Automated User State Detection

Carlotta May[1] and Marlon Spiess[2]

[1] University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany
cmay4@smail.uni-koeln.de
[2] University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany
mspiess1@smail.uni-koeln.de

GitHub Repository: https://github.com/Marbru35/MCQuality

Abstract. Facial expression recognition from video data enables real-time user state detection, crucial for applications in usability testing, customer experience, and human attention analysis. This study systematically reviews existing tools for real-time emotion recognition and evaluates their capabilities. A prototype using the DeepFace library demonstrates real-time facial emotion tracking, highlighting its potential for user experience analysis. While deep learning-based tools can detect strong emotions, they struggle with subtle expressions and environmental challenges. The study emphasizes the need for multimodal approaches to enhance accuracy and robustness.

Keywords: Emotion Detection · Facial Expression Recognition · Real-Time Video Analysis · Usability Testing

1 Introduction

Emotions reveal a person's true nature through their involuntary reactions in situations where they cannot be suppressed or controlled [55]. Consequently, they serve as a reliable indicator of opinions and impressions. Building on this, emotion detection allows companies to go beyond traditional metrics like response time and resolution rates and explore how customers feel during interactions. Negative emotions such as frustration or anger can indicate areas where the service falls short, while positive emotions reflect what works well. Using emotion tracking to see whether users can use a tool without complications and anger, organizations can identify and address weak points in real time, ensuring that issues are resolved promptly and effectively. This proactive approach not only improves the experience of individual users, but also helps to build trust and long-term relationships, both of which are hallmarks of successful Customer Relationship strategies [13].

The importance of user satisfaction extends to almost every business context, underscoring the principle that the user is king [26,42]. Whether in customer service, marketing analysis, or product design, understanding how users feel during interactions provides actionable insights to refine offerings and improve engagement [15,34]. Automatic user state recognition, which uses both verbal and non-verbal cues, plays a crucial role in this process. Verbal cues include speech characteristics such as tone of voice and pitch, while non-verbal cues include facial expressions, eye contact, gestures, and body language [22].

The potential of emotion detection extends far beyond business. In healthcare, these technologies support diagnosis and therapeutic processes by providing real-time insight into the emotional and stress states of a patient [57,53]. In autonomous driving, monitoring the driver by tracking emotions can contribute to safety by detecting fatigue, stress, or distraction of the driver [11]. The recognition of emotions is also advantageous in marketing analysis, as it measures customer reactions to campaigns or products and thus allows data-supported adjustments to be made and fields of interest to be updated [15]. Although people may try to control their expressions, emotions and stress are tied to involuntary physiological changes, such as changes in heart rate or pupil dilation, revealing a person's true state and making them measurable for automated systems [34].

The literature on user state detection lacks a comprehensive review of tools capable of recognizing user emotions and stress in real-time scenarios using speech and video data. This paper addresses this gap by conducting a systematic literature review to identify the state-of-the-art tools and frameworks. The goal is to provide an extensive list of existing tools and to evaluate the identified solutions against the background of improving usability and user satisfaction.

The remainder of this literature review is structured as follows. In Section 2, the underlying concepts of emotions, stress, and facial expression recognition (FER) are briefly explained to provide the fundamentals related to user state detection. This section establishes the necessary theoretical foundation to explore current tools and frameworks capable of detecting user emotions and stress in real-time scenarios, while also reviewing related work to contextualize existing approaches. In Section 3, the methodology describes the search strategy and selection criteria used to identify relevant papers and tools. In Section 4, a comprehensive list of tools is identified and outlined. This list comprises, among others, APIs, SDKs, and libraries for emotion detection on video and speech data. These findings provide the basis for further evaluation. Moreover, this section presents the implementation of a prototype based on one of the identified tools, demonstrating its potential for facial emotion recognition in real-time. In Section 5, the identified tools are discussed using defined comparison criteria. Additionally, possible limitations and future research will be indicated. In Section 6, the conclusion summarizes the key findings and provides closure.

2 Background and Related Work

Emotions and stress play a central role in human perception and interaction with the environment. They influence not only our daily behaviour, but also the way we interact with technology and software. Emotions are automatic, often unconscious reactions to external or internal stimuli that influence our decision-making, facial expressions, speech and body language. Similarly, stress, which is often associated with certain emotional states, affects behaviour and responses to demands. In the context of digital products and applications, both emotions and stress have a significant impact on the user experience (UX). A deeper understanding of these psychological and physiological responses enables designers to create user interfaces that meet users' needs and promote a positive overall experience.

2.1 Emotions

Emotions are fundamental components of the human psyche [55], shaping perception, behaviour, and decision making. They are automatic, involuntary responses of an organism to specific stimuli and are essential for human behaviour. Unlike moods, which are long-lasting emotional states, emotions are short-lived experiences that usually last only seconds or minutes [34]. Emotional states affect both cognitive and physical responses, influencing decision-making, facial expressions, posture, and speech [22]. By shaping how individuals react to internal and external events, emotions play a crucial role in human interaction and engagement with their environment [15,34].

Various theoretical models have been developed to systematically analyse and categorise emotions, with Ekman's theory of basic emotions and Russell's circumplex model being among the most influential and widely used frameworks in emotion research.

Paul Ekman's widely accepted theory of emotions identifies six fundamental emotions: joy, sadness, anger, fear, surprise, and disgust. According to Ekman, these basic emotions are universal across all cultures, meaning that individuals from different backgrounds and societies display similar facial expressions when experiencing emotions. For instance, a smile universally signifies happiness, whereas a frown is commonly associated with sadness. These emotions arise as reaction to external stimuli and are nearly uncontrollable, appearing unconsciously without intentional effort in response to stimuli. Moreover, they dissipate rapidly, which differentiates them overall from more complex emotional states [16].

Russel's Model of Emotions describes emotions as disturbances within a two-dimensional space defined by valence (pleasantness) and arousal (intensity). These dimensions are both independent and bipolar. Independence means that valence and arousal are uncorrelated, allowing for a wide variety of emotional combinations. Bipolarity means that opposing emotions are positioned at opposite ends of each axis. For example, emotions like happy and sad are positioned at opposite poles of the valence dimension, whereas tense and sleepy are opposites along the arousal dimension. This structure means a person cannot simultaneously experience emotions at opposing ends of the same axis, such as being both tense and sleepy [49,55].

Building on these basic theories, practical applications of emotion recognition focus on the analysis of physiological and behavioural signals, where facial expressions and vocal characteristics are important indicators.

The recognition and expression of emotions are central to human communication. Basic emotions, in particular, are expressed through facial expressions that are inborn and consistent across all cultures [16]. These expressions are related to different physiological responses, including variations in breathing patterns, heart rate, pupil dilation, and electrical activity in facial muscles [22].

Facial landmark-based analysis is a key approach to emotion recognition. The movement of the eyebrows, mouth, and chin serve as critical reference points for determining emotional states based on their specific configurations and dynamics [23]. Advanced algorithms combined with face-reading technologies utilise these landmarks to classify emotions with high accuracy [23,34].

Similarly, speech-based emotion recognition systems analyse vocal features to enhance the overall accuracy of emotion detection by incorporating multiple modalities. Studies by Banse and Scherer highlight the relationship between vocal characteristics, such as pitch, loudness, and speech rate, and emotional states. For example, rapid speech or increased vocal energy often indicates excitement or anger [7]. Speech-based systems process real-time data to classify emotions and provide an additional modality for comprehensive emotion analysis [5].

