

[5] Munsif, M., et al. (2022). Monitoring Neurological Disorder Patients via Deep Learning Based Facial Expressions Analysis. 412-423. http://dx.doi.org/10.1016/j.aej.2023.01.017.

region of interest for processing. The first step is to detect the face and then crop it. Face detection is a challenging task due to angles and illumination variations. To avoid such variations, a popular algorithm called **Viola-Jones algorithm** is used for face detection. RGB images are converted to greyscale before being fed into the algorithm. To reduce computational cost, the cropped images are downsampled before being fed into the proposed training model.

The FER AI Toolkit utilizes **Convolutional Neural Networks (CNN)** to analyze facial expressions and estimate the probability scores for different emotional classes, depending on the selected category. By inputting facial images, the system provides predictions for the likelihood of emotions such as happiness,

sadness, neutrality, pain, fear, and disgust, tailored to the chosen emotional category.

Benefits of the ALAMEDA innovative emotion recognition technologies

By integrating the new emotion recognition technologies developed by ALAMEDA into the management of neurological disorders, healthcare providers can better understand the emotional aspects of these conditions and provide more targeted and comprehensive care. Specifically, tangible benefits can be achieved in the following areas:

Monitoring Progress: our technologies can be used to monitor the progress of patients with neurological disorders over time. By regularly assessing emotional states, they can provide objective measurements that help evaluate the effectiveness of treatments and the overall progress of the patient..

Remote Monitoring and Telemedicine: emotion recognition technologies can enrich remote monitoring and telemedicine for patients with neurological disorders. By using cameras, patients can have their emotional states monitored and analyzed from the comfort of their homes. This allows for more accessible and convenient healthcare, reducing the need for frequent in-person visits.

Challenges and future research directions

Future research efforts should focus on

improving model performance, generalization across disorders, and ensuring ethical considerations are upheld to maximize the potential benefits of these systems in clinical practice. With the aim to **improve accuracy and robustness** of the FER AI model, there are plans to collect more challenging datasets from patients, improve the system through attention mechanisms and incorporate temporal information with spatial information of specific facial expressions. Moreover, the performance and reliability of emotional tracking apps such as MEAA could be significantly improved by encompassing additional modalities, such as physiological signals analysis or text sentiment analysis. For future work, further research towards **exploration of multi-modal mood analysis** will be carried out.

Specifically, future research work will delve into the following challenges:

1. Limited and Imbalanced Dataset: Obtaining a diverse and well-labeled dataset for facial expressions, especially for emotions like pain and disgust, can be challenging. The scarcity of open-source datasets for these specific emotions can lead to imbalanced training data, affecting the performance of the model.
2. Model Complexity and Overfitting: Building an effective convolutional neural network (CNN) model for facial emotion recognition is a complex task. Overfitting can occur when the model learns to perform well on the training data but fails to generalize to new, unseen data.

3. Fine-grained Emotion Differentiation: Emotions like pain and disgust can have subtle facial expressions that are challenging to detect accurately. Developing advanced models capable of capturing fine-grained facial cues and nuances specific to neurological disorders is crucial.

4. Variability in Facial Expressions: Neurological disorders can introduce variability in facial expressions due to muscle weakness, paralysis, or other motor impairments. This variability can make it difficult to extract consistent features for emotion recognition. One of the main challenges that we faced during the development of the MEAA, was that due to the neurological disease, the patients' facial muscles often become increasingly rigid; therefore, the richness of facial expressions is significantly decreased. Thus, frequently, facial expressions provide little information related to the actual mood of the patient. To mitigate such phenomena, our computer vision model's results are fused with the ChatBot's text sentiment analysis predictions, and the final decision about the emotional status of the patient is sent to the medical experts. In this way, we aimed to improve the accuracy of the final prediction.

5. Models must be lightweight to be deployed on mobile devices. Commonly, most of the CNN frameworks contain millions of parameters, therefore their deployment on edge devices for real-time applications is not trivial. After experimentation with different state-of-the-art models, Catalink researchers found out that mini-Xception outperforms the other models, both in terms of performance accuracy and model size. Mini-Xception's architecture is lightweight, and its final size is less than 1 MB, thus, the model can easily be deployed on a smart device, without demanding much of the device memory.

6. User privacy and transparency concerns: the MEAA was developed with a privacy-by-design approach. The analysis is conducted only on the device. The extracted video frames are not stored on any external server but are instead discarded after the model estimates the mood. As for transparency, as soon as the app opens, it asks for the user's permission before opening the camera. Moreover, the user is made aware of the background service and the video recording, since the application displays a notification stating "Camera opened". As long as the app is active, a notification at the top of the screen informs the user that he/she is being recorded.

Further suggested readings

The same ALAMEDA researchers from **Norwegian University of Science and Technology (NTNU)** together with colleagues from **Catalink (CTL)** and other international co-authors have produced an **insightful survey of existing literature on FER**[6] where the working flow of FER methods, their integral and intermediate steps, and pattern structures, as well as FER datasets, methods and performance metrics are discussed. This is another not-to-be-missed publication from ALAMEDA for anyone interested in the latest advances and the yet open research avenues in FER.

[6] Sajjad, M., et al. (2023). A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines. Alexandria Engineering Journal. 68. 817-840. **https://doi.org/10.1016/j.aej.2023.01.017.**

Alameda

Bridging the Early Diagnosis and Treatment Gaps of Brain Diseases

This project has received funding from the **European Union's Horizon 2020** research and innovation programme under grant agreement **No 101017558**