2.2 Stress

Stress sharpens the senses and reduces reaction times as a natural reaction to perceived threats [43]. It is broadly defined as a psychological shift from calm to heightened emotional responses due to external or internal stimuli [12]. Historically crucial for survival, stress has become a chronic issue in modern life due to persistent pressures like financial challenges and work overload [43]. Chronic stress is associated with several health risks, including cardiovascular disease, one of the leading causes of death in Western societies [54].

Stress frequently co-occurs with emotions such as fear, anger, and sadness, amplifying its effects on the body and mind [36]. This association suggests the potential for indirect stress detection through emotional states [21].

Unlike universal emotions, stress expressions are highly individualised. Stress-related facial expressions vary significantly from one person to another, influenced by personal appraisal processes [36,35]. These processes depend on factors like values, beliefs, and perceived control, which shape stress responses [20]. For instance, stress levels can vary based on how individuals evaluate a situation and their confidence in their own coping mechanisms [35].

Stress detection employs modalities like facial expressions and speech. Video analysis captures subtle indicators such as pressed lips or raised eyebrows, while speech data reveals stress through changes in pitch, prosody, and intonation [41,47]. Additionally, keywords linked to stressful scenarios provide further insights [41]. However, the effectiveness of these methods and their connection to emotion-related facial activity remains under explored [56].

This review will further explore User State Detection (USD) in the context of emotions.

2.3 User State Detection and Facial Expression Recognition

USD utilises automated systems to identify a user's emotional and physical states without requiring direct input from the user. By analysing multimodal data, such as facial expressions, voice characteristics, and physiological signals, USD offers insights into UX and satisfaction [22].

Focusing on the domain of emotion detection based on video data, facial expression recognition (FER) is a central approach that analyses facial features and expressions to identify emotions in real-time scenarios. In general, the process typically involves three key steps: preprocessing, feature extraction, and emotion classification. Initially, advanced computer vision techniques detect faces in video data and preprocess them to ensure they are suitable for further analysis. Building upon this, features such as facial landmarks-key anatomical points like the corners of the eyes, nose, and mouth-are extracted, enabling a precise detection of movements over time, which collectively form an expression correlated with emotional states [28,32].

Furthermore, the Facial Action Coding System (FACS), introduced by Ekman and Friesen [17], employs a method of encoding specific movements of facial muscles and subtle expressions into AUs [27], which have been scientifically validated as indicators of distinct human emotions [14]. Illustrative examples of AUs and their combinations include raised eyebrows (AU1 + AU2) or tightened lips (AU23) [27].

Once the features are extracted, classification algorithms interpret patterns of these facial actions to assign emotions from a discrete set of categories [28,32]. For instance, happiness is associated with a smile, characterised by the upward movement of mouth corners and wrinkles near the eyes [16].

FER methods can be categorised into two distinct approaches: conventional and deep learning-based. The primary distinction between these approaches lies in their reliance on feature extraction. Conventional approaches entail the manual extraction of features in a structured process, whilst deep learning approaches utilise end-to-end learning to automatically derive features and learn directly from raw data through neural networks [28,32,33].

By translating facial expressions into recognised emotional states, the system achieves an accurate analysis of user emotions, offering a deeper understanding of user states and satisfaction.

In addition to video-based analysis, speech recognition can further aid USD by analysing changes in pitch, tone, and rhythm, which reflect different emotional states [5,7]. Physiological data, such as heart rate or skin conductance, can also provide valuable insights into stress and other user conditions [37]. These multimodal approaches extend the scope of USD beyond emotion recognition to encompass a broader range of user states.

USD is particularly valuable in fields such as healthcare, marketing, and autonomous driving, where the ability to detect and respond to users' emotional and physical states enhances user-centred design. By enabling systems to adapt to the needs of users, USD contributes to more personalised and effective interactions.

2.4 Related Work

In the existing literature, the detection of emotions has already received considerable attention in a wide variety of fields. Affective computing focuses on the development of systems capable of recognising, processing, and analysing human emotions. These systems often extend beyond textual inputs, also incorporating non-verbal cues. Within affective computing, sentiment analysis emerges as an approach that leveraged natural language processing (NLP) and machine learning (ML) techniques to extract and analyse subjective information from textual data. It aims to classify the expressed sentiment as positive or negative in order to gain insights into user emotions [9].

Several publicly available datasets support research in emotion recognition by offering labelled images tagged with specific emotion states. Grounded in Ekman's basic emotions, these datasets often include diverse representations of sex, ethnicity, and age groups, enabling the development of robust and generalisable models. Notable examples include FER2013, a benchmark dataset for evaluating algorithms with basic emotion annotations [24], and AffectNet, which offers annotations for valence and arousal. CK+ focuses on spontaneous expressions [38], while Ascertain provides detailed annotations for nuanced analyses. EMOTIC captures emotions in practical settings, and the Google Facial Expression dataset covers a wide range of expressions [1]. Most of these datasets are conveniently hosted on Kaggle, an online platform that foster collaboration and innovation in data science and machine learning, providing easy access to researchers and developers [46].

Facial recognition and state detection are have been widely adopted across multiple industries. In the context of autonomous driving, for example, driver attention and emotional states are increasingly being monitored to enhance safety. The systems are designed to recognise facial expressions, eye movements, and head posture and enable the vehicle to assess whether the driver is attentive or absent [19]. This technology plays a crucial role in the implementation of safety functions such as automatic braking, lane departure warning, and collision avoidance, which rely on real-time assessments of the driver's condition to be able to intervene in good time.

In psychology and healthcare, patient state recognition is used in therapy sessions to gain a better understanding of emotional states and identify key trigger points of concern. By observing facial expressions, tone of voice, and other behavioural indicators, therapists can better tailor their approach to the emotional needs of their patients, increasing the overall effectiveness of therapy [53].

FER technology has become increasingly integrated into everyday applications, enhancing both security and UX across various sectors. In consumer technology, devices like smartphones and tablets have adopted facial recognition for user authentication. For instance, Apple's Face ID system uses advanced facial mapping to securely unlock devices and authorise transactions. While primarily focused on identification, such systems can potentially incorporate FER to adapt device responses based on user emotions, thereby personalizing UX.

The COVID-19 pandemic underscored significant challenges in FER due to face masks obstructing crucial facial features Studies have shown that face masks can reduce the accuracy of emotion recognition, particularly affecting the identification of emotions like disgust, fear, and surprise [48].

To address these limitations, researchers have developed methods to adapt FER systems for masked faces. One approach involves a hybrid convolutional neural network (CNN) combined with local binary patterns to extract features from the visible regions of the face, particularly the upper facial features like the eyes and eyebrows. This method focuses on recognising basic emotions and has demonstrated improved accuracy in recognising emotions from masked faces [18]. Companies like Apple also adapted their technologies for facial recognition authentication to maintain user convenience and security. Initially, Apple introduced a feature allowing users to unlock their iPhones with Face ID while wearing a mask. Subsequently, with the release of iOS 15.4, Apple enhanced Face ID to recognize users even when masked, by focusing on the unique characteristics around the eyes [2].

Existing research demonstrates that emotion recognition and the underlying facial (feature) recognition already well-established across various applications, yet challenges remain particularly in ensuring robust recognition under varying conditions such as occluded faces. Future developments focus on enhancing current models through multimodal approaches and improved algorithms to further increase accuracy and applicability in real-world scenarios.

3 Method

This research is carried out in two main steps: literature analysis and identification of tools and libraries to detect emotions and stress.

The first step focuses on the analysis of the academic literature to explore existing models for emotion and stress detection. A systematic search strategy was developed using search terms such as 'human emotion recognition', 'automated detection user emotions' or 'pattern emotion recognition' to capture the basic theories, as well as the indicators, associated with emotional states and stress level. Searches were conducted in key academic databases, including Google Scholar and Web of Science. Relevant studies were identified by screening titles and abstracts, followed by a combination of backward reference tracing and forward citation searches to ensure a broad and in-depth exploration of the field.

The literature review examines approaches that combine video, audio and physiological signals to recognize the state of the user. Particular attention is given to identifying frameworks that are designed for video analytics in real-time applications and offer high accuracy. This theoretical foundation serves as a foundation for understanding how tools and libraries can be evaluated and compared.

In the second step, a systematic search is conducted to identify tools and libraries that support emotion detection of users. The primary focus will be on tools that work with video data, employing techniques like facial expression analysis, action unit detection, and facial landmark tracking. To provide a broader perspective, tools based on other modalities, such as voice analysis, text processing, and physiological data, are also included. The aim is to find tools that are suitable for recognizing the emotional states of users when interacting with a website in order to obtain real customer feedback.

The tool identification process is structured as follows.

- **Extensive Web Search:** Conducting comprehensive searches in databases, repositories, and platforms like GitHub and Kaggle to discover current tools and access their documentation.
- **Inclusion Criteria:** Selecting tools based on their ability to detect emotions using video data, functionality (range of detectable emotional states), accessibility (free, licensed or paid), feasibility within project constraints, usability, and suitability for real-time data processing and integration.
- **Feature Analysis:** Collecting information about key features such as supported input types, availability of pre-trained models, supported frameworks, and ease of integration.
- **Prototyping Implementation:** Using one of the tools to create a simulation of video-based emotion recognition

The ultimate aim of this process is to compile a comprehensive list of tools detailing their features and functionalities. This list will serve as the basis for comparing different tools, assessing their strengths and limitations, and identifying those most suitable for the specific use cases.

By combining insights from the literature with practical information about existing tools, this methodology provides a well-rounded understanding of the current technologies available for emotion detection of the user.

4 Results

There are numerous tools for recognizing and detecting emotions based on facial expressions, body movements, eye tracking, speech and gestures. In our study, we identified 56 tools ranging from open source libraries to commercial APIs based on recognizing emotions in videos or speech and emotions. To provide a more holistic overview, we have also included selected tools that specialize in speech-based emotion recognition, as these can complement facial analysis in multimodal approaches. Our selection of tools in Table 1 is a targeted compilation. Although there are many other technologies available, we have deliberately chosen this selection because it best meets our requirements. Their inclusion does not mean that they are superior to the tools not listed, and their exclusion does not mean that they are irrelevant. Our selection includes the most promising solutions for our specific research objective: user state detection.

Of the tools identified, 42 are geared towards analysing facial expressions. However, not all of them support real-time emotion recognition. Some focus on post-processing static images or videos and are therefore more suitable for retrospective analysis than for real-time emotion classification. Others function as platforms for user testing, recording users as they interact with a website for subsequent analysis. Since our goal is to dynamically capture a user's emotional state during interaction, real-time processing and emotion detection is a crucial factor for our evaluation.

To systematically refine our selection, we applied three important filter criteria to the collected tools: *Facial Expression-Based Emotion Recognition* to ensure that the tools directly analyse facial expressions and do not rely on additional modalities. The *free availability* was considered to favour tools that are accessible for research or academic use without commercial restrictions. Finally, the criterion of real-time processing was crucial to enable dynamic classification of emotions using live video data rather than limiting the analysis to pre-recorded media. Using these criteria, the selection focused on 13 tools, which we further analysed based on their methodological approaches, integration flexibility and suitability for real-time emotion recognition tasks. These tools vary in implementation, ranging from independent frameworks for comprehensive facial analysis to solutions integrating multiple features such as face detection, landmark tracking and head pose estimation.

4.1 Deep Learning-Based Tools

Our collection includes several stand-alone libraries that rely on deep learning for emotion recognition, utilizing convolutional neural networks (CNNs) and pre-trained models to enhance accuracy and efficiency. These frameworks range from lightweight real-time solutions to more advanced architectures for detailed facial analysis. **DeepFace** is a lightweight Python library that provides a high-level API for face detection, face recognition and emotion recognition. It collects several state-of-the-art models such as VGG-Face, OpenFace, DeepID, ArcFace and Dlib, which are also used by other emotion classification tools [52]. **OpenFace** provides an advanced toolkit that includes facial feature recognition, head pose estimation and emotion analysis, focusing on a detailed analysis of facial behaviour. It serves as a foundation for traditional machine learning-based emotion recognition. Many emotion recognition systems integrate OpenFace's facial mark extraction for pre-processing [6]. **Mini-Xception**, another deep learning-based solution, is a lightweight architecture optimised for emotion recognition and designed for efficient performance in real-time applications [4]. **HSEmotion**, a Python library for recognising emotions, enables easy integration into existing workflows, but does not offer more comprehensive functions for facial analysis [50]. **TensorFlow 101** is a deep learning-based framework for analysing facial emotions built on a custom CNN architecture. It has been trained on the FER2013 dataset and optimises speed and performance for both image-based and real-time applications. The repository provides step-by-step documentation, making it particularly useful for researchers and developers who want to customise their own emotion recognition models [51]. Similarly, **CLCM** (Custom Lightweight CNN-based Model) utilises CNN architectures and has been trained with the FER2013 and AffectNet datasets, focusing on real-time emotion classification but with less documentation available for easy implementation [25]. **FFEM** (Fast Facial Emotion Monitoring) is an open-source tool developed for facial emotion detection and monitoring. It integrates MediaPipe for facial recognition and utilises DeepFace for emotion classification by using deep learning based facial analysis. Developed with OpenCV, it enables real-time emotion recognition and acts as a lightweight framework for analysing facial expressions [31].

4.2 Multiple modalities

Some tools integrate multiple modalities beyond facial recognition combining speech, text, EEG, or additional visual data. These approaches enhance emotion recognition analysis by leveraging diverse input sources for a more comprehensive understanding of user states. **Multimodal Emotion Recognition** is a web-based platform that analyses facial expressions in addition to speech and text input. It uses the XCeption architecture and depth-separable convolutions to analyse facial expressions [39]. Another multimodal solution, **MindLink-Eumpy**, combines facial expression recognition with EEG data to improve the accuracy of emotion recognition. This open-source Python toolbox uses CNN and transfer learning to reduce the common problem of limited annotated data. By combining the classification decisions from EEG and face models, a more precise emotion analysis is made possible [37]. **PAZ** (Perception for Autonomous Systems) is a Python library for various computer vision tasks, including emotion recognition, face recognition, object recognition and key point estimation. Using HaarCascades and mini-Xception, it performs emotion recognition trained on FER2013, but does not primarily focus on facial expressions. It supports hierarchical processing and enables 2D and 3D key point recognition as well as broader applications in autonomous systems and machine vision [3].

4.3 Web- and API-based solutions

Several tools serve as API-based solutions for browser and web application integration, enabling real-time facial and emotion recognition. **Face-API.js**, a JavaScript library that uses TensorFlow, provides real-time face recognition and emotion detection for browser-based environments. It uses lightweight CNN models and supports face recognition and facial expression recognition with relatively small model size [45]. **Moody** also captures a multimodal approach and serves as a web application for recognising emotions in video conferences such as Zoom or Microsoft Teams. The application captures the emotional reactions of participants in real time by recording the screen during the session and analysing the emotions. The recorded emotions are then available for statistical analyses [44]. **Human API** is a web and Node.js-enabled solution designed for comprehensive human analysis, including 3D face recognition, rotation tracking, posture analysis and iris tracking. Unlike tools that focus solely on recognising emotions, Human API offers a modular architecture where different machine learning models can be applied depending on the use case. Although it also includes the recognition of facial expressions, its main focus is on the multi-faceted analysis of human behaviour [40].

4.4 Prototype Implementation

Building upon the evaluation of tools for emotion recognition, we implemented a prototype based on the DeepFace Python library due to its ease of use, lightweight nature, and comprehensive facial analysis capabilities [52]. The prototype explores the potential of emotion recognition tools in providing valuable insights into user emotions and their applicability, particularly in the context of usability and UX during interaction with a website.

The implementation focuses on real-time emotion detection, accessible through a graphical user interface (GUI). It processes video input captured via a webcam, analysing facial expressions in each frame. Detected faces are highlighted with bounding boxes, and the dominant emotion is displayed in real time (Figure 1 Appendix A). This provides immediate feedback on user reactions, demonstrating how such a tool could be used in usability testing to identify, for example, moments of frustration or confusion during interaction (Figure 2 Appendix A). The goal is to illustrate how real-time emotion tracking could offer actionable data for evaluating user satisfaction levels.

To further explore the capabilities of emotion recognition, the prototype temporarily logs data in a CSV file, which is cleared after each session. This log includes timestamps, the respective dominant emotion, and the likelihood of all recognised emotions - neutral, happy, fear, surprise, angry, sad, and disgust. Utilising this data, the prototype generates two key visualisations: a bar chart that summarises the distribution of dominant emotions to reflect overall user satisfaction and a line graph that shows the intensity of each emotion over time to identify critical moments in user interactions. For instance, sudden spikes in negative emotions such as "anger" or "fear" during specific tasks could indicate usability pain points, whereas consistent positive emotions like "happiness" might reflect satisfaction. These visualisations illustrate the potential of emotion data to enhance UX research by highlighting key emotional responses.

The prototype's GUI is intentionally designed to be simple and accessible, ensuring ease of use while emphasising its emotion detection capabilities. By integrating automated data handling and visualisations, the prototype demonstrates how such tools could enhance UX research by making emotional states more visible and measurable. It is important to note that this prototype is not a fully developed solution but rather a proof of concept, illustrating the potential applications of emotion recognition from video data. The objective is to demonstrate the theoretical potential of these tools and their contribution to a more profound comprehension of user emotions in digital interactions.

Name	Tool Category	Real-time	Modality	Pricing
Evalyzer[1]	User testing	No	Facial/Speech	Limited
lookback.io[2]	User research platform	Yes	Video, Audio, Screen Sharing	Limited
Userlook[3]	User testing	No	Video, Audio, Screen Recording	limited
Affectiva[4]	SDK / API	Yes	Facial	Limited
BeEmotion.ai[5]	Software	Yes	Facial	Limited
CERT[6]	Software	Yes	Facial	Limited
CLCM[7]	(Python) model	Yes	Facial	Free
DeepFace[8]	Lightweight Python library	Yes	Facial	Free
EmoPy[9]	Python toolkit	No	Facial	Free
EmoVu[10]	SDK / API	Yes	Facial	Limited
EmotionalDAN[11]	Model	No	Facial	Free
Emotiva[12]	API	Yes	Facial	Limited
FFEM[13]	Python library	Yes	Facial	Free
Face-API.js[14]	API	Yes	Facial	Free
FaceReader Noldus[15]	API	Yes	Facial	Limited
Face++[16]	API / SDK	No	Facial	Limited
face_classification[17]	Model	Yes	Facial	Free
FindFace[18]	SDK	Yes	Facial	Limited
HSEmotion[19]	Python library	Yes	Facial	Free
Human[20]	API	Yes	Facial	Free
iMotions (Facial Expression Module)[21]	Software module	Yes	Facial	Limited
itracks[22]	Plugin / App	No	Facial	Limited
Kairos[23]	API	No	Facial	Limited
Luxand.cloud FaceAPI[24]	API	No	Facial	Limited
Luxand FaceSDK[25]	SDK	Yes	Facial	Limited
MindLink-Eumpy[26]	Python toolbox	Yes	Facial/EEG	Free
Mini-Xception[27]	Framework / Architecture	Yes	Facial	Free
MorphCast[28]	SDK / API / Web app	Yes	Facial	Limited

OpenFace[29]	Advanced, open-source toolkit	Yes	Facial	Free
PAZ[30]	Python library	Yes	Facial	Free
SHORE[31]	SDK / API	Yes	Facial	Limited
SkyBiometry[32]	API	Yes	Facial	Limited
SmartClick[33]	API	Yes	Facial	Limited
TensorFlow 101[34]	Open-source GitHub repository (model)	Yes	Facial	Free
Touchy Feely[35]	(Python) framework / architecture / model	No	Facial	Free
VeriLook[36]	SDK	Yes	Facial	Limited
Visage Technologies[37]	SDK	Yes	Facial	Limited
Imentiv AI[38]	Web application	No	Facial/Speech	Limited
Microsoft Cognitive Services[39]	API	Yes	Facial/Speech	Limited
MixedEmotions[40]	Toolbox / Platform	No	Facial/Speech	Free
Moody[41]	Web application	Yes	Facial/Speech	Free
Multimodal-Emotion-Recognition[42]	Web application	Yes	Facial/Speech	Free
DELTA[43]	Platform / Toolkit	No	Speech	Free
ERTK[44]	Python library / Toolkit	No	Speech	Free
EmoBox[45]	Python toolkit	No	Speech	Free
emotion2vec[46]	Python model	No	Speech	Free
emotion2vec+[47]	Python model	No	Speech	Free
FunASR[48]	Toolkit	Yes	Speech	Free
Good Vibration[49]	API	Yes	Speech	Limited
I-Motions (Voice Analysis Module)[50]	Software module	Yes	Speech	Limited
Multimodal-Speech-Emotion[51]	Open-source GitHub repository (model)	No	Speech	Free
Multimodal-Speech-Emotion-Recognition[52]	Open-source GitHub repository (models)	No	Speech	Free

OpenSMILE[53]	C++ toolkit / Python library	Yes	Speech	Limited
PyAudioAnalysis[54]	Python library	Yes	Speech	Free
SenseVoice[55]	Python API	No	Speech	Free
Vokaturi[56]	SDK / API	Yes	Speech	Limited

Table 1: List of tools sorted by modality and alphabetically

5 Discussion

This discussion critically evaluates the tools identified in our study, comparing their strengths and limitations to determine their applicability for real-time user state detection. Based on this evaluation, we justify the selection of DeepFace as the most appropriate tool for our prototype implementation. Furthermore, we reflect on the main limitations encountered in real-world applications of emotion recognition and outline future research directions to address these challenges.

5.1 Selection Criteria

In addition to the filtering criteria introduced in Section 4, the selection of a suitable emotion recognition tool for our prototype depends on several factors. These include classification accuracy, ease of integration, and computational efficiency. While various tools offer emotion recognition capabilities, their applicability for real-time usability testing requires careful consideration.

First, classification accuracy is critical for reliable emotion detection, especially in real-time interactions where immediate feedback is important. Misclassifications can distort insights into user satisfaction and usability, potentially leading to incorrect conclusions. Thus, selecting a tool that can accurately predict emotions is critical to minimising false positives and negatives in dynamic testing environments.

Second, ease of integration is a decisive factor, as the chosen tool must fit seamlessly into existing architectures without excessive development overhead. This favours solutions with well-documented and straightforward implementation processes that minimise the complexity of building an emotion recognition system from scratch.

Finally, computational efficiency is a key consideration, as real-time emotion recognition requires minimal latency to avoid disrupting the UX and to ensure that all expressed emotions are captured. The tool should process incoming video frames quickly and provide near-instantaneous emotion classification without causing noticeable delays or frame drops. Excessive processing time leads to a delay in the display of detected emotions, reducing the effectiveness of real-time feedback in usability testing. Therefore, selecting a tool optimised for real-time performance is essential to maintain a smooth and responsive user evaluation process.

5.2 Evaluation of Presented Tools

Among the deep learning-based tools, DeepFace, OpenFace, Mini-Xception, TensorFlow 101, CLCM, HSEmotion, and FFEM stand out for their ability to perform facial expression-based emotion recognition using CNN-based architectures [52,6,4,51,25,50,31]. DeepFace integrates several state-of-the-art models [52], allowing for flexibility in model selection while maintaining a user-friendly API for easy integration.

OpenFace, on the other hand, offers advanced facial behaviour analysis, including facial feature recognition, head pose estimation, and gaze tracking [6]. However, it does not include built-in emotion recognition, requiring additional models or external processing to infer emotional states from facial behaviour.

Mini-Xception, TensorFlow 101, CLCM, and FFEM are not ready-to-use emotion recognition tools but rather model architectures or frameworks that require additional training and integration efforts into larger emotion classification systems [4,51,25,31]. While TensorFlow 101 provides extensive documentation [51], CLCM lacks sufficient documentation [25], making implementation more complex. FFEM integrates MediaPipe for facial recognition and utilises DeepFace for emotion classification [31], acting as a framework rather than a fully independent solution. These factors make it less adaptable than DeepFace, which offers pre-trained models that can be easily switched and fine-tuned for specific use cases [52].

In terms of multimodal approaches, MindLink-Eumpy and Multimodal Emotion Recognition incorporate multiple input sources, such as EEG data and speech, to improve classification accuracy [37,39]. While these approaches provide valuable insights into user states, their reliance on additional hardware and specialised datasets introduces complexity that makes them less practical for a stand-alone, lightweight prototype.

Web-based and API-driven solutions such as Face-API.js, Human API, and Moody are optimised for seamless online integration, making them suitable for browser-based applications [45,40,44]. However, their dependence on cloud processing or JavaScript environments reduces their applicability for standalone Python implementations. In particular, Moody is a pre-configured web application developed specifically for analysing user emotion during video conferencing on platforms such as Zoom or Microsoft Teams [44]. This limits its adaptability to other use cases.

In conclusion, DeepFace emerges as the most suitable tool for the prototype implementation, taking into account the defined selection criteria. Its classification accuracy benefits from a range of state-of-the-art models, allowing easy adaptation to specific use cases. The ability to switch between models such as VGG-Face, OpenFace, DeepID, ArcFace, and Dlib with minimal code changes enhances its flexibility and ensures optimal performance in different applications. In addition, its well-documented Python library enables seamless integration with minimal effort, while its computational efficiency ensures real-time processing with no noticeable latency, making it ideal for usability testing and UX evaluation.

5.3 Limitations & Future Research

While DeepFace provides a strong foundation for real-time facial emotion recognition, several limitations remain.

In this study, we adopt the assumption proposed by Ekman [16] that basic emotions are universally expressed and recognised across different cultures, regardless of a person's background or appearance. However, more recent research challenges this notion, arguing that emotional expressions are not universally distinct or rigidly separable but are instead shaped by cultural learning and contextual interpretation [29,30]. It suggests that individuals from different cultural backgrounds rely on varying facial cues to interpret emotions, resulting in different recognition patterns. This raises the question of whether automated emotion recognition tools that rely on predefined emotion categories, can accurately capture the full complexity of human emotions.

The implemented prototype further illustrates these challenges, where neutral facial expressions are often misclassified as slightly negative. Certain individuals are consistently categorised as 'sad' more often than others, despite displaying similarly neutral expressions. In addition, emotional transitions are not always clearly distinguishable, as small changes in facial expression often lead to fluctuating predictions between multiple categories. This suggests that subtle variations in muscle relaxation or tension are not reliably detected by the tool. Furthermore, the accuracy of emotion recognition is highly dependent on environmental factors, such as adequate lighting and minimal background distraction. Poor lighting or shadows lead to inconsistent recognition, whereas the presence of additional objects in the frame or excessive movement leads to misclassifications or instability in predictions. While real-time emotion recognition is effective at identifying strong and expressive emotions, it is unable to adequately handle the subtleties of human facial expressions. This limitation is particularly relevant in contexts such as usability testing and user satisfaction analysis, as explored in the prototype use case.

Many available emotion detection tools categorise emotions based on broad labels such as 'happy', 'sad', or 'fear'. However, in real-world contexts, emotional expressions tend to be more nuanced, especially in digital interactions. To illustrate this point, consider the scenario of an online shopping experience. In such a context, a customer who is unable to purchase a preferred product may not visibly show signs of extreme sadness, but may instead express disappointment—a more subtle variation that is more difficult to accurately identify. This highlights the need to refine emotion recognition tools to capture micro-expressions and more context-specific emotional states, ensuring a more accurate and meaningful interpretation of user emotions. Additionally, improving robustness to environmental factors would increase the reliability of these tools, making them more adaptable to diverse real-world conditions.

Beyond technical limitations, data privacy concerns pose a significant challenge to real-time emotion recognition. As the technology relies on the capture and analysis of facial expressions, it cannot be deployed in public environments without the explicit consent of the user. Users must be fully aware that their facial data is being recorded and processed, as failure to inform them raises ethical and legal concerns. Due to these limitations, usability testing with facial emotion recognition is currently only feasible under controlled laboratory conditions, where participants can give informed consent prior to data collection. This limits the ability to conduct large-scale, real-world evaluations in natural user environments, such as analysing customer emotions during private online shopping sessions.

These privacy concerns also underscore the need for transparency and explainability in emotion recognition tools to foster user trust and ensure the ethical use of such tools in real-world applications. Users should not only be informed about the collection and processing of their emotional data but also have a clear understanding of how the system interprets their expressions. However, achieving this level of transparency is particularly challenging due to the widespread use of deep learning approaches such as CNNs. These models act as opaque black boxes [8], making it difficult to explain why facial expressions are classified in a particular way. Without clear insights into how decisions are made, users may find it difficult to trust the fairness and reliability of the tools. Future research should therefore focus on developing more interpretable and transparent models to ensure that emotion recognition can be used responsibly, taking into account ethical and legal issues.

Finally, another important area for future research is the integration of multimodal approaches to emotion recognition [10]. This could help to alleviate some of the limitations observed in the prototype described above. As facial expressions alone may not always fully convey and represent an emotional state, the inclusion of additional modalities could improve classification accuracy [10], especially in ambiguous cases. Examples of this are the aforementioned speech analysis or physiological signals. Additionally, multimodality could reduce the tool's dependence on optimal environmental conditions. Complementary cues from speech or biometric sensors help to maintain reliable emotion classification when facial analysis is negatively affected by poor lighting conditions. By using a combination of modalities, emotion recognition tools could become more adaptive and context-aware, leading to a more precise and meaningful interpretation of user emotions.

6 Conclusion

This study investigated the potential of automatic emotion detection in the context of usability testing and UX research by analysing user emotions and their impact on marketing and product development. Understanding user emotions can provide valuable insights into UX and usability issues and allows the identification of problematic design elements in websites or software applications. By observing users as they perform certain tasks, areas of confusion, frustration or dissatisfaction can be recognised. This process can be automated using classification algorithms, enabling real-time emotion detection to highlight points of user dissatisfaction.

However, the use of such technology raises data protection concerns and requires the explicit consent of the user. Therefore, its application is currently most likely in controlled usability test environments and not in large-scale, real-world deployments. Aside from usability testing, emotion recognition technology is already being used in other areas, such as tracking attention span in autonomous driving or assessing emotional sensitivity in psychology. Recognising emotions is therefore an important topic.

To increase accuracy and robustness, the integration of emotion recognition with complementary UX research methods such as eye tracking, screen tracking and click behaviour analysis could improve the reliability of emotion classification. These additional inputs enable a more holistic understanding of the user's emotions and reduce the reliance on pure facial expressions, which can be influenced by factors such as lighting conditions, camera resolution and obstacles such as glasses or masks. Customised classification models tailored to specific user characteristics such as glasses or facial disguises could further improve recognition accuracy.

This study addressed a gap in the literature by systematically examining existing real-time emotion recognition tools in video-based UX research. The results suggest that while current tools effectively recognise stringent emotional expressions, they struggle with subtle emotions and rapid micro-expressions that are common in digital interactions. Given the rapid advances in the field of artificial intelligence, future developments are expected to improve the accuracy, reliability and interpretability of emotion recognition tools, particularly in UX research and usability testing. Future research should focus on refining these models to improve their ability to capture contextual emotional nuances and user frustration patterns in UX environments.

A multimodal approach that integrates facial analysis, eye movement tracking and audio sentiment recognition has the potential to improve the accuracy of emotion recognition. By utilising multiple data sources, emotion recognition can become a more effective tool for identifying usability issues, improving website and software design, and ultimately increasing user satisfaction.

A Appendix: Simulation

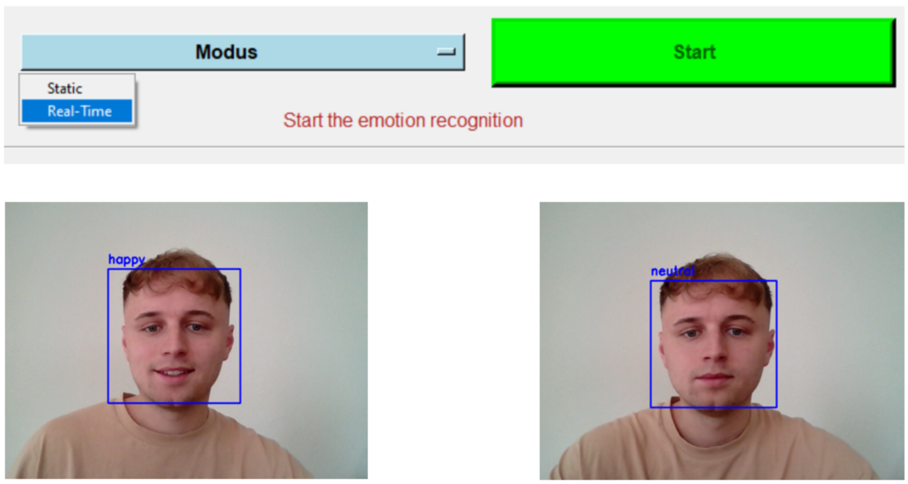

Fig. 1. Simulation of real-time emotion recognition

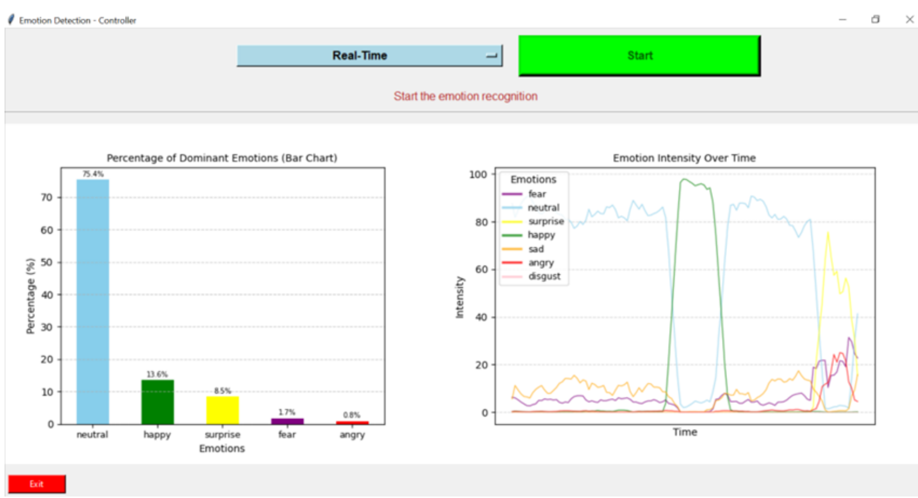

Fig. 2. Statistics after running real-time emotion detection

B Appendix: Tool References

1. Evalyzer: https://www.evalyzer.com/en/product/
2. lookback.io: https://www.lookback.com/usability-testing-with-lookback
3. Userlook: https://userlook.co/
4. Affectiva: https://www.affectiva.com/
5. BeEmotion.ai: https://www.beemotion.ai/discovery-ai
6. CERT: CERT Paper
7. CLCM: CLCM Paper
8. DeepFace: https://github.com/serengil/deepface
9. EmoPy: https://github.com/thoughtworksarts/EmoPy
10. EmoVu: https://www.eyeris.ai/technology
11. EmotionalDAN: https://github.com/IvonaTau/emotionaldan
12. Emotiva: https://emotiva.it/en/emotion-recognition-2/
13. FFEM: https://github.com/WiseGeorge/Fast-Facial-Emotion-Monitoring-FFEM-Package
14. Face-API.js: https://github.com/justadudewhohacks/face-api.js
15. FaceReader Noldus: https://www.noldus.com/facereader
16. Face++: https://www.faceplusplus.com/emotion-recognition/
17. face_classification: https://github.com/oarriaga/face_classification
18. FindFace: https://ntechlab.com/findface-sdk/
19. HSEmotion: https://github.com/av-savchenko/face-emotion-recognition
20. Human API: https://github.com/vladmandic/human
21. iMotions (Facial Expression Module): https://imotions.com/products/imotions-lab/modules/fea-facial-expression-analysis/
22. itracks: https://www.itracks.com/
23. Kairos: https://face.kairos.com/
24. Luxand.cloud FaceAPI: https://luxand.cloud/face-api
25. Luxand FaceSDK: https://www.luxand.com/facesdk/
26. MindLink-Eumpy: MingLink-Eumpy Paper
27. Mini-Xception: https://github.com/AmrElsersy/Emotions-Recognition
28. MorphCast: https://www.morphcast.com/
29. OpenFace: https://github.com/TadasBaltrusaitis/OpenFace
30. PAZ: https://github.com/oarriaga/paz
31. SHORE: https://shore-dl.iis.fraunhofer.de/dc4d8d4008793ad5642da33d2e4fc16a9255490e/
32. SkyBiometry: https://skybiometry.com/demo/face-detect/
33. SmartClick: https://smartclick.ai/api/emotion-detection/
34. TensorFlow 101: https://github.com/serengil/tensorflow-101
35. Touchy Feely: Touchy Feely Paper
36. VeriLook: https://neurotechnology.com/verilook.html
37. Visage Technologies: https://visagetechnologies.com/demo/
38. Imentiv AI: https://imentiv.ai/
39. Microsoft Cognitive Services: https://azure.microsoft.com/de-de/solutions/ai, https://azure.microsoft.com/en-us/blog/cross-channel-emotion-analysis-in-microsoft-video-indexer/

40. MixedEmotions: https://github.com/MixedEmotions/MixedEmotions
41. Moody: https://github.com/COINS-SS21/moody
42. Multimodal-Emotion-Recognition: https://github.com/maelfabien/Multimodal-Emotion-Recognition
43. DELTA: https://github.com/Delta-ML/delta
44. ERTK: Emotion Recognition ToolKit Paper
45. EmoBox: https://github.com/emo-box/EmoBox
46. emotion2vec: https://github.com/ddlBoJack/emotion2vec
47. emotion2vec+ : https://github.com/ddlBoJack/emotion2vec
48. FunASR: https://github.com/modelscope/FunASR
49. Good Vibration: https://goodvibrations.nl/
50. iMotions (Voice Analysis Module): https://imotions.com/products/imotions-lab/modules/voice-analysis/
51. Multimodal-Speech-Emotion: https://github.com/david-yoon/multimodal-speech-emotion
52. Multimodal-Speech-Emotion-Recognition: https://github.com/Demfier/multimodal-speech-emotion-recognition
53. OpenSMILE: https://www.audeering.com/research/opensmile/
54. PyAudioAnalysis: https://github.com/tyiannak/pyAudioAnalysis
55. SenseVoice: https://github.com/FunAudioLLM/SenseVoice
56. Vokaturi: https://vokaturi.com/

References

1. AI, G.: Google research datasets (2024), https://research.google/resources/datasets/, accessed: 2024-04-28
2. Apple Inc.: Use face id while wearing a mask with iphone 12 and later (2024), https://support.apple.com/en-us/102452, accessed: 2025-01-30
3. Arriaga, O., Valdenegro-Toro, M., Muthuraja, M., Devaramani, S., Kirchner, F.: Perception for autonomous systems (PAZ). CoRR **abs/2010.14541** (2020), https://arxiv.org/abs/2010.14541
4. Arriaga, O., Valdenegro-Toro, M., Plöger, P.: Real-time convolutional neural networks for emotion and gender classification. CoRR **abs/1710.07557** (2017), http://arxiv.org/abs/1710.07557
5. Avci, U., Akkurt, G., Unay, D.: A pattern mining approach in feature extraction for emotion recognition from speech. In: Proceedings of the International Conference on Engineering and Technology. Yasar University, Izmir, Turkey (2018)
6. Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.P.: Openface 2.0: Facial behavior analysis toolkit. In: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018). pp. 59–66 (2018). https://doi.org/10.1109/FG.2018.00019
7. Banse, R., Scherer, K.R.: Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology **70**(3), 614–636 (1996). https://doi.org/10.1037/0022-3514.70.3.614
8. Buhrmester, V., Münch, D., Arens, M.: Analysis of explainers of black box deep neural networks for computer vision: A survey. Machine Learning and Knowledge Extraction **3**(4), 966–989 (2021). https://doi.org/10.3390/make3040048, https://www.mdpi.com/2504-4990/3/4/48
9. Cambria, E.: Affective computing and sentiment analysis. IEEE Intelligent Systems **31**(2), 102–107 (2016). https://doi.org/10.1109/MIS.2016.31
10. Castellano, G., Kessous, L., Caridakis, G.: Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech, pp. 92–103. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85099-1_8, https://doi.org/10.1007/978-3-540-85099-1_8
11. Chen, Z., Feng, X., Zhang, S.: Emotion detection and face recognition of drivers in autonomous vehicles in iot platform. Image and Vision Computing **128**, 104569 (2022). https://doi.org/https://doi.org/10.1016/j.imavis.2022.104569, https://www.sciencedirect.com/science/article/pii/S0262885622001986
12. Chyan, P., Achmad, A., Nurtanio, I., Areni, I.S.: A deep learning approach for stress detection through speech with audio feature analysis. In: 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). pp. 1–5 (2022). https://doi.org/10.1109/ICITISEE57756.2022.10057845
13. Curtis, T., Abratt, R., Dion, P., Rhoades, D.L.: Customer satisfaction, loyalty and repurchase: Some evidence from apparel consumers. Review of Business **32**(1) (2011), https://commons.erau.edu/db-management/20
14. Donato, G., Bartlett, M., Hager, J., Ekman, P., Sejnowski, T.: Classifying facial actions (1999). https://doi.org/10.1109/34.799905
15. Dzedzickis, A., Kaklauskas, A., Bucinskas, V.: Human emotion recognition: Review of sensors and methods. Sensors (Basel) **20**(3), 592 (2020). https://doi.org/10.3390/s20030592

16. Ekman, P.: An argument for basic emotions. Cognition and Emotion **6**(3-4), 169–200 (1992). https://doi.org/10.1080/02699939208411068
17. Ekman, P., Friesen, W.V.: Facial action coding system. In: Environmental Psychology & Nonverbal Behavior. Consulting Psychologists Press, Palo Alto, CA, USA (1978). https://doi.org/10.1037/t27734-000
18. ELsayed, Y., ELSayed, A., Abdou, M.A.: An automatic improved facial expression recognition for masked faces. Neural Computing and Applications **35**(20), 14963–14972 (2023). https://doi.org/10.1007/s00521-023-08498-w
19. Eyeris: Technology overview. https://www.eyeris.ai/technology (2025), accessed: 2025-01-05
20. Folkman, S.: Stress: Appraisal and coping. In: Gellman, M.D., Turner, J.R. (eds.) Encyclopedia of Behavioral Medicine, pp. 1913–1915. Springer, New York, NY (2013). https://doi.org/10.1007/978-1-4419-1005-9_215
21. Gao, H., Yüce, A., Thiran, J.P.: Detecting emotional stress from facial expressions for driving safety. In: 2014 IEEE International Conference on Image Processing (ICIP). pp. 5961–5965 (2014). https://doi.org/10.1109/ICIP.2014.7026203
22. Garcia-Garcia, J.M., Penichet, V.M.R., Lozano, M.D.: Emotion detection: a technology review. In: Proceedings of the XVIII International Conference on Human Computer Interaction. Interacción '17, Association for Computing Machinery, New York, NY, USA (2017), https://doi.org/10.1145/3123818.3123852
23. Giroux, F., Léger, P.M., Brieugne, D., Courtemanche, F., Bouvier, F., Chen, S.L., Tazi, S., Rucco, E., Fredette, M., Coursaris, C., Sénécal, S.: Guidelines for collecting automatic facial expression detection data synchronized with a dynamic stimulus in remote moderated user tests. In: Kurosu, M. (ed.) Human-Computer Interaction. Theory, Methods and Tools. pp. 243–254. Springer International Publishing, Cham (2021)
24. Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., et al.: Challenges in representation learning: A report on three machine learning contests. In: International Conference on Neural Information Processing (2015)
25. Gursesli, M.C., Lombardi, S., Duradoni, M., Bocchi, L., Guazzini, A., Lanata, A.: Facial emotion recognition (fer) through custom lightweight cnn model: Performance evaluation in public datasets. IEEE Access **12**, 45543–45559 (2024). https://doi.org/10.1109/ACCESS.2024.3380847
26. Hallowell, R.: The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. International Journal of Service Industry Management **7**(4), 27–42 (1996). https://doi.org/10.1108/09564239610129931
27. Hamm, J., Kohler, C.G., Gur, R.C., Verma, R.: Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. Journal of Neuroscience Methods **200**(2), 237–256 (2011). https://doi.org/10.1016/j.jneumeth.2011.06.023
28. Huang, Y., Chen, F., Lv, S., Wang, X.: Facial expression recognition: A survey. Symmetry **11**(1189) (2019). https://doi.org/10.3390/sym11101189
29. Jack, R.E.: Culture and facial expressions of emotion. Visual Cognition **21**(9-10), 1248–1286 (2013). https://doi.org/10.1080/13506285.2013.835367
30. Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G., Caldara, R.: Cultural confusions show that facial expressions are not universal. Current Biology **19**(18), 1543–1548 (2009). https://doi.org/https://doi.org/10.1016/j.cub.2009.07.051, https://www.sciencedirect.com/science/article/pii/S0960982209014778
31. Jorge Martínez Pazos, William Gómez Fernández, A.O.G.y.D.B.L.: Monitoring emotional response during mental health therapy. Revista Cubana de Ciencias

Informáticas **17**(4) (2023), https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=2806
32. Khan, A.R.: Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges. Information-an International Interdisciplinary Journal **13**(268) (2022). https://doi.org/10.3390/info13060268
33. Ko, B.C.: A brief review of facial emotion recognition based on visual information. Sensors **18**(401) (2018). https://doi.org/10.3390/s18020401
34. Kolakowska, A., Landowska, A., Szwoch, M., Szwoch, W., Wróbel, M.: Emotion recognition and its applications. In: Hippe, Z., Kulikowski, J., Mroczek, T., Wtorek, J. (eds.) Human-Computer Systems Interaction: Backgrounds and Applications 3, pp. 51–62. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-08491-6_5
35. Lazarus, R.S., Folkman, S.: Stress, Appraisal, and Coping. Springer, New York (1984)
36. Lerner, J.S., Dahl, R.E., Hariri, A.R., Taylor, S.E.: Facial expressions of emotion reveal neuroendocrine and cardiovascular stress responses. Biological Psychiatry **61**, 253–260 (2007), https://api.semanticscholar.org/CorpusID:52800733
37. Li, R., Liang, Y., Liu, X., Wang, B., Huang, W., Cai, Z., Ye, Y., Qiu, L., Pan, J.: Mindlink-eumpy: An open-source python toolbox for multimodal emotion recognition. Frontiers in Human Neuroscience **15**, 621493 (2021). https://doi.org/10.3389/fnhum.2021.621493, https://www.frontiersin.org/articles/10.3389/fnhum.2021.621493/full
38. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z.: Ck+ dataset (2023). https://doi.org/10.34740/KAGGLE/DSV/4957114, https://www.kaggle.com/dsv/4957114
39. Mael, F.: Multimodal emotion recognition (2024), https://github.com/maelfabien/Multimodal-Emotion-Recognition/tree/master, accessed: 2024-02-11
40. Mandic, V.: Human: Real-time face and body analysis api (2024), https://vladmandic.github.io/human/demo/index.html, accessed: 2024-02-11
41. Matsuo, N., Hayakawa, S., Harada, S.: Technology to detect levels of stress based on voice information. Fujitsu Sci. Tech. J. **51**(4), 48–54 (2015), https://api.semanticscholar.org/CorpusID:11680031
42. McBain, H.: Are customers always right. The Rotarian pp. 32–33 (November 1944), https://books.google.com
43. McEwen, B., Sapolsky, R.: Stress and your health. The Journal of Clinical Endocrinology and Metabolism **91** (2006). https://doi.org/10.1210/JCEM.91.2.9994
44. Moody: Moody: Emotion analysis for online meetings (2024), https://www.moody.digital/meetings, accessed: 2024-02-11
45. Mühler, V.: Face-api.js: Javascript face recognition and emotion detection library (2024), https://github.com/justadudewhohacks/face-api.js, accessed: 2024-02-11
46. or Organization Name, A.: Dataset title. https://www.kaggle.com/datasets (Year), accessed: 2024-12-21
47. Pisanski, K., Sorokowski, P.: Human stress detection: Cortisol levels in stressed speakers predict voice-based judgments of stress. Perception **50**(1), 80–87 (2021). https://doi.org/10.1177/0301006620978378
48. Rinck, M., Primbs, M.A., Verpaalen, I.A.M., Bijlstra, G.: Face masks impair facial emotion recognition and induce specific emotion confusions. Cognitive Research: Principles and Implications **7**(1), 83 (2022). https://doi.org/10.1186/s41235-022-00430-5, https://doi.org/10.1186/s41235-022-00430-5

49. Russell, J.A., Lewicka, M., Niit, T.: A cross-cultural study of a circumplex model of affect. Journal of Personality and Social Psychology **57**(5), 848–856 (1989). https://doi.org/10.1037/0022-3514.57.5.848
50. Savchenko, A.V.: Facial expression and attributes recognition based on multi-task learning of lightweight neural networks. In: Proceedings of the 19th International Symposium on Intelligent Systems and Informatics (SISY). pp. 119–124. IEEE (2021), https://arxiv.org/abs/2103.17107
51. Serengil, S.I.: tensorflow-101. https://github.com/serengil/tensorflow-101 (2021)
52. Serengil, S.I., Ozpinar, A.: Hyperextended lightface: A facial attribute analysis framework. In: 2021 International Conference on Engineering and Emerging Technologies (ICEET). pp. 1–4. IEEE (2021). https://doi.org/10.1109/ICEET53442.2021.9659697, https://ieeexplore.ieee.org/document/9659697/
53. Steppan, M., Zimmermann, R., Fürer, L., Southward, M., Koenig, J., Kaess, M., Kleinbub, J.R., Roth, V., Schmeck, K.: Machine learning facial emotion classifiers in psychotherapy research: A proof-of-concept study. Psychopathology **57**(3), 159–168 (11 2023). https://doi.org/10.1159/000534811, https://doi.org/10.1159/000534811
54. Steptoe, A., Kivimäki, M.: Stress and cardiovascular disease: an update on current knowledge. Annual review of public health **34**, 337–54 (2013). https://doi.org/10.1146/annurev-publhealth-031912-114452
55. Teixeira, A., Rodrigues, M., Carneiro, D., Novais, P.: Horus: An emotion recognition tool. In: Bi, Y., Bhatia, R., Kapoor, S. (eds.) Intelligent Systems and Applications. pp. 126–140. Springer International Publishing, Cham (2020)
56. Viegas, C., Maxion, R.A., Hauptmann, A.G.: The seven faces of stress: Understanding facial activity patterns during cognitive stress. In: Proceedings of the 2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG). pp. 1–9 (2024). https://doi.org/10.1109/FG59268.2024.10581960
57. Vimaleswaran, B., Ratnayake, G.S.: E - therapy improvement monitoring platform for depression using facial emotion detection of youth. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). pp. 950–955 (2021). https://doi.org/10.1109/ICICV50876.2021.9388447